University of Rhode Island

DigitalCommons@URI

2017

Tools for Anti-Microbial Peptide Editable Database

Sunandha Sharma Acharya
University of Rhode Island, sunu.acharya@gmail.com

Follow this and additional works at: https://digitalcommons.uri.edu/theses

Recommended Citation

TOOLS FOR ANTI-MICROBIAL PEPTIDE EDITABLE DATABASE

BY

SUNANDHA SHARMA ACHARYA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

COMPUTER SCIENCE

UNIVERSITY OF RHODE ISLAND

2017

MASTER OF SCIENCE THESIS IN COMPUTER SCIENCE

OF

SUNANDHA SHARMA ACHARYA

APPROVED:

Thesis Committee:

Major Professor   Joan Peckham

Lenore M. Martin

Lisa Dipippo

Clinton Chichester III

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2017

ABSTRACT

The goal of this thesis is to create tools for Anti-Microbial Peptide Database AMPed, developed by Professor Lenore Martin's microbial peptide research group at University of Rhode Island. The database is for Researchers and Biologists, as they will be the main users. To develop a functioning and efficient database there should be tools that are relevant to the users of the database. In our case, the primary users are biologists, needing to calculate specific results or to visualize their data. The tools that are developed as a part for this thesis are the MIC (Minimal Inhibitory Concentration) and MLC (Minimal Lethal Concentration) calculator, the Mass spectrometry fragmentation prediction tool.

With the help of these tools, researchers can upload their lab data, perform the calculation and see results using graphical visualizations.

Also, in addition to this, normalization of the database is performed, as this is an editable database and we may need to change the schema as necessary (add and delete attributes) and re-normalize the database. We use the agile software development process to support these iterative changes.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

"**Antimicrobial peptides**(AMPs) are part of the innate immune response found among all classes of life. Fundamental differences exist between prokaryotic and eukaryotic cells that may represent targets for antimicrobial peptides. These peptides are potent, broad spectrum antibiotics that demonstrate potential as novel therapeutic agents. Antimicrobial peptides have been demonstrated to kill Gram negative and Gram positive bacteria, enveloped viruses, fungi and even transformed a cancerous cell. Unlike many conventional antibiotics, it appears that antimicrobial peptides may also have the ability to enhance immunity by functioning as immunomodulators [1]".

AMPed (Anti-Microbial Peptide editable database) [2], developed by Professor Lenore Martin's microbial peptide research group at the University of Rhode Island, was designed to provide a user-friendly resource for researchers seeking information regarding Antimicrobial Peptides. AMPed will serve as a catalogue of all known AMPs and other peptides, populated by information gathered from journals, research papers, and other databases. Existing protein databases, with broader subject matter, are often difficult and time-consuming to search for AMPs effectively. AMPed will organize and manage available AMP information. AMPed, in order to most effectively collect and disseminate pertinent information, will allow for researchers to upload

relevant research data directly to the database. Our database, still in the data curation phase, seeks to collect a variety of information sets on each peptide.

For example, the database encompasses information on peptides including their name, sequence, and molecular weight, and microbes including species name, type, and wall classification. Additionally, the database will glean data from other sources in order to further strengthen its capabilities. Structural information will be gathered from the Protein Data Bank, a repository with over 130,000 entries [3]. Through manual curation of entries onto our database and the pooling of data from established data banks, AMPed will prove a valuable research tool for a variety of applications.

Following are the goals of this thesis:

- Create MIC (minimal inhibitory concentration) and MLC (minimal lethal concentration) calculator tools.

- Create a Mass spectrometry prediction tool.

- Normalize the Database.

- Add important enhancements to the website, such as mission and vision, privacy policy and terms of use.

The MIC and MLC of an antibiotic or therapeutic is the lowest concentration of a drug that will inhibit the visible growth of an organism after overnight incubation or kill an organism, respectively [4]. These values can be influential in guiding the choice of antimicrobials used in a therapy situation. Additionally, regular surveillance of MICs is required due to a continuing decrease in susceptibility to the commonly used antibiotics in critically ill patients [5]. A tool that would be able calculate the MIC of a

given peptide against a given organism would be valuable in this context. Moreover, within the laboratory MICs can serve as a research tool to determine the in vitro activity of new antimicrobials. Our MIC calculator tool will serve a dual purpose in both the guidance of use of antimicrobes in a therapy situation, as well as the characterization of new antimicrobials in the research laboratory.

Our other tool, mass spectrometry is an analytical technique that is used in a variety of applications throughout many branches of science. For example, in chemistry, mass spectrometry can be used to determine the identity of an unknown compound. Specifically, relevant to the applications in biotechnology, mass spectrometry can be used as a technique to determine the sequence of amino acids in a peptide. With respect to AMPs, mass spectrometry can be best applied to two major areas. Firstly, mass spectrometry can be used to determine the amino acid sequence of an unknown AMP. This information is crucial in the discovery of novel antimicrobial agents. Secondly, when undergoing the synthesis of AMPs for the purpose of research and drug characterization, mass spectrometry can be used to validate that the purified peptide has the intended sequence. It is important that the amino acid sequence is that of the peptide of interest as often the sequence and structure determine its efficacy as an AMP. The mass spectrometry tool would serve as a tool in both of these areas. It would assist a researcher in identifying novel AMPs, as well as validate the isolation and purification of peptides of interest.

REFERENCES

[1] Wikipedia, Antimicrobial Peptides. URL:

https://en.wikipedia.org/wiki/Antimicrobial_peptides, Last accessed 10/10/2017

[2] AMPED, Database initially developed by the Dr. Martin's research Group, URL:

https://amped.uri.edu, Last accessed 10/10/2017

[3] PDB Current Holdings Breakdown. URL:

https://www.rcsb.org/pdb/statistics/holdings.do , Last accessed 10/10/2017

[4] O'Neill AJ, Chopra I. Expert Opin Investig Drugs. 2004 Aug;13(8):1045-63.

"*Preclinical evaluation of novel antibacterial agents by microbiological and

molecular techniques*."[Online] Available:

https://www.ncbi.nlm.nih.gov/pubmed/15268641

[5] **Turnidge J, Ferraro M, Jorgensen J.** 2011. Susceptibility Test Methods: General

Considerations, p 1115-1121. *In* Versalovic J, Carroll K, Funke G, Jorgensen J,

Landry M, Warnock D (ed),*Manual of Clinical Microbiology, 10th Edition*. ASM

Press, Washington, DC. doi: 10.1128/9781555816728.ch67 [Online]

Available: http://www.asmscience.org/content/book/10.1128/9781555816728.chap67

CHAPTER 2

Goal

Understand current environment of AMPed, and integrate changes to the current environment.

AMPed Current Environment

The AMPed database has been expanded for a couple of years and is under the supervision of Professor Lenore Martin. Now, a vast project, it was initiated as small class project by David Ryder, a biochemist, and Daniel Ducharme, a computer scientist, in CMB/CSC 522. George Konstantinidis then designed the first general database structure [1]. After that Tripti Garg, enhanced this structure on her thesis project and also created AMPed's first Web Interface [2]. Following her, Abraham Herrera worked on this project with the help and guidance of Professor Martin and students in subsequent CMB/CSC 522 classes. He created a data uploading feature for the user of the website, and also enhanced the database and GUI [3]; referenced here as DBS (Database) version 3.0 and AWI (Amped Web Interface) version 2.0. The enhanced DBS version of this project is 4.0 and AWI version is 3.0.

The site is written in PHP scripts with MySQL commands improved with Cascading Style Sheets (CSS), Asynchronous JavaScript and XML (AJAX), a client-side script that communicates to and from a server/database. In spite of all this work we still needed some tools for the websites which were relevant to the main users of the website; research biologists. For this project with help of Professor Martin and

students in the CMB/CSC 522 we were able to identify important tools, normalize the database and add more new features to the website. This chapter describes these enhancements.

AMPed Current System

Domain Name

AMPed is a repository, so to access it we need to have an interface. We created a customized domain name amped.uri.edu with a new IP address, 131.128.20.101, hosted in a secure GUI front end as https://amped.uri.edu (https)[4].

Hosting Server

As mentioned before my predecessor on this project Abraham Herrera, helped to move the AMPed website to a new virtualized Linux server running GNU/Linux.

The hosting server technical specifications are as follows:

- Processor: Pentium Xeon 3.3Ghz, 4 cores.

- Memory: 8 GB

- Disk space: 1TB

Biological data is ever increasing and accumulating. For the time being we have enough space to accumulate the required data. These specifications can be updated if a need arises. The download speed is how fast you can pull data from the server whereas upload speed is how fast you can upload data to server. The download Internet speed for this server is about 500 Mbps download and the upload speed is 150 Mbps with a wide bandwidth.

The following software is installed and running on the server:

- R 3.3.3 → One of the tools uses R programming language.

- Perl 5.16.3 → My Predecessors used Perl to code various aspects of the AMPed database.

- PHP 5.6.27 → The Graphical User Interface (GUI) is developed with PHP.

- MySQL 5.5 → MySQL is a freely available open source Relational Database Management System (RDBMS) that uses Structured Query Language (SQL).

- Apache 2.4.6 → Apache is free and open-source cross-platform web server software.

- Python 2.7.5 → Python is a scripting language like Perl; PHP used for web programming and also for desktop applications.

GUI front

Any further addition to the current GUI should follow the tone and structure of its current environment. Every new page that is added should follow the same color scheme and use the same icons to be uniform with the chosen AMPed schema. When the user accesses the database it should not look like pages are randomly added with a different tone. This would make it look like whole setup is out of sync. Generally, these characteristics are not very user friendly.
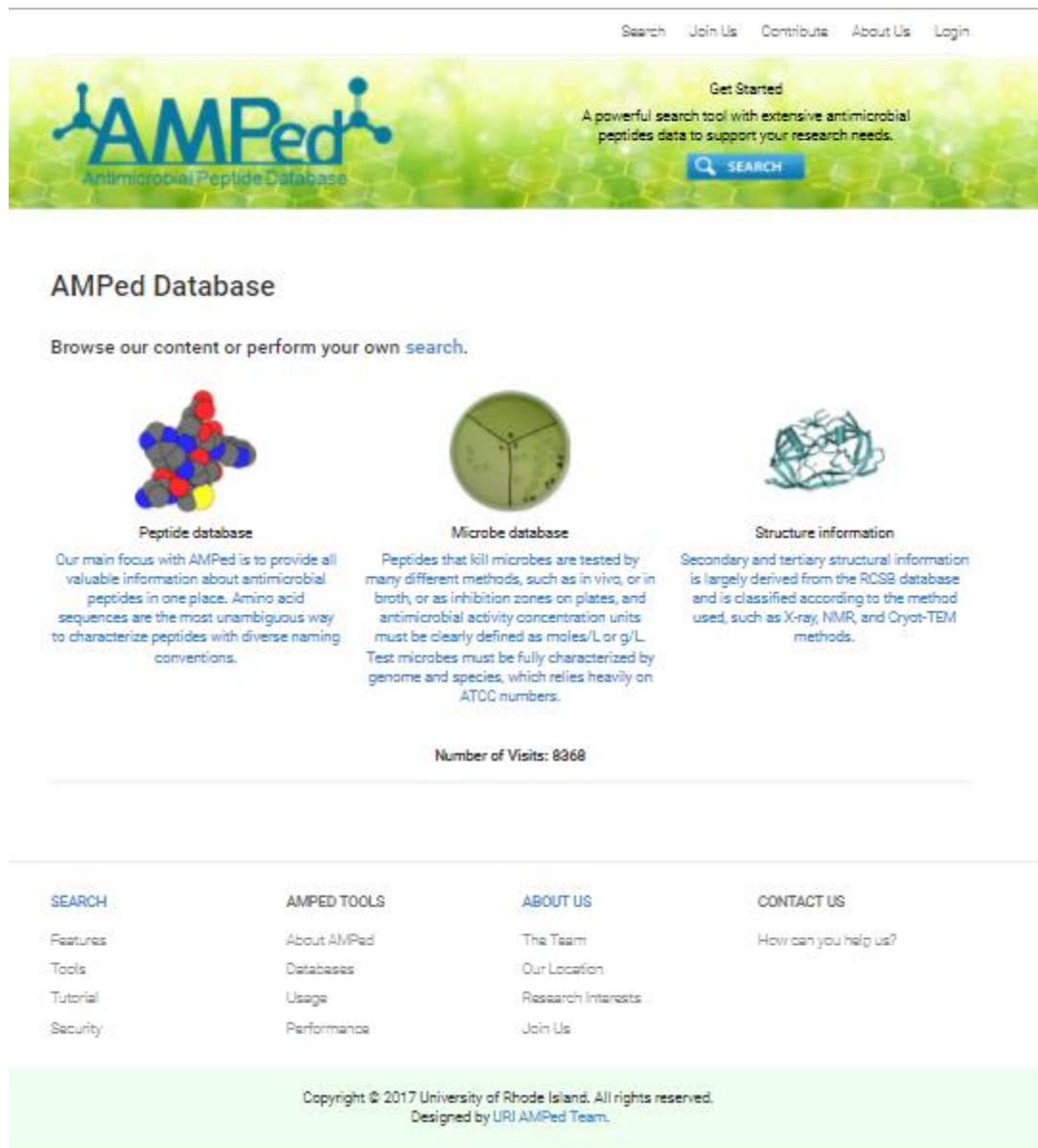
Figure 1: Homepage of AMPed

