Exploring the Molecular Basis of Insecticide Resistance

Part I: Identifying the Cause

Although Zika virus was first identified in monkeys in Uganda in 1947, it was not discovered in humans until 1952. This mosquito-borne virus typically causes mild symptoms, including low fever, achiness, headaches, and skin rash, lasting 2-7 days. The first large outbreak in humans occurred in 2007. Ongoing outbreaks were recorded worldwide, but symptoms were mild. However, in 2015, a large outbreak in Brazil became internationally known because infection was associated with Guillain-Barré syndrome (an autoimmune disorder attacking the nervous system) in adults and microcephaly in newborns. In the most recent outbreak, it was determined that Zika can also be transmitted sexually.

Background Research (https://www.cdc.gov/zika/index.html)
Find out about Zika Virus – how it is transmitted, symptoms and severity, and prevention and treatment. Also determine if Zika is currently causing infections where you live. Even if the virus is not being transmitted via mosquitoes where you live, why is it important to be aware of *where* the virus is currently active? *Write a paragraph summarizing your findings.*

Many diseases, including malaria and Zika, are transmitted to humans by mosquitoes. As a way of controlling the spread of these mosquito-borne illnesses, insecticides began to be used widely in the 1960s. Many insecticides, such as DDT and Malathion, target the nervous system.

Neurotransmitters are small molecules that send signals from transmitting neurons to adjacent neurons or muscle cells. Acetylcholine is a neurotransmitter that is produced from acetyl-CoA and choline within the transmitting neuron, and, once released, it binds to cholinergic receptors on other neurons or muscle cells. After the signal is received and the nerve impulse is transmitted, acetylcholine diffuses away from the receptor and is broken down by acetylcholinesterase (AChE) in the synaptic space. Choline and acetic acid are then recycled to the transmitting neuron.

AChE is the target of many insecticides which act as inhibitors of the enzyme. These insecticides were very effective when they were first introduced, but soon mosquito populations became predominantly resistant to insecticides. This is because some mosquitoes possess variations in their DNA sequences coding for AChE that are still able to break down acetylcholine normally, but are resistant to inhibition by insecticides. As insecticides kill mosquitoes that are susceptible to the drugs, those that are resistant are still able to reproduce, and the mosquito population becomes resistant to treatment with insecticides.

Your task is to explore the molecular basis for insecticide resistance, then design a new insecticide that will be effective against the resistant strain of mosquitoes.

Predictions
Mosquito acetylcholinesterase is a 702 amino acid protein. *How many mutations do you predict are necessary to confer resistance to insecticides?* Note that this is a prediction and there is no right or wrong answer.

Looking at the DNA
Your colleagues have isolated the mRNA from neurons of mosquitoes that are resistant (R) or sensitive (S) to insecticides. From the mRNA, they constructed cDNA and sequenced the genes.

Of what advantage is it to use mRNA instead of sequencing the gene directly?

The cDNA sequences can be found in DataFile1, in FASTA format. FASTA format is a standard format used by a variety of bioinformatics tools. Files in FASTA format always begin with a definition line (preceded by a greater than symbol ">"), followed by sequence information. Identify the differences in these sequences by using a sequence alignment tool (http://www.ebi.ac.uk/Tools/msa/clustalo/). Copy and paste the entire file into the sequence box, use the standard parameters and be sure to select the correct type of data before submitting. *Note: Since you may wish to refer back to your data later, you should save a copy of all the data you collect!*

How many differences did you find between the two sequences? Is this what you expected?
Discuss the location of the variations between the two sequences. Are they evenly distributed in the sequence? Why or why not?

Comparing Protein Sequences

Your focus is on the *expression* of the DNA in the proteins. So your next step is to translate the sequences (http://bio.lundberg.gu.se/edu/translat.html) then align the protein sequences, using the same alignment tool you used for the nucleotide sequences. Note that you will need to translate each sequence separately, and you should copy ONLY the DNA sequence, not the definition line. Save the protein sequences in a Word file, and add a definition line to each sequence. (Save the file; you'll need it later!) You do NOT have to remove the numbers and spaces from the output, as these are recognized in FASTA. You MUST remove the asterisk at the end of the sequence, as this is NOT recognized in FASTA.

How many differences are there between the two protein sequences? Is this what you expected?

Part II: Exploring Further

A visitor to your lab is puzzled that there can be so many differences in the DNA sequences between the sensitive and resistant strains, but only a single difference in the protein. You have compiled the data in the following table in preparation for publishing your results:

Figure 1: Acetylcholinesterase mRNA compared in SLA-B (sensitive to insecticides) and SR (resistant to insecticides). Nucleotides that differ between the two strands are listed by number. SLA-B codon sequences are listed on top, with SR sequences below, with the nucleotide which differs highlighted in red. Only one nucleotide difference, at base 739, results in a change in amino acid sequence, from glycine to serine. This residue is highlighted in yellow.

Nucleotide number	Codon	Nucleotide number	Codon	Nucleotide number	Codon	Nucleotide number	Codon
402	GUC GUA	696	CCC CCG	846	GGG GGU	1314	GGG GGU
450	CCU CCA	714	GCC GCU	864	ACA ACU	1518	AUU AUC
573	CCG CCA	732	UUC UUU	876	CCC CCU	1554	CUG CUC
624	GGG GGC	739	GGC AGC	1077	CCG CCA	1599	GAA GAG
627	GCC GCG	763	CUG UUG	1182	UGC UGU	1632	AAC AAU
651	CCG CCC	777	GAC GAU	1218	AAC AAU	1689	ACC ACU
691	CGG AGG	813	GUA GUG	1272	GGA GGU	1827	CCC CCA

Using your knowledge about the genetic code and the data above, explain to your visitor why most of the differences in the DNA are not expressed in the protein.

