Brandon Laufer

CDS 302 Final Project

**Objective**

Given that bacteria is becoming more drug resistant as time passes, there is a need for other types of antimicrobials. One such is cationic antimicrobial peptides. Using nanogel particle harvests, a large amount of data has been obtained on peptide sequences contained within the blood of several species of animal.

To better the ability to find similar sequences and keep track of all of the data, a database is necessary. This database tracks all harvests and the mass spectrometry data from the harvests (which includes the peptide sequences).

Missing discussion of why MS Access was used.

**Description of Data and Data Preparation**

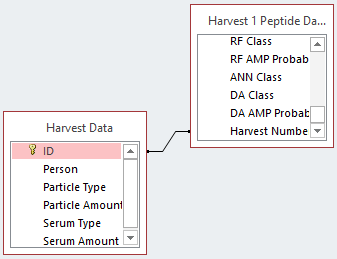
The harvest data comes from the (hand-written or electronic?) laboratory notebooks of those performing peptide harvesting within Dr. Barney Bishops’ lab. Mass spectrometry data was obtained from the runs performed on a Thermo Scientific MS/MS with Orbitrap. This data is then passed through PEAKS de novo software to check the peptide sequences against the genome of the given species. Once this has been performed, the data is run through Python code developed by Shaylyn Scott and myself to add in data pertaining to the sequence. This data includes, but is not limited to, charge, molecular weight, and isoelectric point. The output of this is an Excel spreadsheet (\*.xlsx).

Microsoft Access was then used to create a main table entitled Harvest Data. This table includes an auto number column, who performed the harvest, type of harvest, amount of particles used, type of serum, and amount of serum. To facilitate input into this table, a form was generated via Access. The auto number field was removed from the form so that it is not changed accidently and a button was added that automatically adds the entered information into a new line on that table and blanks the form. The code for this can be found in project\_vba\_I.txt.

To import the Excel spreadsheets, another form was created. Only a title and Import Excel Data button were added to this form. The button runs VBA code that allows the user to enter the number that corresponds to the auto number column on the Harvest Data table. It then opens a dialog box that allows the user to select the file that will be imported.

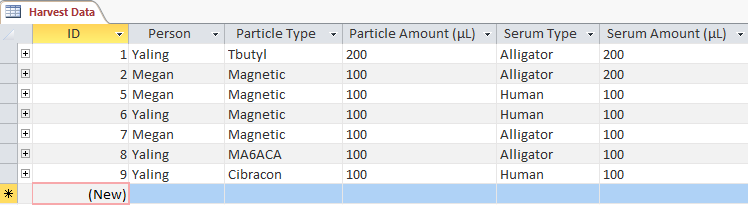
When the file is imported, the code removes header lines and adds column names. It also adds an additional column with the previously chosen number so that it can be linked to the Harvest Data table. While importing the file, it names the new table “Harvest x Peptide Data” where x is the number chosen. This code can be found in project\_vba\_II.txt.

**Overview of Data**

 From the sequence and information about it, what can be gained is the ability to search through and find those with similar properties to known cationic antimicrobial peptides. On the other side of this, it can also help to eliminate some of the sequences based upon properties found from peptides that do not show antimicrobial activity.

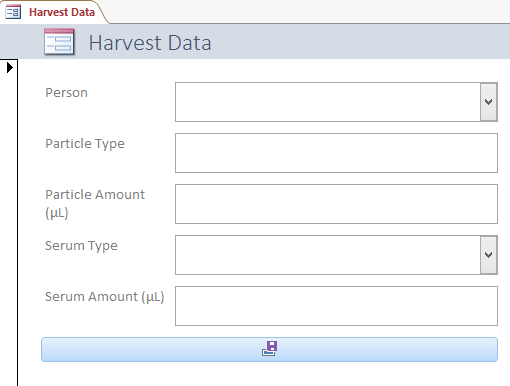
Using the information about the harvests and sequences, future harvests will be able to target specific features. The relationships that are formed from using a table with all of the harvest details and linking it to its peptide data should facilitate better control and quicker access to this feature. Figure 1 shows how these relationships work.

**Figure 1.** Relationship example



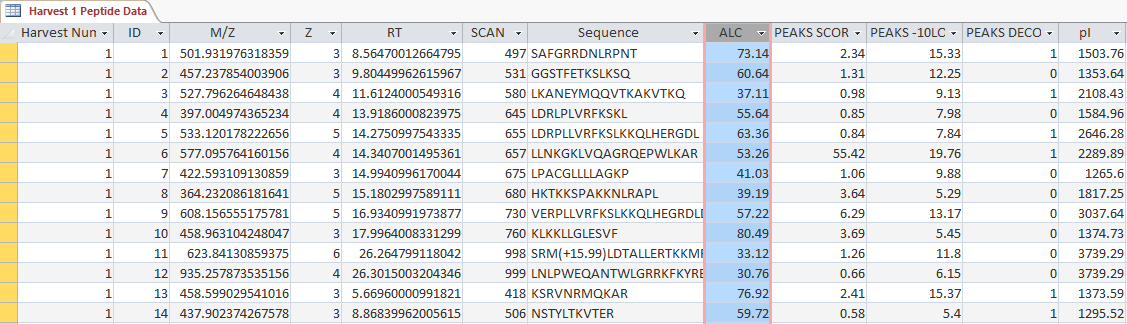
**Figure 2.** Harvest Data table example

This figure represents only one such relationship but others are formed each time a new set of peptide data is entered.

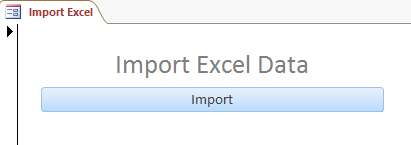
 Figure 2 gives an overview of how the data is represented for the Harvest Data. The corresponding form for this table can be seen in Figure 3. Setup of the form is incomplete as only some fields have dropdown boxes allowing for all data to be entered in the same format, future additions will be to standard-ize all fields. The button seen on the bottom of the form uses the code in project\_vba\_I.txt to allow the user to enter data into the table and then blank the form for the next set of data.

**Figure 3.** Harvest Data form

Although the peptide data usually contains hundreds to thousands of lines, Figure 4 gives an example of the table created. In order to input these tables, a form is again used and can be seen in Figure



**Figure 4.**  Peptide Data table example

5. This form also contains a button that runs code as stated previou-sly and can be found in project\_vba\_II.txt.

**Figure 5.** Import excel data form

**Analysis**

Due to time limitations and the amount of data needing to be entered, data analysis has not been performed. After more data has been entered, analysis will include queries that allow for the search of all sequences with specific patterns and or trends based upon the additional information regarding that sequence.

**Discussion**

Since analysis has not yet been performed, the future possibilities of this database is speculative. What is hoped to be found by this is quicker reference to whether or not a peptide has antimicrobial properties. This should also allow for the ability to create sequences that have not as of yet been found within organisms but show potential for having the necessary properties.