AMPed Web Flow

A web flow or web flowchart is a diagram showing a sequence of actions or movements performed by a user within the website. AWI version 2.0 as mentioned earlier is by Abraham Hererra [3]. Figure: 1 shows the set of options from the homepage: a search option with results, the About Us page, Join Us, Contribute, Contact Us, Logout and Login options with a validation process. Similarly, in AWI

version 3.0 the flow chart shows the Tools options, created for this project that can be accessed by the user, if user has a valid account. If a user does not have an account, it redirects them to the login page where they can create account.
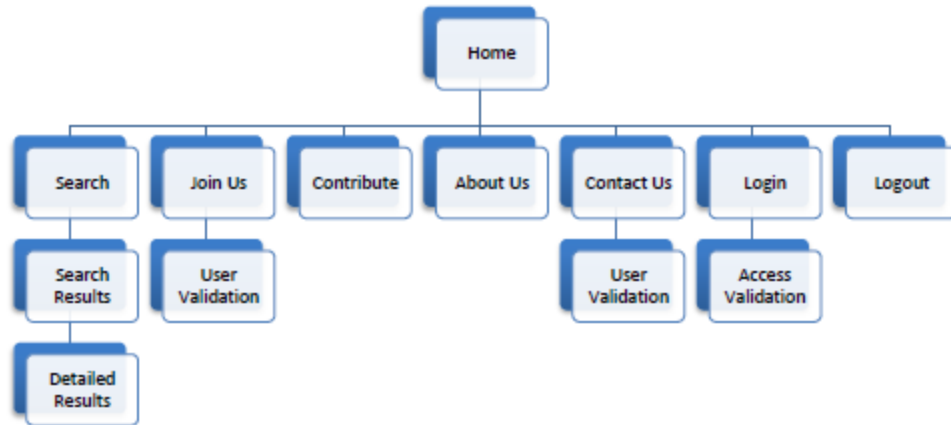
Figure 2: AWI version 2.0
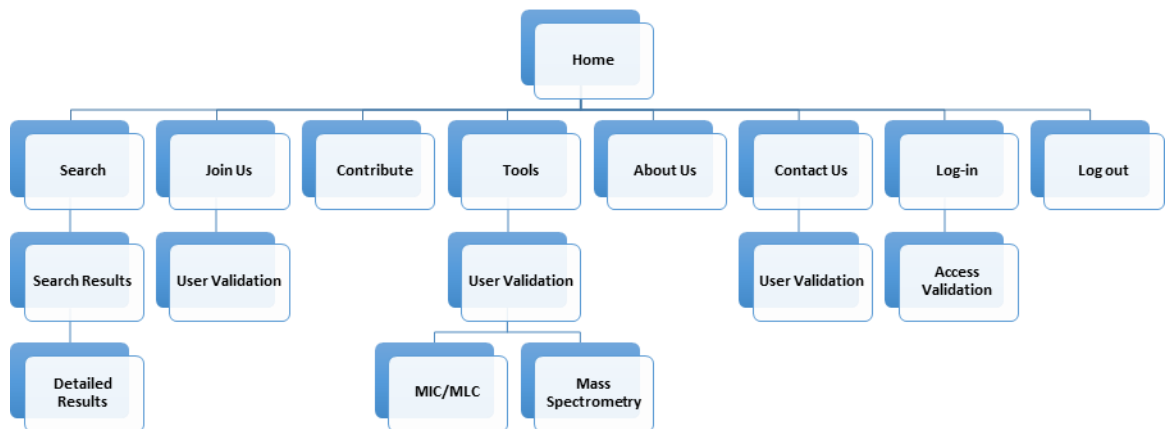
Figure 3: AWI version 3.0

AWI version 3.0 shows the all the options that are available for the user the existing web interface is merged with the new added interface which contains the MIC/MLC calculator and mass spectrometry tools (details in Chapter 3). Some options are easily accessed by the user without creating an account, but for some features such as tools or the contribute section, creating an account is necessary. This feature is for safety and security and controls who contributes data to the database, and who uses tools so that there will be no contamination or integrity problems with the data.

Portability

The AMPed graphical user interface has a mobile portability design [3], which also integrates new content (tools) developed for this project. This means that the user can upload a properly formatted data file, from their mobile device to the user interface where server side scripting is done. They can also download the result to the mobile device as well.

Privacy Policy

A privacy policy is an important legal document on any website that collects the user's personal information. It provides information about how a website visitor's personal information will be used.

Privacy policies are different according to the nature of how information is collected and used [5][6]. Our privacy policy can be found at the footer part of the website. The full privacy policy in appendix D.

Figure: 4 Privacy Policy footer page.

"In the US, the California Online Privacy Protection Act (CalOPPA) [7] dictates that if you collect any personal information from any California-based users, such as email addresses, GPS location, phone numbers, or mailing addresses, you are required to have a legal statement available for users to review that discloses the privacy practices of your business" [8]. We are adhering to the California standard.

Terms of Use

Terms of Use are set of rules that user has to abide to use the website. "Terms of service can also be merely a disclaimer, especially regarding the use of websites" [9]. Sometimes information on the website can be derived from different sources and the website owner cannot be held liable for that information. User should not assume that information is always correct. To make user understand these things, terms of use is important. Nowadays, upon creating accounts or to access some functions and resources in websites "Agree" or "Do not Agree" signs comes up. This kind of feature is not implemented in our case but we do have complete "Terms of Use" in place if users wants to review it.

Generally, "Terms of use" can be found in the footer section of the AMPed website. Our terms of use are given in Figure 5.

11

Figure: 5 Terms of Use

Summary

Goals set for this chapter were met as we understand the current underlying environment of AMPed. The new features were integrated to the existing system: The Tools page which holds the new calculation tool, privacy policy and terms of use in GUI.

REFERENCES

[1] Konstantinidis, George, "*Antimicrobial Peptide Editable Database*" (2015).*Open
Access Master's Theses. University of Rhode Island, Paper 694. Available:
http://digitalcommons.uri.edu/theses/694

[2] Garg, Tripti, "*Graphical User Interface for Antimicrobial Peptide Database*"
(2016). *Open Access Master's Theses.* University of Rhode Island, Paper 831.
Available:
http://digitalcommons.uri.edu/theses/831

[3] Herrera, Abraham, "*Web Based Bioinformatics Tool for Contributing Content to
the Anti-Microbial Peptide Editable Database*" (2016). *Open Access Master's
Theses.* University of Rhode Island, Paper 1042. Available:
http://digitalcommons.uri.edu/theses/1042

[4] Comodo, what is HTTPS? URL: https://www.instantssl.com/ssl-certificate-
products/https.html , Last accessed 10/12/2017

[5] Privacytrust. Importance of Privacy Policy.
URL: https://www.privacytrust.com/guidance/privacy_policy.html , Last accessed
10/23/2017

[6] katalystsolutions. The Importance of Privacy Policies. URL:
https://www.katalystsolutions.com/newsletter/57-the-importance-of-privacy-policies.html
, Last accessed 10/23/2017

[7] California Law Online Privacy. URL: https://oag.ca.gov/privacy/privacy-laws
, Last accessed 10/23/2017

[8] Reasons you need a privacy policy.

13

URL: https://termsfeed.com/blog/top-4-reasons-you-need-privacy-policy/ , Last

accessed 10/23/2017

[9] Reasons you need a privacy policy. URL:

https://en.wikipedia.org/wiki/Terms_of_service , Last accessed 10/23/2017.

CHAPTER 3

MIC/MLC Calculator

Goal

One of the main goals of this thesis is to develop minimal inhibitory concentration (MIC) and minimum lethal concentration (MLC) calculators, so that biologist/researchers can upload their raw data according to the specified template format and get the required calculation with a plot to visualize their data. By uploading their data onto our platform to use our software tool, we also begin to add new data to the AMPed database from researchers around the world. One main goal of the broader AMPed effort is to facilitate collaboration between researchers worldwide.

Design Process

Sequence Diagram

The sequence diagram is used to show the interactions between software objects in sequential order as the interactions occur. The sequence diagram helps developers and related people/business staff to understand how the software functions or data flows. It also serves to communicate the requirements for system implementation. Figure 4, shows the General sequence diagram of user interaction with the AWI 3.0 version of AMPed.
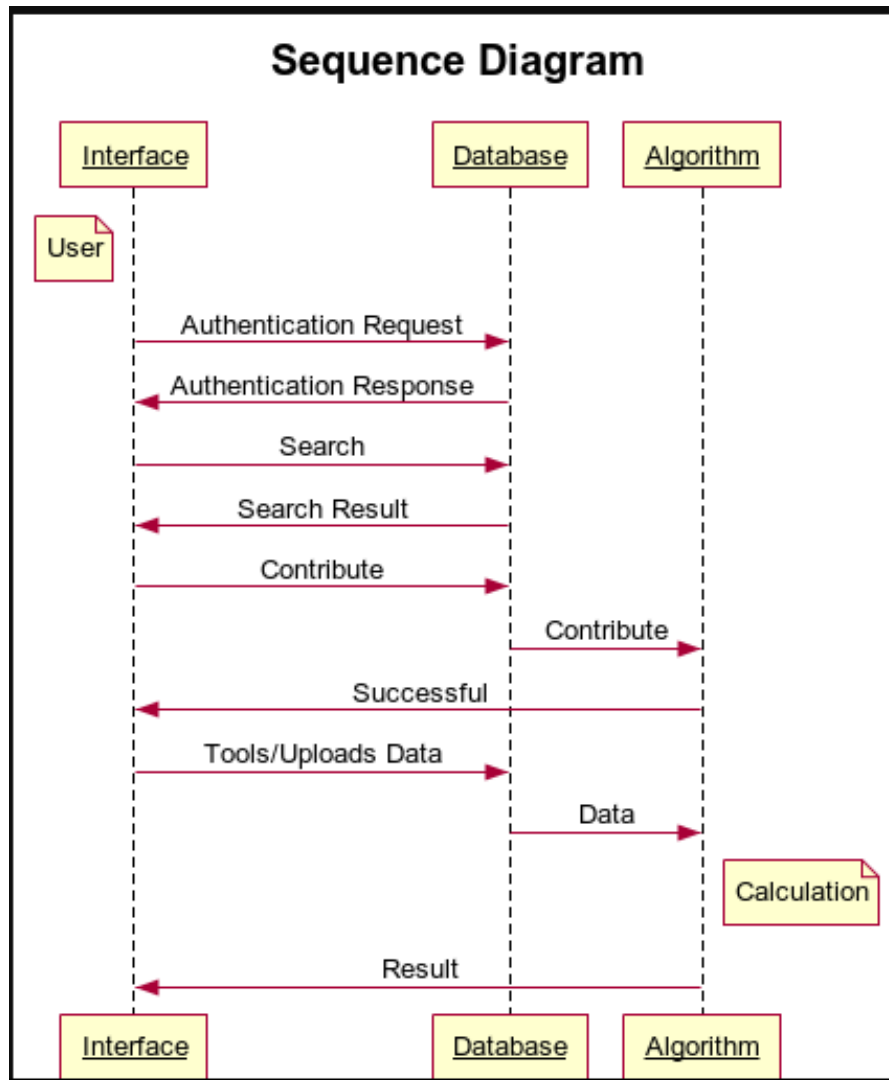
Figure 6: sequence diagram, the order of typical events is shown from top to bottom

In the above diagram arrow heads show the directions and sequences of functions and how users interact with the database and algorithms. In order to use the tools, the user must be logged in so an "Authentication Request" is sent to the database where it checks to see if that the account is set up or not, then it sends back an "Authentication Response". User can request "Search" to locate information in the database as well; the "Search Result" shows the appropriate result response. Another function is "Contribute". Authorized users can contribute experimental data from their research

groups to the database. The user data is parsed using an algorithm to determine if the data is acceptable and in the right format. It then generates an appropriate response. Similarly, when using tools, the user uploads data that goes to the server and, using the chosen algorithm, the tool generates calculation result and shows the plots in the user interface and also allows the user to choose to save a copy of their result.