Confirming the Results

In a lab meeting, you and your colleagues discuss your initial findings. You realize that it is possible that there may be other mutations that could confer resistance. You decide to collect mosquitoes from around the world, some that are resistant and others that are sensitive to insecticides. (Fortunately, you have collaborators around the world who are willing to send samples!) You decide to compare sensitive and resistant strains of *Culex pipiens pipiens* (the Northern house mosquito found in temperate regions) and *Culex pipiens quinquefasciatus* (a tropical subspecies). You also have access to a single strain of *Culex torrentium* (a mosquito prevalent in Europe).

You want to stretch your research funding dollars, so instead of isolating mRNA and making cDNA copies of the entire AChE gene, you decide to use PCR to amplify the region of the gene (exon 5) in which you have identified the important mutation.

Do a multi-sequence nucleotide alignment and conduct a phylogenetic tree from the data collected in the lab: (http://www.ebi.ac.uk/Tools/msa/clustalo/). Note that *Culex pipiens* samples include a letter and number; these represent strains that are known to be either resistant (R) or sensitive (S) to insecticides. The status of the *C. torrentium* fly is unknown. The sequences to use are in DataFile2.

Because you are only looking at a part of the gene, the first nucleotide contained in all the sequences in the alignment is 451. Though it may be easier to print a copy of the results for the next step of the analysis, please be aware that not all printers will maintain the nice alignment of the sequences. You may prefer to copy and paste the results in a Word document, then highlight various features in the Word document.

Identify which sequences are from mosquitoes that are sensitive and resistant to insecticides. (You may also want to distinguish between the different subspecies.) Note that the order of the sequences in the output varies from the order given in the handout! Then look at the nucleotides that differ among the aligned sequences.

You may wish to generate and save a copy of the phylogenetic tree that shows the relatedness of these sequences in a visual diagram. At the top of the output you'll see a dropdown menu with "select tree menu". Select "Rooted phylogenetic tree with branch length (UPGMA)" then click the Exec button. Data will be displayed in a new window.

Which nucleotide differences are due to differences between species? Which nucleotide differences are due to random variation within a species – but not related to insecticide resistance? Which nucleotide differences are found that distinguish mosquitoes sensitive and resistant to insecticides? Predict whether the C. torrentium mosquito is likely to be resistant or sensitive to insecticides.

Part III: Molecular Basis of Insecticide Resistance

At this point, you and your collaborators are convinced that you are onto something quite interesting. But now you want to know *why* a change from glycine to serine at amino acid 119 can lead to insecticide resistance.

Differences between Glycine and Serine

Consider the structures of glycine and serine. (You may wish to consult a biochemistry textbook or pull out the Amino Acid Starter Kit to help.)

What are the differences in the structure of the sidechains of these two amino acids? Compared to all 20 amino acids, do you think this is a significant change in SIZE of the sidechain?

Now consider the chemical properties of these two amino acids. Are they the same or different?

Based on your knowledge of protein structure, discuss whether you think the differences in size or chemical properties of these two amino acids will contribute more to resistance to insecticides.

Exploring AChE Protein Structure

Though you know where the G119S change occurs in the linear amino acid sequence of the AChE protein, you want to explore where it is positioned in the folded protein. A quick review of structures in the Protein Data Bank (http://pdb.org) reveals numerous structures, but none from the mosquito. You discover AChE is featured in David Goodsell's Molecule of the Month (http://www.pdb.org/pdb/101/motm.do?momID=54). Read the article to get some background information on AChE structure and function.

You and your colleagues finally decide to work with two PDB structures, 1amn and 1qon. Though AChE is conserved, with the same catalytic triad in each organism, protein length and numbering system vary among species. You'll have to do a little digging to figure out how to compare sequences. Using information from the Molecule of the Month article, the Protein Data Bank structural summary page for each pdb file, and the FASTA sequences for each molecule (click on "Download Files" on the right side of the PDB page and then select FASTA to access these), complete the table below. Include the names of the catalytic site residues in column 1. You might start by filling in the information you know about each protein, then align the sequences to determine the corresponding residue numbers in other species.

PDB File	Not available	1amn	1qon
Species	*Culex pipiens*		
Common name	mosquito		
Active Site residues:			
Mutant residue			

This information will be very useful as you explore AChE protein structure in other organisms and apply this information to the mosquito.

While you were gathering the background information, your colleagues were busy modeling AChE. Online computer visualizations are available at: http://mgl.scripps.edu/people/goodsell/jmol/ach/. You may also have access to physical models of AChE, based on 1QON. After exploring these models and making observations about the location of the catalytic triad and the mutation, you should be ready to write the discussion for the paper you are going to publish about your discoveries.

Be sure to address the following in your discussion:

- *The location of the catalytic triad – both linearly in the amino acid sequence and spatially in the folded protein. [Hint: use the electric eel structure for your discussion.]*
- *How the structure of the insecticide inhibitor compares with that of the substrate, and whether the insecticide acts as a competitive or non-competitive inhibitor.*
- *An analysis of the G119S mutation in terms of whether a change in size or chemical properties of the sidechain impacts:*
    - *Normal substrate binding and enzyme function.*
    - *Inhibitor binding.*
- *Based on your findings, propose a better design for a new insecticide that would function well regardless of whether G or S is present at residue 119.*