Data

Bacterial growth data that is to be used to calculate the MIC/MLC values is the output of bacterial plate scanners that are used in the lab. An example of scanner software is softmax pro [1]. In Dr. Martin's Lab [2] a 96-well plate is setup for the experiment. The wells are setup according to a specific pattern. Plates are set up with test peptide dilutions in an ascending or descending order. To standardize data points, we have set up our plate templates to contain peptide dilutions in descending order. The overall peptide concentration decreases from left to right. Figure 5, shows how the plates need to be arranged to seamlessly upload growth data to the AMPed database.

	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
Pep1	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
Pep2	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
antibiotic	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M
NaOH	Blank	Dil1	Dil2	Dil3	Dil4	Dil5	Dil6	Dil7	Dil8	Dil9	Dil10	M

Figure 7: 96-well plate Data format

17

In above figure Pep1=test Peptide1, Pep2=test Peptide2, Dil#= Dilution, M= Microbe only, Blank = broth only, antibiotic =dilution series of an antibiotic that has a well-established MIC against the test microbe. NaOH= a 5M solution of sodium hydroxide in water is capable of killing all microbes.

For example, for the standard microdilution broth assay [3] dilution, the peptide in the first three rows would be 128,64,32,16,8,4,2,1,0.5,0.25 meaning, according to schema shown in Figure: 6 Dil1 is a dilution of 128 mg/L, Dil2 is a dilution of 64 mg/L and so on upto Dil 10 ranges. Dilutions can be changed according to the needs of the experiment. It depends upon what is necessary for the investigation and how potent the researchers are expecting the peptide to be. The overall pattern of serial dilutions should be maintained, however.

The MIC and MLC calculators determine the minimum concentration of a given AMP that will either be inhibitory or lethal against some specified species of bacteria. Peptides are tested in triplicate, so the first 3 rows will all have the same peptide concentrations as Pep1 shown in Figure 5. As Pep1 rows 4-6 will contain dilutions of a second test peptide, let's say Pep2, set up analogous to the first 3 rows with Pep1. Tests are repeated 3 times for each peptide (here we are supposing pep1 and pep2) to see what kind of changes the dilution brings to each peptide, and the two peptides are compared with the standard antibiotic in row 7.

In the above figure only one block is shown There can be many blocks like this depending upon the experiment. Each block represents one single time point in the growth curve of the bacteria. The antibiotic Row 7 will have a known antibiotic at the same dilutions as the test peptides.

Last row 8 will have NaOH as a Positive lethal control. Column 1 or Blank is broth only, no bacteria (contamination control). From Column 2-11 Dil1-Dil10 will be varying dilutions of test peptides. Column 12 or M will be bacteria and broth only as the negative control (no peptide/antibiotic added).

Data for upload

The file that we can get from Softmax Pro, is saved in .txt format but the non-binary format does not have list of the time points in each block. We require time points to calculate MIC/MLC for a minimum of 12 hours growth and also to monitor the entire growth curve for each bacterial culture. So, the output file will have the dilutions and all settings as described above from the lab experiment. That data is exported in Microsoft Excel in .csv format, and then are added time points for each block. Suppose we have 50 blocks, there will be 50 time points in the data, one for each block. Figure 6, below shows a data template format in .csv format which now can be uploaded to the web to get the required calculations and visualizations.

	Blank	X128	X64	X32	X16	X8	X4	X2	X1	X0.5	X0.25	Microbe	Timepoint
A													
B													
C													
D													
E													
F													
G													
H													
	Blank	X128	X64	X32	-	-	-	-	-	-	-	-	-
A													
-													

one time point

Figure: 8 upload data .csv format

Each row and column has a header like shown in figure 6 columns are labeled Blank, X128 and so on (one can have any name in the header) with time point as the last column on the right. After each block (after row H) there is one blank row and other block follows which have same number of rows and columns. "Timepoint" the cell underneath that label will have one value, in hour time format for each block the minimum time point (0 hours) and the maximum time point (preferably 12-24 hours). While generating the results, the algorithm uses the same "Timepoint" for each whole block. It is important for uploaded data to follow this format for a proper output of an accurate MIC/MLC and growth curve plot. We realize that this format may differ from the format that the data was originally collected, however, reorganizing the data prior to uploading it is essential for a successful result. For a sample of a whole dataset see appendix A.

Algorithm Steps

General Outline of algorithm for MIC/MLC(will be explain with code in details after this)

Step 1: Any user who wants to use our tools must have an account with us. Tools reside in the footer area of the AWI, if the user is not logged in, they are redirected to the log in page or, if needed, to create a new account.

Step 2: A few questions are asked related to the data to be uploaded to gather information about the test peptides used in the experiments like what bacteria the test peptide fights against, the assay used and so on.

Step 3: A sample template is provided, if needed, with more information. The file should be in the proper format to get the right output.

Step 4: Select and upload the data file.

Step 5: The PHP script runs in the background creating a unique name for every file uploaded and saves the file in the server. The system executes an R-script to generate the result.

Step 6: The algorithm calculates the average of the first 3 rows (pep1) for every column and puts the results in a new table(pep1) for that time point.

The same is done for rows 4-6 (pep2) of every column. The result into a new table(pep2) for that time point.

Step 7: Graphs are plotted for Pep1 and Pep2 and a result table containing the calculations is published in the server.

Step 8: Results are also displayed on the web interface; these can be downloaded to the user's device.

The Average calculations of the pep1 and pep2 are displayed and by looking at those scientist can easily find the MIC/MLC of those peptides.

AMPed Web Interface(AWI) for MIC/MLC

Tools

Tools can be found in the footer of the web interface as shown in Figure 7. The footer was updated as a part of AWI version 3.0. The "Tools" page was created to hold the tools created for this project. Figure 8 shows the footer part of the web interface where

along with other footer title "AMPed Tools" is also added. As mentioned above this part holds the tools created for this project.
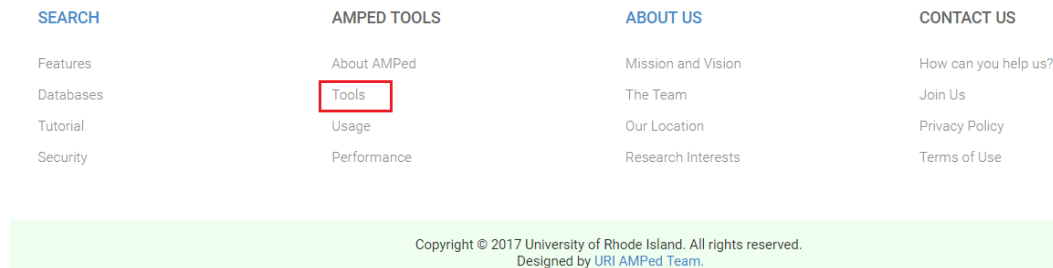
Figure: 9 Footer with Tools holding tools created.

PHP side coding in footer part "AMPED TOOLS"

```html
<div class="col-sm-3">
  <div class="single-widget">
    <h2>AMPed Tools</h2>
    <ul class="nav nav-pills nav-stacked">
      <li><a href="/aboutamped.php">About AMPed</a></li>
      <li><a href="/2toolspage.php">Tools</a></li>
      <li><a href="#">Usage</a></li>
      <li><a href="#">Performance</a></li>
    </ul>
  </div>
</div>
```

Figure: 10 Coding in footer

This snippet of code checks if the user is logged or not, if the user is logged in, it takes user to the page where they can choose tools to use (Figure 10), if not, it redirects to the login page (Figure 9).

```php
<?php session_start();

    if ($_SESSION['is_auth'] != true) {

    $_SESSION['redirect'] = "2toolspage.php";

    header ("location: ../login.php"); exit();

    }?>
```

22

Figure: 11   log in page

After login, the user can choose which one to use (figure 10).

Figure: 12

MIC/MLC Calculator Page

In order to gather the information related to the data a few questions (see Figure 11) are asked via GUI which helps the system to display the results according to the need of the researcher. Questions included are:

- What bacteria does your peptide fight against?

- What kind of Assay you are using?

- What is your medium?

- What would you like to name Test Peptide 1?

- What would you like to name Test Peptide 2?

23

- What control antibiotic did you use?

- What concentration of the control antibiotic did you use?

- How many time points does your data have?

Figure: 13 MIC/MLC Calculator page

Also, if the user needs more information to understand the questions a "title" attribute is added to the text field. When the cursor hovers around the text field, more information is given, as shown in Figure 11. For example," *What species/strain of bacteria are you trying to inhibit? E.g., S. aureus, E. coli, etc.*" pops up when the cursor hovers over the text box associated with "*What bacteria does your peptide fight against?*"

24

Upload file

As shown in the above figure, "Choose File" can be displayed using attributes (specific words written in code corresponds to what we want to see in the user interface) in PHP, described as below;

<div align="center">type="file"</div>

inside the input tag and another attribute

<div align="center">accept= " .csv"</div>

allow you to choose only .csv file format.

Snippet of the code

<input style="height: 35px;" name="upload_file" type="file" accept= " .csv" id="upload_file" value=""/>

When writing client-side code, attribute

<div align="center">enctype="multipart/form-data" [4]</div>

should be used if the form includes any <input type="file"> **elements** as used in our code. The "enctype" attribute of the form element specifies how the form-data should be encoded when submitting it to server. The enctype attribute can be used only if

<div align="center">method="post"[5]</div>

POST is used to collect form data from client browser and send it to the web server. It has advanced functionality which supports multipart binary input - this allows the client (browser) to send multiple chunks of binary input (files) to the server.

Snippet of code

```
<form action="mic_upload_process.php" method="post" enctype="multipart/form-
data" name="form1" id="form1" onSubmit="return valForm(this);">
```

To create submit button we can use "type" attribute inside input tag

```
"type="Submit"
```

Creating dynamic .r file through PHP

Files uploaded by users and saved on the server should have unique names. A function is created in PHP with global variable $Filename which has a time() function [6]. The time() function returns the current time in the number of seconds since the Unix Epoch (January 1 1970 00:00:00 GMT). So every time a file is uploaded it is saved with a unique name in the server. To execute that unique file using an R script, there should be a passing of variables between PHP and R, so that R can read the file saved by PHP [7][8]. For every data file that is uploaded, an R-script is created, which results in the R script having the same time () name as the data file but of .r file type. The plots and result file also have same time () name but are of .png type.

In summary for every instance of data file uploaded, scripts, image and .csv files are created with a unique name that uniquely corresponds to that instance.

Suppose for one instance,

1509732603-data.csv (file uploaded by user saved with unique name)

1509732603-mic.r (r script created for that particular file)

26

1509732603-pep1.png (pep1 plot with unique name generated by r script)

1509732603-pep2.png (pep2 plot with unique name generated by r script)

1509732603-result.csv (pep2 plot with unique name generated by r script)

```php
<?php
//function to open .r from php and write rscript everytime a file is uploaded with dyanmic/unique name.
function writeFile($idcode) {
global $fileName;
$text1 = "#m_m\nMyData <- read.csv(\"mic-data/".$fileName. "\", header=TRUE, sep=\",\", colClasses = \"character\")\n";
$constantTXT = <<<END
                colnames(MyData) <- c("FirstCol",colnames(MyData)[2:14])
                Mycl_data <- subset(MyData, FirstCol == "A" | FirstCol == "B" | FirstCol == "C" |
                FirstCol == "D" | FirstCol == "E" | FirstCol == "F")

                ## Fill in time point will valid time
                for(i in 1:nrow(Mycl_data))
                {
                    ifelse(Mycl_data\$Time.point[i]=="", Mycl_data\$Time.point[i] <- val_time,
                    val_time <- Mycl_data\$Time.point[i])
                }

                cl_names <- colnames(Mycl_data[2:ncol(Mycl_data)])## Extracting column names

                Mycl_data <- Mycl_data[,c(-1)] ## Get rid of the first irrelevant column

                Pep1 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)
                Pep2 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)
                temp = 3
                count = 1
                for (i in 1:nrow(Pep1))
                {
                    for (j in 1:ncol(Pep1))
                    {
                        Pep1[i,j] = round(mean(as.numeric(Mycl_data[c(count:temp),j])),digits = 5)
                        Pep2[i,j] = round(mean(as.numeric(Mycl_data[c((count+3):(temp+3)),j])),digits = 5)
                    }

                    count = temp + 4
                    temp = temp + 6
                }
```

Figure: 14 PHP script (showing how R script is running inside PHP)

```php
                colnames(Pep1) <- cl_names
                colnames(Pep2) <- cl_names

                rownames(Pep1) <- c(1:nrow(Pep1))
                rownames(Pep2) <- c(1:nrow(Pep2))

                Pep1 <- as.data.frame(Pep1)
                Pep2 <- as.data.frame(Pep2)
                print(Pep1)
                print(Pep2)

END;
$text2 = "png(file = \"mic-output/".$fileName. "-pep1.png\")\n
matplot(Pep1\$Time.point, Pep1[,c(1:12)],type = c(\"b\"),lty = 1,lwd =2,  xlab = \"Time Points\",ylab = \"OD\",main = \"MIC_MLC\",pch=c(15,16,17),col = 1:12)\n
legend(\"topleft\", legend = colnames(Pep1), col=1:12, box.lwd =par(\"lwd\") ,pch=c(15,16,17)) # optional legend\n
dev.off()\n";
$text3 = "png(file = \"mic-output/".$fileName. "-pep2.png\")\n
matplot(Pep2\$Time.point, Pep2[,c(1:12)],type = c(\"b\"),lty = 1,lwd =2,  xlab = \"Time Points\",ylab = \"OD\",main = \"MIC_MLC\",pch=c(15,16,17),col = 1:12)\n
legend(\"topleft\", legend = colnames(Pep2), col=1:12, box.lwd =par(\"lwd\") ,pch=c(15,16,17)) # optional legend\n
dev.off()\n";
$text4 = "TotalPep <- rbind(Pep1,Pep2)\n
write.csv(TotalPep,\"mic-output/".$fileName. "-result.csv\")";

$myfile= fopen(getcwd()."/R-scripts/" . $idcode . "-mic.R", 'w') or die("Could not create the file.");
fwrite($myfile, $text1);
fwrite($myfile, $constantTXT);
fwrite($myfile, $text2);
fwrite($myfile, $text3);
fwrite($myfile, $text4);
fclose($myfile);
}
```

Figure: 15 PHP script (showing how R script is running inside PHP)

As shown in figure 12 and 13, (1,3,4 ,5) has dynamic text (or generated r script) in variable whereas constant text (r script) is in different variable.

Similarly, Figure 13 (6) shows creating .r file and writes the dynamic as well as constant text/r script in the file. Full PHP coding in appendix B

Snippet of code

This code shows how a user uploaded file is saved with unique time stamp.

$idcode= time();   #variable is time() which is passed through out code.

    $fileName = $idcode. "-" . "data.csv"; # $idcode is used to create a unique name for this instance of data file uploaded by the user with .csv extension..

    $finalLoc = getcwd()."/mic-data/".$fileName; # the data file is saved on the server with folder name "mic-data".

    $URLloc = "/mic-data/".$fileName; #location for the file to be saved

$result =move_uploaded_file($_FILES["upload_file"]["tmp_name"],$finalLoc);

Now, using this function it creates the R script, and exec command [9] executes the R script (Full R script in appendix C).

writeFile($idcode);

    $Rcommand = "Rscript R-scripts/" . $idcode . "-mic.R";

    exec($Rcommand);

Snippet of code for calculation

This code shows how the different rows are selected from the excel file and the steps

that taken to separate different rows and put them in different variable and how after

calculation saved as data frames. Below is the part of R script for calculation.

```
Pep1 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)
Pep2 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)
       temp = 3
       count = 1
       for (i in 1:nrow(Pep1))
       {
               for (j in 1:ncol(Pep1))
               {
               Pep1[i,j] = round(mean(as.numeric(Mycl_data[c(count:temp),j])),digits
= 5)
       Pep2[i,j]=
round(mean(as.numeric(Mycl_data[c((count+3):(temp+3)),j])),digits = 5)
               }

               count = temp + 4
               temp = temp + 6
       }

       colnames(Pep1) <- cl_names
       colnames(Pep2) <- cl_names

       rownames(Pep1) <- c(1:nrow(Pep1))
       rownames(Pep2) <- c(1:nrow(Pep2))

       Pep1 <- as.data.frame(Pep1)
       Pep2 <- as.data.frame(Pep2)
```

Result

The result contains the average bacterial growth in the presence of peptide 1 (as the mean of first 3 row) of every column in the block sorted according with different dilutions.

The two plots generated by the MIC/MLC calculator show the average of peptide1 vs. time point and average of peptide2 vs. time point respectively.

	Blank	X128	X64	X32	X16	X8	X4	X2	X1	X0.5	X0.25	Microbe	Time.point
1	0.043	0.04267	0.04633	0.044	0.045	0.045	0.04233	0.045	0.04333	0.04333	0.04333	0.05367	0
2	0.042	0.046	0.04433	0.04533	0.045	0.05067	0.08633	0.08367	0.07	0.06967	0.07833	0.07933	0.5
3	0.045	0.046	0.04633	0.04533	0.045	0.05267	0.09167	0.09133	0.08567	0.07167	0.08667	0.088	1
4	0.043	0.046	0.04633	0.04533	0.045	0.05667	0.10633	0.09933	0.07567	0.07567	0.09567	0.09767	1.5
5	0.042	0.046	0.04633	0.04533	0.045	0.06167	0.12667	0.12	0.10667	0.08133	0.11667	0.12	2
6	0.04267	0.046	0.04633	0.04533	0.045	0.06433	0.151	0.14833	0.14533	0.08833	0.145	0.147	2.5
7	0.04167	0.046	0.04633	0.04533	0.045	0.06833	0.157	0.179	0.10133	0.092	0.17467	0.17767	3
8	0.038	0.046	0.04633	0.04533	0.045	0.73333	0.98667	1.135	1.14333	1.13933	1.147	1.165	24
11	0.04767	0.04533	0.04367	0.045	0.04433	0.04333	0.044	0.044	0.045	0.044	0.05667	0.06233	0
21	0.04167	0.04533	0.04367	0.044	0.04533	0.053	0.09033	0.08367	0.07	0.07113	0.07967	0.07967	0.5
31	0.04167	0.04533	0.04367	0.047	0.04533	0.05633	0.093	0.087	0.07067	0.07233	0.084	0.085	1
41	0.04333	0.04533	0.04367	0.047	0.04533	0.05967	0.10367	0.098	0.09933	0.08	0.09633	0.09767	1.5
51	0.042	0.04533	0.04367	0.047	0.04533	0.06067	0.11667	0.11733	0.09467	0.087	0.11567	0.119	2
61	0.04233	0.04533	0.04367	0.047	0.04533	0.063	0.14787	0.145	0.08833	0.09433	0.143	0.14567	2.5
71	0.03767	0.04533	0.04367	0.047	0.04533	0.07633	0.167	0.17167	0.092	0.10133	0.16867	0.17167	3
81	0.037	0.04533	0.04367	0.047	0.04533	0.72733	0.99267	1.07933	1.09367	1.114	1.12467	1.16333	24

Download excel file: 1510183410-data.csv.

Figure: 16 Calculated Result

MIC_MLC Peptide1

MIC_MLC eptide2
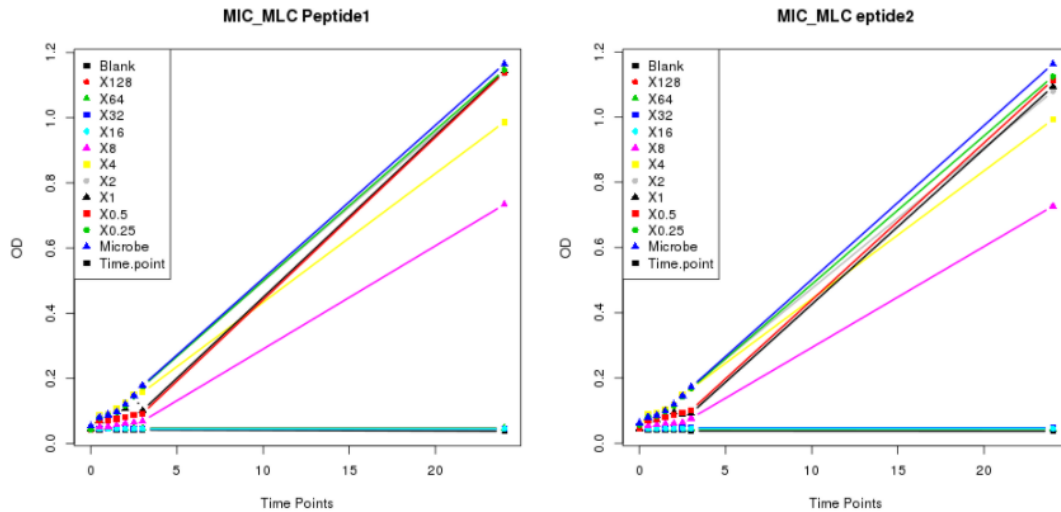
Download-PEP1
Download-PEP2
Bacteria Name :E.coli   Peptide2 Name : Beta-defensin 3

Assay Name :X ray   Antibiotic Name :gentamicin

Medium :LB   Antibiotic concentration :12

Peptide1 Name :Beta-defensin 1   Timepoint :5

Figure: 17 Plot Results

In the above plots of bacterial growth treatment with pep1 and pep2, each average of 3 measurements in a given area of the plate is plotted against the time point of that block according to peptide concentration added to that area of the plate. If we study the plots with the calculations carefully we can find that the minimum inhibitory concentration is at a X16 dilution of both peptides tested on a single 96-well plate. Plots are simply there to make data come alive. We can visualize the bacterial growth data i.e. pep1 and pep2 shown as above and this makes it easier to understand the data and conclude the MIC/MLC results.

Summary

Goals set were met as a tool was created to upload raw lab data in a specified .csv format to the web interface and get the required calculations and plot visualization which could be downloaded for the reference.

31

REFERENCES

[1] Softmax pro. URL: https://www.moleculardevices.com/systems/microplate-readers/softmax-pro-7-software , Last accesses 3/6/2018

[2] Dr. Lenore Martin. URL:http://web.uri.edu/cmb/lenore-martin/ , Last accessed 10/20/2017.

[3] J. M. Andrews "*Determination of minimum inhibitory concentrations*."

J Antimicrob Chemother. 2001 Jul; 48 (Suppl 1): 5–16. Avaliable:

https://www.ncbi.nlm.nih.gov/pubmed/11420333

[4] Form content Type. URL:

http://www.w3.org/TR/html401/interact/forms.html%23h-17.13.4#h-17.13.4 , Last accessed 10/20/2017

[5] Post Method. URL: https://www.tutorialspoint.com/php/php_get_post.htm , Last accessed 10/20/2017

[6] Time function. URL: https://www.tutorialspoint.com/php/php_date_and_time.htm, Last accessed 10/20/2017

[7] Pass Arguments and Execute R script from PHP Forms. URL:

https://www.codeproject.com/Articles/1119237/Pass-Arguments-and-Execute-R-script-from-PHP-Forms , Last accessed 10/20/2017

[8] Integrate R with PHP. URL: http://www.listendata.com/2017/04/how-to-integrate-r-with-php.html , Last accessed 10/20/2017

[9] exc. URL: http://php.net/manual/en/function.exec.php , Last accessed 10/20/20

CHAPTER 4

MASS SPECTROMETRY PREDICTION TOOL

Goal

Here we lay out the algorithm needed to create a mass spectrometry (MS) prediction tool, where biologists can upload their predicted peptide sequence along with its measured mass spectrum. In return they will get back the predicted isotope distribution with the possible masses of the desired peptide sequence. By comparing the calculated spectrum with the measured spectrum, users can evaluate the integrity of their peptide.

Introduction

Mass spectrometry is an analytical technique used to characterize the structure of a molecule, by using gas-phase momentum to directly measure the mass/charge ratios of individual molecules. MS can be used to determine or confirm the amino acid sequence of an unknown or known AMP.

MS is the most commonly used technique to check that the amino acid composition of a synthetic peptide is correct or not [1]. This test records the mass-to-charge ratio of each peptide and outputs a plot of the relative abundances of each isotopic mass peak with high resolution. Because some of the atoms in each peptide molecule moving through the spectrometer have a certain probability of containing a less-abundant heavier mass isotope therefore they will be viewed at one or two atomic mass units heavier than the lowest mass or monoisotopic mass of that peptide. MS output gives a

distribution of mass peaks at different intensity depending upon their compositions different isotopes in each peptide [2].

This information is crucial in the discovery of novel antimicrobial drugs. Also, when performing the synthesis of AMPs for the purpose of research and drug characterization, mass spectrometry can be used to validate that a synthesized peptide has its asserted amino acid composition, as a gold standard characterization technique.

Data

Input for this tool will be a peptide sequence consisting of as a strong amino acid of known atomic composition. The program gives the isotope distribution with all the possible masses from the input peptide sequence.

Example: Peptide with atomic composition $C_{47}H_{83}N_{15}O_{16}S$

user will enter the peptide sequence in the form CvHwNxOySz (usual elements that compose peptides), as well as a parameter k, for the number of peaks to be shown and then the program will output an k-by-2 matrix with the possible molecular masses in the left column and their corresponding abundances in the right column.

Programming Language

The Perl programming Language will be used, integrated with PHP. R will be used for graphical functions. The weight of each amino acid will be defined in the program i.e. looked up in a table, and the then weight of the peptide chain can be calculated.

Algorithm

- User will input values in numbers for v,w,x,y,z,k to specify the peptide and the number of peaks. After that it will calculate the mass values of the peaks assuming that all of the atoms are low-mass isotopes and storing that in the first entry of the *mass* vector.

- Next, it will add 1 to that and stores that number in the next mass vector entry, for the case where there is exactly one heavier isotopic atom in the molecule.

- This process proceeds until *mass* contains k entries.

- Now, the probabilities of the masses of each peptide isotope needs to be stored so the vector p, is initialized with zeros in each entry. The probability of having the lowest mass isotope of a peptide randomly is as follows:

$$\mathbf{p^{12}C^v \times p^1H^w \times p^{14}N^x \times p^{16}O^y \times p^{32}S^z}\,[1].$$

- Here, $p^{12}C$ is the probability of a random carbon atom being the minimum mass isotope, ^{12}C, and the other probabilities are defined similarly.

- Then algorithm will store this value into the first entry of p.

- Now, we will consider the case where there is exactly one heavier isotopic element of each element, meaning, each ^{12}C in the initial probability with one ^{13}C, then two ^{13}C, up to v ^{13}C, which corresponds to the case where all carbons in the molecule are heavier isotopes and all of the other atoms are low-mass isotopes. Then it will store each of the values in the 2nd, 3rd, and $v+1^{th}$ entries of the *prb* vector.

- Same process will repeat to other elements i.e. ^2H, ^{15}N, ^{17}O, ^{18}O, ^{33}S, ^{34}S.

- After that we will consider the cases where two elements, such as C and H, are isotopic. It will give us probability of the peptide with only one instance of ^{13}C and one ^{2}H, followed by one ^{13}C and two ^{2}H, up to one ^{13}C and w ^{2}H.

- Again same process will be used for instances of ^{2}H, but with two ^{13}C. Will continued until there are v ^{13}C and w ^{2}H.

- Algorithm will go on for three instances of heavier isotope elements, then four, and finally the cases where all five elements are heavier isotopes, adding each probability to the entry in the *prb* vector that corresponds to the mass associated with that isotope.

- Now, each of the probabilities in the *prb* vector will be divided by the largest entry in the *prb* vector and stored in a new vector as intensities, which is a list of the abundances for each mass relative to the highest mass (set to be 100%).

- Lastly, the *mass* vector and the vector with intensities are combined into a matrix with the entries of mass in the left column and the corresponding entries of the intensities in the right column.

This table will be the output of the program and resembles an MS data set.

There might be some limitations to the algorithm that needs to be addressed as being developed as a bioinformatic tool. The computational load increases exponentially with the size of the molecule. Peptides could be up to one hundred amino acids long and the tool should be able to handle any size of peptide; it is in fact a very huge project, currently algorithm can only handle peptides with fifteen long amino acids.

Output

Figure 18 shows how the predicted Spectrum would look like.

#of atom in peptide	Isotope	Atomic mass	%Abundance		Relative Mass	Intensity
47	12C	12	98.93		1145.586	100
	13C	13.00335	1.07		1146.589	58.73899
83	1H	1.00782	99.9885		1147.59	24.76994
	2H	2.0141	0.0115		1148.591	7.75869
15	14N	14.00307	99.632		1149.592	1.97077
	15N	15.0001	0.368		1150.594	0.42152
16	16O	15.99491	99.757		1151.595	0.07821
	17O	16.99913	0.038		1152.597	0.01285
	18O	17.99916	0.205		1153.599	0.00189
1	32S	31.97207	94.93		1154.601	0.00025
	33S	32.97145	0.76		1155.603	0.00003
Total 162						

Figure 18: Predicted Mass Spectrum

Figure:19 is the sample distribution plot of the peptide Intensity vs Mass-to-charge ratio (mz).
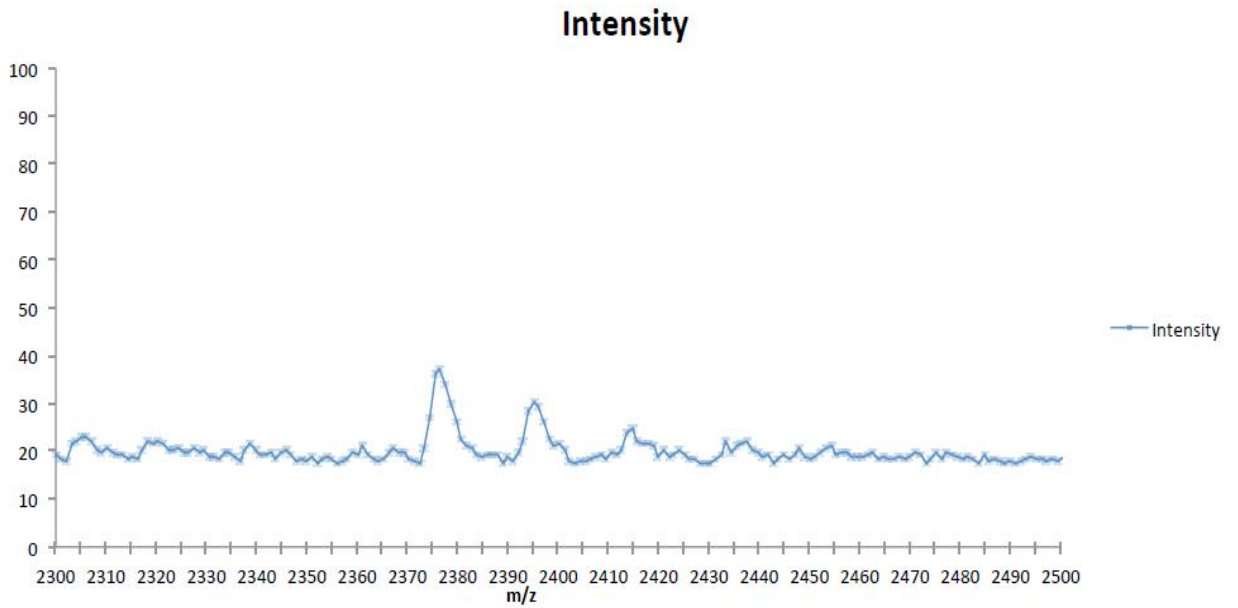
Figure 19: Measured Mass Spectrum of a peptide graph

37

REFERENCES

[1] Nefedov AV, Sadygov RG. *"A parallel method for enumerating amino acid compositions and masses of all theoretical peptides."* BMC Bioinformatics. 2011;12:432. doi:10.1186/1471-2105-12-432. Available:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3270061/

[2] Yamamoto H., McCloskey James A.*" Calculations of Isotopic Distribution in Molecules Extensively Labeled with Heavy Isotopes"* [Online] Available:

http://pubs.acs.org/doi/abs/10.1021/ac50010a025

 [3] Roman A. Zubarey, * Per Håkansson, and, and Bo Sundqvist

 "Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements." Analytical Chemistry **1996** *68* (22), 4060-4063

doi: 10.1021/ac9604651. Available: https://pubs.acs.org/doi/pdf/10.1021/ac9604651

CHAPTER 5

The AMPed DATABASE

A database is a collection of information, data that is organized in a manner so it can be easily accessed, managed and updated. AMPed uses a relational database, so data is organized into rows, columns and tables [1].

The Amped database is an editable database where values, attributes, can be updated as we add information and need to make changes. The database design process is very important to integrate all of the content and to assure that the data remains consistent and correct and can be accessed efficiently. The Entity Relationship (ER) diagram is a graphical representation of the data entities and their relationships to each other. The logical representation of entities is mapped tables, where the attributes become column names in the physical design of the database. The design of the AMPed database has gone under a number of changes by students who worked on this project before. They are George Konstantinidis Tripti Garg and Abraham Hererra. Part of the updated, a fragment of the current ER diagram of AMPed, is shown below as Figure 20. See Appendix D for the detailed diagram.
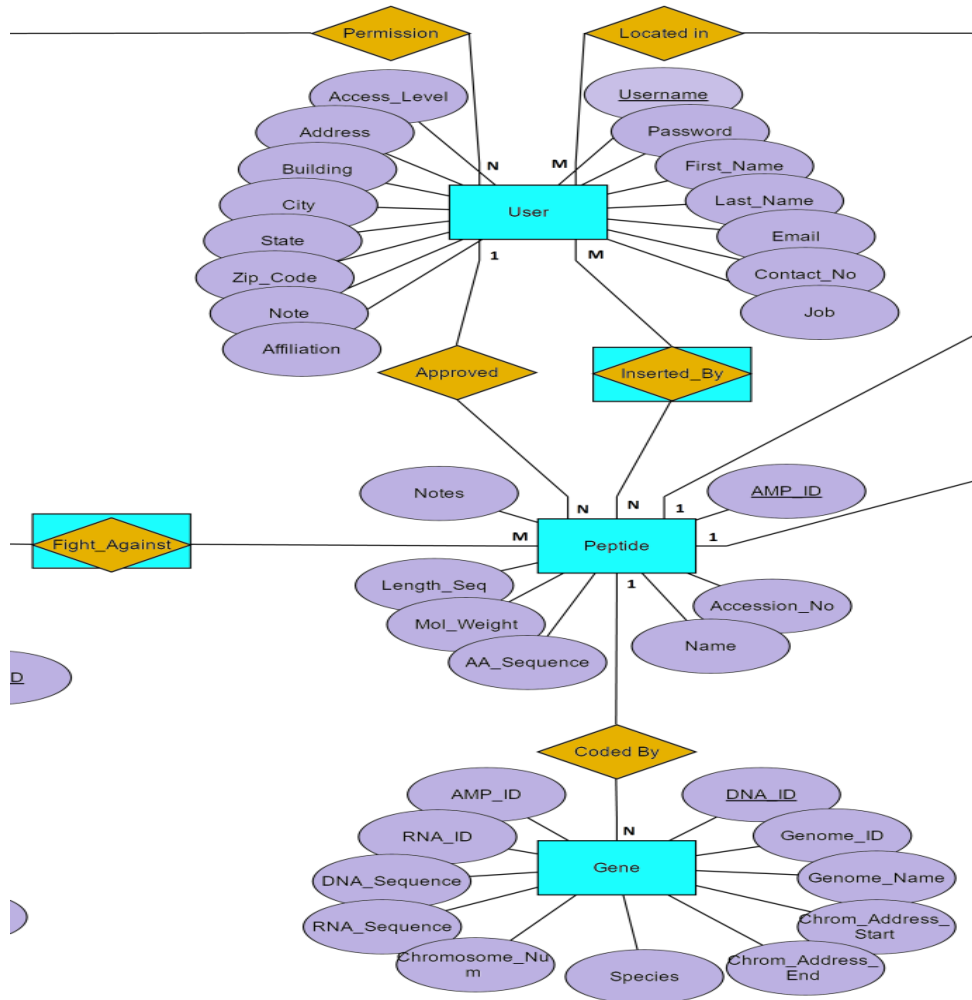
Figure 20: ER diagram of a part of the AMPed database.

An entity is an object or concept about how you want to store information. Database
entities are converted into tables that hold attributes in columns.
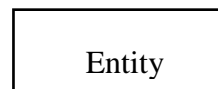
Entity

Relationships are represented by diamond shapes and show associations between or among entities.
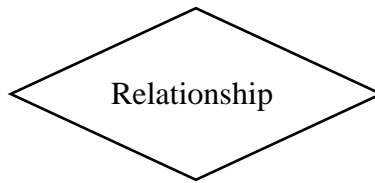
Associative entities are transformed into tables that hold the foreign keys (a means of reference among tables) associated with it or connected with it. It is represented in an ER diagram by diamond shape within the rectangle shape.
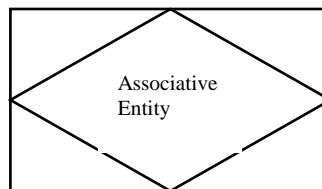
Attributes, represented by ovals, are characteristic/properties of the entity. A key attribute, represented with an underline, is a uniquely identifies and entity.
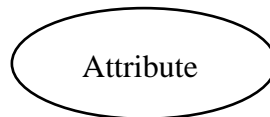
A multivalued attribute can have more than one value and is represented as shown below.
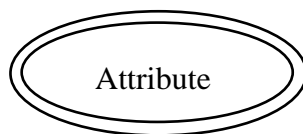
Solid lines connecting attributes and entities show the relationships between them i.e. which attributes are associated with which entities.

There are 3 types of relationships among entities in the ER diagram, one-to-one, one-to-many, and many-to-many relationships. For examples in the AMPed ER diagram, one of the relationships is between the Peptide entity and the Gene entity. This is a one to many relationship, as shown in Figure 20 represented by 1-N. This means that one peptide can be related to many genes, or that many genes can related to one peptide.

Normalization

Normalization is the process of organizing data in a database; which includes creating tables and relationships between those tables, setting rules that define those relationships to protect the data and to make the database more flexible, correct and consistent by eliminating redundancy and inconsistent data. One of the goals of this project is normalize the existing AMPed database. In other words, normalization is necessary to avoid data redundancy, insertion anomaly, update anomaly & deletion anomaly. Most commonly used normal forms are:

- First Normal Form (1NF)

  First normal form states that an attribute of a table cannot hold multiple values. It should hold only atomic(single) values. In AMPed database, Atomic_Coordinates table has an attribute Atomic_Names which was identified as a multi valued attribute and is correctly represented in the ER diagram. To normalize the multivalued attribute, new table, Atom, is created with primary key as Atom_Id which have two attributes Atom_Nmae and Seq_AA_No as the foreign key. All the tables in AMPed database complies with 1NF.

42

- Second Normal Form (2NF)

A table is in 2NF if both of the conditions hold:

- Table is in 1NF.

- No non-prime (not a part of a key) attribute is dependent on a proper subset of any candidate key of the table. In other words, it should not have a partial dependency, i.e. an attribute in a table should never be uniquely determined by only a part of the primary key. It must be fully dependent upon the key.

All attributes of AMPed complies with 2NF none of the attributes are partially dependent on the any of the primary keys.

- Third Form (3NF)

For a table to be in the third normal form;

- It should be in the Second Normal form, and

- It should not have a transitive dependency, i.e. a non-prime attribute that depends upon, or is uniquely determined by other non-prime attributes. All attributes should depend upon the prime attributes or primary key. All of the attributes in the AMPed database comply with t 3NF.

If attributes are added to the database in the future, the resulting new schema should be checked for compliance with these normal forms to reduce data redundancy and to achieve data integrity.

Tables

To accommodate the needs of the users, and to create a correct representation of the relationships, tables needed to be updated with additional attributes. For the detail information about table refer to the appendix E. The Microbe, Amino_Acid_Address, Amino_Acid_Address tables (See Figure 21 below) were modified as follows:

- Removed Species_Name from the table Microbe and then added Cell_Wall_Type to table Microbe

- Added Atom_Address attribute to table Atomic_Coordinates

- Removed PDB_ID from the Amino_Acid_Address table. Added AMP_ID, AA_Address to the table Amino_Acid_Address table.

All Species of microbe have either + or – for cell wall type so we added Cell wall type to be more specific. The Atomic Coordinates table was missing a unique primary key, so we added Atom_Address, which is unique, as the primary key. PDB_ID was removed as a foreign key from Amino_Acid_Address as there was no need for direct connection between the Amino_Acid_Address table and the 3D_Structure table. If needed, we can we can still join 3D_Structure table and Amino_Acid_Address table through the Peptide table.

After these changes were made, we checked the database schema for compliance with normal forms. It is currently in 1st, 2nd, and 3rd normal form.

Schema

The schema defines the database tables. The fields (attributes) in each table show the datatypes of the associated data. The schema also shows the relationships between the tables. The keys (unique identifiers of the rows of the table) and the foreign keys (references from one table to another used to combine information among tables) are also identified in the schema. The schema imposes integrity constraints that ensure that the changes made by authorized users do not result in loss of data and that information derived from multiple tables is correct and consistent.

See appendix E for updated schema of Amped database.
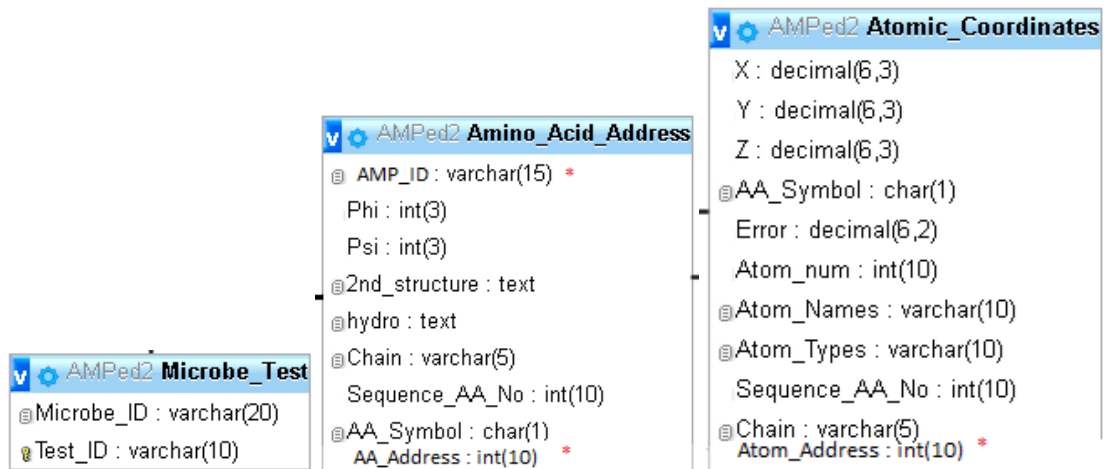
Figure 21: Few tables.

REFERENCES

[1] Database. URL: http://searchsqlserver.techtarget.com/definition/database , Last accessed 10/23/2017.

CHAPTER 6

CONCLUSION AND FUTURE WORK

The Anti-Microbial Peptide Editable Database (AMPed) is a collection of data describing antimicrobial peptides, their sources, and their targets. It has been in development for several years under the direction of Professor Lenore Martin. It is a joint effort of many dedicated undergraduate and graduate students, with guidance from many Professors.

The primary goal for this project was to build tools that would be helpful for researchers to calculate and visualize analysis results, normalize the database, and add important enhancements to the website, such as mission and vision, privacy policy and terms of use.

The MIC/MLC calculator was developed and is being beta tested by the research group in Dr. Martin's lab. The tool is working as expected. It was developed and tested using sample data with unexpected constraints that may occur while testing on real data. The tool was developed keeping that in mind that authorized users may need to extend or modify it. The algorithm for the mass spectrometry tool was also developed. It will guide future development of the system by students. Time constraints prohibited the implementation this tool; other phases of the project such as MIC/MLC tool development, schema evolution and interface development took longer than expected.

The field of biology is ever growing and accumulates large volumes of data, Sorting data to uncover research results is a huge task. One of the contributions of these tools

to manipulate the data for better visualization and to save time. These are the qualities that researchers look for in any tools.

The collection of data is not enough; data should also be easy to access when needed. Our database has a huge collection of peptides and related information, so the interface to access that data should be user friendly as well. No one wants spend more time than necessary to just access the data, so the GUI should be easy to maneuver. One of the goals was to add new features to the existing interface in such a way that it does not look like they were randomly added. The flow of the interface should be maintained. While adding features all these important points were considered. Our GUI is better and user friendly, and easy to maneuver. Complex looking interfaces confuse the user which is not considered a good quality.

Identifying correct relationships between data is a very crucial part of developing a database. While creating and updating our database, careful observations were made regarding the relationships between entities to avoid any redundancies and inconsistencies that may occur while inserting, deleting and updating the data in the database. Changes in the database design were made to prevent these problems.

The original goals of this work were achieved. We have furthered our efforts to make our database a hub for sharing resources and data between researchers in the same field. These goals have been achieved with the guidance of Dr. Martin. There are several other features that presently reside in AMPed, developed by students who worked on this project before. The work of this thesis helps to make the database more useful to researchers interested in anti-microbial peptides. The ultimate goal is a full functioning database with all required user and administrative features to support a

repository of anti-microbial peptides. This would be similar to BRENDA, a comprehensive Enzyme information system [1].

In the future, a new tool also needs to be developed to automatically download content from a variety of sources and parse it to the AMPed site. This would continue populating the site with the ever-expanding amount of newly discovered and validated content

REFERENCES

[1] Technische Universität Braunschweig, BRENDA. URL: https://www.brenda-enzymes.org/ , Last accessed 10/23/2017.

GLOSSARY

AMPed

> The Antimicrobial Peptide Editable Database developed at the University of
> Rhode Island under the direction of Professor Lenore Martin.

Antimicrobial

> An antimicrobial is an agent that kills microorganisms or inhibits their growth,
> which includes antibacterial, antiviral, antifungal and antiprotozoal agents.

Amino acid

> Organic compound that serves as the building blocks of proteins.

AWI

> AMPed Web Interface; the group of web pages and forms that access the
> AMPed database.

Broth

> The liquid media in which bacteria grow. The media contains nutrients and
> minerals needed for allowing the grow of the organism.

DBS

> Database Structure; the structure of the tables conforming the AMPed
> database.

MIC

> Minimum inhibitory concentration (MIC) is the lowest concentration of the
> peptide that will inhibit the growth of a targeted microorganism.

MLC

Minimal Lethal Concentration (MLC)is the lowest concentration of a drug at which bacteria are killed.

MOLECULAR WEIGHT

Molecular weight is the total mass of a given peptide as determined by the sum of the composite amino acids (as determined by NIST) minus the mass of water for each peptide bond in the peptide.

MySQL

MySQL is an open-source relational database management system (RDBMS) created by a Swedish company, MySQL AB.

MASS SPECTROMETRY

An analytical technique used to characterize the structure of a molecule. During the process of mass spectrometry, the molecule of interest is fragmented and the mass/charge ratio of these fragments are determined via instrumentation.

NaOH

Sodium hydroxide, a basic chemical used as a control in the experiment. The NaOH will be lethal to the bacteria.

NCBI

National Center for Biotechnology Information (NCBI) is a US government resource that develops, distributes, supports, and coordinates access to a variety of databases and software for the scientific and medical communities.

PDB

Protein Data Bank (PDB) is a repository of curated 3D biological structures, usually proteins, populated by direct input from an international community of scientists. PDB utilizes different file formats (mmCif/XML) to display protein structure information.

Peptide

A small protein usually less than 100 amino acids in length.

Perl

Perl (short for "Practical Extraction and Report Language") is a highly capable, feature-rich programming language with over 29 years of development. Perl runs on over 100 platforms from portables to mainframes and is suitable for both rapid prototyping and large scale development projects.

Pseudomonas aeruginosa (PA)

A Gram negative rod-shaped bacteria that may be used to assess the properties of the AMP.

R

R is a language and environment for statistical computing and graphics. It is available as free software, compiles and runs on a wide variety of UNIX platforms.One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed.

R Studio

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well

as tools for plotting, history, debugging and workspace management. It is available in open source.

Staphylococcus aureus (SA)

A Gram positive round-shaped bacteria that may be used to assess the inhibitory/lethal properties of the AMP.

Appendix A

Sample data file in excel, there can be many blocks as generated by the experimental data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Blank	128	64	32	16	8	4	2	1	0.5	0.25	Microbe	Time point
2	A	0.041	0.043	0.046	0.047	0.043	0.042	0.042	0.047	0.041	0.042	0.042	0.052	0
3	B	0.043	0.044	0.046	0.044	0.046	0.046	0.044	0.042	0.047	0.046	0.044	0.052	
4	C	0.045	0.041	0.047	0.041	0.046	0.047	0.041	0.046	0.042	0.042	0.044	0.057	
5	D	0.048	0.047	0.042	0.047	0.047	0.041	0.047	0.047	0.046	0.044	0.041	0.063	
6	E	0.049	0.042	0.047	0.042	0.042	0.047	0.044	0.044	0.047	0.041	0.047	0.064	
7	F	0.046	0.047	0.042	0.046	0.044	0.042	0.041	0.041	0.042	0.047	0.082	0.06	
8	G	0.05	0.047	0.042	0.047	0.042	0.047	0.044	0.044	0.047	0.041	0.041	0.041	
9	H	0.045	0.049	0.049	0.042	0.048	0.049	0.042	0.0491	0.042	0.041	0.05	0.0492	
10														
11		Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Time point
12	A	0.041	0.046	0.046	0.047	0.043	0.047	0.082	0.084	0.072	0.072	0.078	0.08	0.5
13	B	0.043	0.046	0.046	0.042	0.046	0.052	0.087	0.083	0.069	0.069	0.079	0.079	
14	C	0.042	0.046	0.041	0.047	0.046	0.053	0.09	0.084	0.069	0.068	0.078	0.079	
15	D	0.041	0.047	0.042	0.04	0.047	0.049	0.089	0.087	0.072	0.073	0.081	0.079	
16	E	0.042	0.042	0.047	0.046	0.042	0.053	0.091	0.082	0.069	0.069	0.079	0.079	
17	F	0.042	0.047	0.042	0.046	0.047	0.057	0.091	0.082	0.069	0.0714	0.079	0.081	
18	G	0.042	0.078	0.078	0.076	0.078	0.076	0.078	0.076	0.078	0.078	0.076	0.078	
19	H	0.041	0.049	0.049	0.042	0.048	0.044	0.044	0.041	0.04	0.053	0.051	0.062	
20														
21		Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Time point
22	A	0.044	0.046	0.046	0.047	0.043	0.051	0.09	0.091	0.074	0.074	0.086	0.089	1
23	B	0.046	0.046	0.046	0.042	0.046	0.055	0.091	0.092	0.091	0.071	0.088	0.088	
24	C	0.045	0.046	0.047	0.047	0.046	0.052	0.094	0.091	0.092	0.07	0.086	0.087	
25	D	0.043	0.047	0.042	0.049	0.047	0.052	0.094	0.091	0.071	0.071	0.089	0.087	
26	E	0.043	0.042	0.047	0.046	0.042	0.058	0.09	0.085	0.07	0.073	0.082	0.084	
27	F	0.039	0.047	0.042	0.046	0.047	0.059	0.095	0.085	0.071	0.073	0.081	0.084	
28	G	0.049	0.078	0.08	0.086	0.08	0.078	0.08	0.086	0.08	0.078	0.086	0.09	
29	H	0.049	0.049	0.049	0.04	0.048	0.044	0.044	0.04	0.04	0.051	0.042	0.097	
30														
31		Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Time point
32	A	0.042	0.046	0.046	0.047	0.043	0.054	0.098	0.1	0.073	0.073	0.096	0.1	1.5
33	B	0.043	0.046	0.046	0.042	0.046	0.056	0.1	0.099	0.075	0.075	0.096	0.097	
34	C	0.044	0.046	0.047	0.047	0.046	0.06	0.121	0.099	0.079	0.079	0.095	0.096	
35	D	0.042	0.047	0.042	0.049	0.047	0.056	0.11	0.1	0.1	0.079	0.098	0.098	
36	E	0.045	0.042	0.047	0.046	0.042	0.062	0.099	0.097	0.099	0.08	0.096	0.096	
37	F	0.043	0.047	0.042	0.046	0.047	0.061	0.102	0.097	0.099	0.081	0.095	0.099	
38	G	0.042	0.084	0.085	0.082	0.084	0.084	0.085	0.082	0.084	0.084	0.084	0.095	
39	H	0.041	0.049	0.049	0.042	0.048	0.05	0.048	0.04	0.042	0.051	0.041	0.104	

Table 1: Sample data in excel in .csv format

Appendix B

PHP side coding after data file is submitted for mic/mlc calculator

```php
1    <?php session_start(); ?>
2    <!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
3    <html>
4      <head>
5        <meta content="text/html; charset=windows-1252" http-equiv="content-type">
6        <link href="css/bootstrap.min.css" rel="stylesheet">
7        <link href="css/main.css" rel="stylesheet">
8        <meta name="viewport" content="width=device-width, initial-scale=1">
9        <link rel="shortcut icon" href="/favicon.ico" type="image/x-icon">
10       <link rel="icon" href="/favicon.ico" type="image/x-icon">
11      </head>
12      <body>
13       <?php
14    $bacteriaName   = $_POST["Bacteria_Name"];
15    $assay          = $_POST["Assay_Name"];
16    $medium         = $_POST["Medium_Name"];
17    $peptide1       = $_POST["Peptide1_Name"];
18    $peptide2       = $_POST["Peptide2_Name"];
19    $antibiotic     = $_POST["Antibiotic_Name"];
20    $con            = $_POST["Con_Name"];
21    $timepoint      = $_POST["Time_Point"];
22       ?>
23
24       <?php
25    //function to open .r from php and write rscript everytime a file is uploaded with dyanmic/unique name.
26    function writeFile($idcode) {
27    global $fileName;
28    $text1 = "#m_m\nMyData <- read.csv(\"mic-data/".$fileName. "\", header=TRUE, sep=\",\", colClasses = \"character\")\n";
29    $constantTXT = <<<END
30                    colnames(MyData) <- c("FirstCol",colnames(MyData)[2:14])
31                    Mycl_data <- subset(MyData, FirstCol == "A" | FirstCol == "B" | FirstCol == "C" |
32                    FirstCol == "D" | FirstCol == "E" | FirstCol == "F")
33
34                    ## Fill in time point will valid time
35                    for(i in 1:nrow(Mycl_data))
36                    {
37                        ifelse(Mycl_data$Time.point[i]=="", Mycl_data$Time.point[i] <- val_time,
38                        val_time <- Mycl_data$Time.point[i])
39                    }
```

```r
41                    cl_names <- colnames(Mycl_data[2:ncol(Mycl_data)])## Extracting column names
42
43                    Mycl_data <- Mycl_data[,c(-1)] ## Get rid of the first irrelevant column
44
45                    Pep1 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)
46                    Pep2 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)
47                    temp = 3
48                    count = 1
49                    for (i in 1:nrow(Pep1))
50                    {
51                        for (j in 1:ncol(Pep1))
52                        {
53                            Pep1[i,j] = round(mean(as.numeric(Mycl_data[c(count:temp),j])),digits = 5)
54                            Pep2[i,j] = round(mean(as.numeric(Mycl_data[c((count+3):(temp+3)),j])),digits = 5)
55                        }
56
57                        count = temp + 4
58                        temp = temp + 6
59                    }
60
61                    colnames(Pep1) <- cl_names
62                    colnames(Pep2) <- cl_names
63
64                    rownames(Pep1) <- c(1:nrow(Pep1))
65                    rownames(Pep2) <- c(1:nrow(Pep2))
66
67                    Pep1 <- as.data.frame(Pep1)
68                    Pep2 <- as.data.frame(Pep2)
69                    print(Pep1)
70                    print(Pep2)
71
72
73    END;
74    $text2 = "png(file = \"mic-output/".$fileName. "-pep1.png\")\n
75    matplot(Pep1$Time.point,Pep1[,c(1:12)],type = c(\"b\"),lty = 1,lwd =2,  xlab = \"Time Points\",ylab = \"OD\",main = \"MIC_MLC Peptide1\",pch=c(15,16,17),col = 1:12)\n
76    legend(\"topleft\", legend = colnames(Pep1), col=1:12, box.lwd =par(\"lwd\") ,pch=c(15,16,17)) # optional legend\n
77    dev.off()\n";
```

56

```php
78      $text3 = "png(file = \"mic-output/".$fileName. "-pep2.png\")\n
79      matplot(Pep2\$Time.point,Pep2[,c(1:12)],type = c(\"b\"),lty = 1,lwd =2,  xlab = \"Time Points\",ylab = \"OD\",main = \"MIC_MLC Peptide2\",pch=c(15,16,17),col = 1:12)\n
80      legend(\"topleft\", legend = colnames(Pep2), col=1:12, box.lwd =par(\"lwd\") ,pch=c(15,16,17)) # optional legend\n
81      dev.off()\n";
82      $text4 = "TotalPep <- rbind(Pep1,Pep2)\n
83      write.csv(TotalPep,\"mic-output/".$fileName. "-result.csv\")";
84
85      $myfile= fopen(getcwd()."/R-scripts/" . $idcode . "-mic.R", 'w') or die("Could not create the file.");
86      fwrite($myfile, $text1);
87      fwrite($myfile, $constantTXT);
88      fwrite($myfile, $text2);
89      fwrite($myfile, $text3);
90      fwrite($myfile, $text4);
91      fclose($myfile);
92      }
93
94      include("top_header_start.php"); include("top_header_logo.php"); include("top_header_menu.php");
95      echo "<div class=\"container\"><div class=\"row\"><div class=\"col-sm-12\"><br/><br/><table>\n\n";
96          //uploaded file parameters
97      $_FILES["upload_file"]["name"];     // file's original filename
98      $_FILES["upload_file"]["type"];     // file's mimetype
99      $_FILES["upload_file"]["tmp_name"]; // temporary filename
100     $_FILES["upload_file"]["size"];     // file's size
101     $_FILES["upload_file"]["error"];    // error code
102
103
104     if ($_FILES["upload_file"]["error"] > 0) {
105     $thisMsg = "<p>Error transfer";
106     }
107     $idcode= time();
108     $fileName = $idcode. "-" . "data.csv";
109     $finalLoc = getcwd()."/mic-data/".$fileName;
110     $URLloc = "/mic-data/".$fileName;
111     $result = move_uploaded_file($_FILES["upload_file"]["tmp_name"],$finalLoc);
```

```php
113     if (!$result) {
114         $thisMsg = "<p>Unfortunately the file could not be uploaded: $fileName.</p>";
115     } else {
116         $thisMsg = "<p>Download excel file: <a href=\"$URLloc\" target=\"_blank\">$fileName</a>.</p>\n";
117     }
118
119     writeFile($idcode); // create R script
120     $Rcommand = "Rscript R-scripts/" . $idcode . "-mic.R";
121     exec($Rcommand); // execute R script
122
123     $f = fopen("mic-output/" .$fileName. "-result.csv", "r");
124     while (($line = fgetcsv($f)) !== false) {
125         echo "<tr>"."  &nbsp";
126         foreach ($line as $cell) {
127             echo "<td>"."  &nbsp" . htmlspecialchars($cell) . "</td>";
128         }
129         echo "</tr>\n";
130     }
131     fclose($f);
132     $pep1png = "mic-output/".$fileName."-pep1.png";
133     $pep2png = "mic-output/".$fileName."-pep2.png";
134     echo "  &nbsp                 
135             <br/>";
136     echo "<br/><br/>";
137     echo "\n</table>";
138     echo "<br/>";
139     echo $thisMsg;
140     echo "<br/><br/>";
141     echo '<img src="'.$pep1png.'">';
142     echo '<img src="'.$pep2png.'">';
143     echo "<br/>";
144     echo '<a href="'.$pep1png.'" download>Download-PEP1</a><br/>';
145     echo '<a href="'.$pep2png.'" download>Download-PEP2</a>';
146     ##echo user input
147     echo "<p>Bacteria Name :$bacteriaName"."          "."Peptide2 Name : $peptide2</p>";
148     echo "<p>Assay Name :$assay"."          "."Antibiotic Name :$antibiotic</p>";
149     echo "<p>Medium :$medium"."          "."Antibiotic concentration :$con</p>";
150     echo "<p>Peptide1 Name :$peptide1"."          "."Timepoint :$timepoint</p>";
151     ?>
```

```php
150     echo "<p>Peptide1 Name :$peptide1"."          "."Timepoint :$timepoint</p>";
151     ?>
152     <?php include("footer.php"); ?>
153     </body>
154     </html>
```

PHP Hypertext Preprocessor file   length : 6,233 lines : 154   Ln : 101 Col : 54 Sel : 0 | 0   Unix (L

Figure 22: PHP coding in page after submitting data file for mic/mlc

Appendix C

One instance (1510000255) of code in dynamically created .r script.

```r
## R script to calculate MLC/MLC ##

## Written by: Sunandha Acharya ##

MyData <- read.csv("mic-data/1510000255-data.csv", header=TRUE, sep=",",

colClasses = "character")

colnames(MyData) <- c("FirstCol",colnames(MyData)[2:14])

Mycl_data <- subset(MyData, FirstCol == "A" | FirstCol == "B" | FirstCol == "C" |

FirstCol == "D" | FirstCol == "E" | FirstCol == "F")

## Fill in time point will valid time

for(i in 1:nrow(Mycl_data))

{

        ifelse(Mycl_data$Time.point[i]=="", Mycl_data$Time.point[i] <- val_time,

        val_time <- Mycl_data$Time.point[i])

}

cl_names <- colnames(Mycl_data[2:ncol(Mycl_data)])## Extracting column names

Mycl_data <- Mycl_data[,c(-1)] ## Get rid of the first irrelevant column

Pep1 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)

Pep2 <- matrix(0, nrow = (nrow(Mycl_data)/6), ncol = 13)

temp = 3

count = 1

for (i in 1:nrow(Pep1))
```

```r
{
        for (j in 1:ncol(Pep1))

        {

Pep1[i,j] = round(mean(as.numeric(Mycl_data[c(count:temp),j])),digits = 5)

Pep2[i,j] = round(mean(as.numeric(Mycl_data[c((count+3):(temp+3)),j])),digits = 5)

        }


        count = temp + 4

        temp = temp + 6

}

colnames(Pep1) <- cl_names

colnames(Pep2) <- cl_names

rownames(Pep1) <- c(1:nrow(Pep1))

rownames(Pep2) <- c(1:nrow(Pep2))

Pep1 <- as.data.frame(Pep1)

Pep2 <- as.data.frame(Pep2)

print(Pep1)

print(Pep2)

png(file = "mic-output/1510000255-data.csv-pep1.png")

matplot(Pep1$Time.point,Pep1[,c(1:12)],type = c("b"),lty = 1,lwd =2,  xlab = "Time

Points",ylab = "OD",main = "MIC_MLC",pch=c(15,16,17),col = 1:12)

legend("topleft", legend = colnames(Pep1), col=1:12, box.lwd =par("lwd")

,pch=c(15,16,17)) # optional legend
```

```
dev.off()


png(file = "mic-output/1510000255-data.csv-pep2.png")

matplot(Pep2$Time.point,Pep2[,c(1:12)],type = c("b"),lty = 1,lwd =2,  xlab = "Time

Points",ylab = "OD",main = "MIC_MLC",pch=c(15,16,17),col = 1:12)

legend("topleft", legend = colnames(Pep2), col=1:12, box.lwd =par("lwd")

,pch=c(15,16,17)) # optional legend

dev.off()

TotalPep <- rbind(Pep1,Pep2)

write.csv(TotalPep,"mic-output/1510000255-data.csv-result.csv")
```
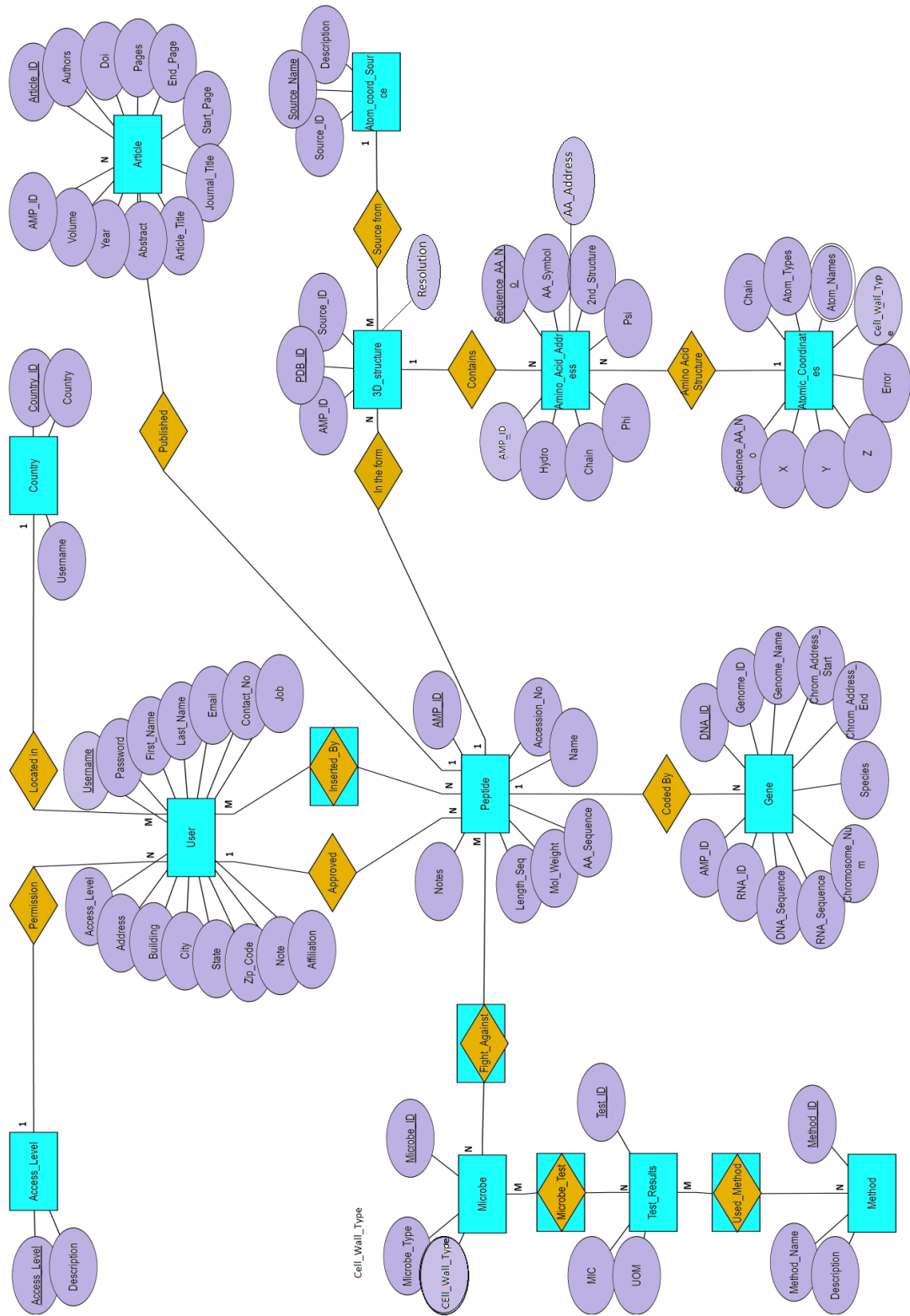
Appendix D

Figure 23: AMPed ER diagram
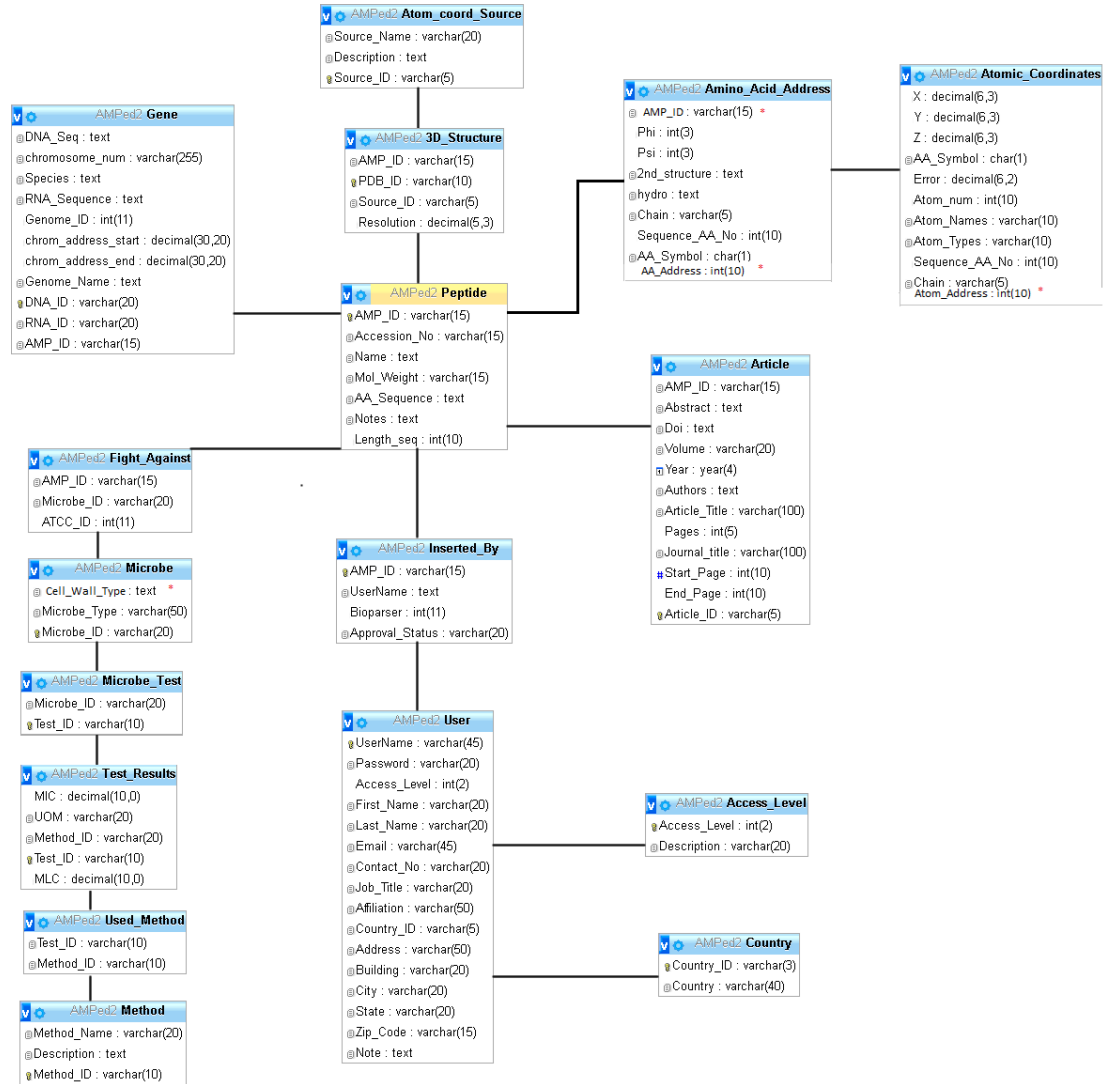
Appendix E

Figure 24: Updated Schema of AMPed Database

BIBLIOGRAPHY

[1] Antimicrobial Peptides. URL:

https://en.wikipedia.org/wiki/Antimicrobial_peptides, Last accessed 10/10/2017

[2] BRENDA. URL: https://www.brenda-enzymes.org/ , Last accessed 10/23/2017.

[3] California Law Online Privacy. URL: https://oag.ca.gov/privacy/privacy-laws

, Last accessed 10/23/2017

[4] Database initially developed by the Dr. Martin's research Group, URL:

https://amped.uri.edu, Last accessed 10/10/2017

[5] Dr. Lenore Martin. URL:http://web.uri.edu/cmb/lenore-martin/ , Last accessed

10/20/2017.

[6] exec. URL: http://php.net/manual/en/function.exec.php , Last accessed 10/20/2017

[7] Form content Type. URL:

http://www.w3.org/TR/html401/interact/forms.html%23h-17.13.4#h-17.13.4 , Last

accessed 10/20/2017

[8] Garg, Tripti, "*Graphical User Interface for Antimicrobial Peptide Database*"

(2016). *Open Access Master's Theses.* Paper 831. Available:

http://digitalcommons.uri.edu/theses/831

[9] Herrera, Abraham, "*Web Based Bioinformatics Tool for Contributing Content to

the Anti-Microbial Peptide Editable Database*" (2016). *Open Access Master's

Theses.* Paper 1042. Available:

http://digitalcommons.uri.edu/theses/1042

[10] The Importance of Privacy Policies. URL:

https://www.katalystsolutions.com/newsletter/57-the-importance-of-privacy-policies.html

, Last accessed 10/23/2017

[11] Importance of Privacy Policy.

URL: https://www.privacytrust.com/guidance/privacy_policy.html , Last accessed

10/23/2017

[12] Integrate R with PHP. URL: http://www.listendata.com/2017/04/how-to-

integrate-r-with-php.html , Last accessed 10/20/2017

[13] J. M. Andrews "*Determination of minimum inhibitory concentrations*."

J Antimicrob Chemother. 2001 Jul; 48 (Suppl 1): 5–16. Avaliable:

https://www.ncbi.nlm.nih.gov/pubmed/11420333

[14] Konstantinidis, George, "*Antimicrobial Peptide Editable Database*" (2015).*Open*

 Access Master's Theses. Paper 694. Available:

http://digitalcommons.uri.edu/theses/694

[15] O'Neill AJ, Chopra I. Expert Opin Investig Drugs. 2004 Aug;13(8):1045-63.

"*Preclinical evaluation of novel antibacterial agents by microbiological and*

molecular techniques."[Online] Available:

https://www.ncbi.nlm.nih.gov/pubmed/15268641

[16] Pass Arguments and Execute R script from PHP Forms. URL:

https://www.codeproject.com/Articles/1119237/Pass-Arguments-and-Execute-R-

script-from-PHP-Forms , Last accessed 10/20/2017

[17] PDB Current Holdings Breakdown. URL:

https://www.rcsb.org/pdb/statistics/holdings.do , Last accessed 10/10/2017

[18] Post Method. URL: https://www.tutorialspoint.com/php/php_get_post.htm , Last accessed 10/20/2017

[19] Reasons you need a privacy policy.

URL: https://termsfeed.com/blog/top-4-reasons-you-need-privacy-policy/ , Last accessed 10/23/2017

[20] Roman A. Zubarey, * Per Håkansson, and, and Bo Sundqvist

 "Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements." Analytical Chemistry **1996** *68* (22), 4060-4063

doi: 10.1021/ac9604651. Available: https://pubs.acs.org/doi/pdf/10.1021/ac9604651

[21] Terms of Service. URL: https://en.wikipedia.org/wiki/Terms_of_service , Last accessed 10/23/2017

[22] Time function. URL:

https://www.tutorialspoint.com/php/php_date_and_time.htm, Last accessed 10/20/2017

 [23] Turnidge J, Ferraro M, Jorgensen J. 2011. Susceptibility Test Methods: General Considerations, p 1115-1121. In Versalovic J, Carroll K, Funke G, Jorgensen J, Landry M, Warnock D (ed), *Manual of Clinical Microbiology, 10th Edition.* ASM Press, Washington, DC. doi: 10.1128/9781555816728.ch67 [Online] Available: http://www.asmscience.org/content/book/10.1128/9781555816728.chap67

[24] What is HTTPS? URL: https://www.instantssl.com/ssl-certificate-products/https.html , Last accessed 10/12/2017

[23] Yamamoto H., McCloskey James A.*" Calculations of Isotopic Distribution in Molecules Extensively Labeled with Heavy Isotopes*" [Online] Available: http://pubs.acs.org/doi/abs/10.1021/ac50010a025