Development of a System to Rapidly Identify Molecular Interactions

Project Number: DRA 1402

A Major Qualifying Project Report

Submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

By

Xuan He

Peter Lam

Rebecca Lee

Chang Li

April 30, 2015

Approved:

Professor Dirk Albrecht
Biomedical Engineering Department
Biology & Biotechnology Department

Keywords:
1. Ligand-receptor interaction
2. Deorphanization
3.*C. elegans*

This report represents the work of WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on its website without editorial or peer review.

Authorship for Writing

Section	Primary Author
Acknowledgement	Chang Li and Rebecca Lee
Abstract	Rebecca Lee
Chapter 1: Introduction	All Members
Chapter 2: Literature Review	
2.1 Clinical Significance	Xuan He
2.2 General Background	Xuan He
2.2.1 Neurotransmitters	Chang Li
2.2.2 G-protein Coupled Receptors	Chang Li
2.2.3 *C. elegans*	Peter Lam
2.3 Current Systems	Xuan He
2.4 Previous Major Qualifying Project	Rebecca Lee
2.5 System Components	Xuan He
2.5.1 ASH Neuron	Peter Lam
2.5.2 Genetically Modified *C. elegans*	Xuan He and Chang Li
2.5.3 Behavior Recording and Worm Tracking	
2.5.4 Machine Learning	Peter Lam
2.6 Development of a System to Rapidly Identify Molecular Interactions	Xuan He
Chapter 3: Project Strategy	
3.1 Initial Client Statement	Rebecca Lee
3.2 Objectives	Xuan He
3.3 Constraints	Peter Lam

3.4 Revised Client Statement	Rebecca Lee
3.5 Project Approach	Xuan He
3.5.1 Technical	Xuan He
3.5.2 Management/Financial	Xuan He
Chapter 4: Alternative Designs	
4.1 Functions	Xuan He
4.1.1 Worm Delivery	Xuan He
4.1.2 Chemical Spotting	Xuan He
4.1.3 Behavior Identification	Xuan He
4.2 Alternative Designs	Xuan He
4.2.1 Chemical Spotting Methods	Xuan He
4.2.2 Chemical Configuration Options	Xuan He
4.2.3 Worm Strain Options	Xuan He
4.2.4 Analytical Methods	Xuan He
4.3 Preliminary Data	Peter Lam and Xuan He
4.3.1 Stamp vs. Pipetting	Peter Lam and Xuan He
4.3.2 Shape of Plate	Peter Lam and Xuan He
4.3.3 *C. elegans* Strain	Peter Lam and Xuan He
4.3.4 Manual Analysis vs. Machine Learning Analysis	Xuan He
4.3.5 Supplemental Experiments	Peter Lam
Chapter 5: Design Verification	
5.1 Final Design Decisions	Xuan He
5.1.1 Overview of System Process	Xuan He
5.2 Experimental Results	
5.2.1 Final System Experiment	Chang Li

8.1 Overview of project and Objectives	Xuan He
8.2 Limitations	Xuan He
8.3 Future Recommendations	Xuan He and Peter Lam
References	Peter Lam
Appendix	
Appendix I	Chang Li and Xuan He
Appendix II	Peter Lam
Appendix III	Peter Lam
Appendix IV	Peter Lam
Appendix V	Rebecca Lee
Appendix VI	Rebecca Lee

All team members contributed to editing each section of this report. Rebecca Lee was the final editor of this report.

Authorship for Experiments and Program Design

Section	Primary Credit
Section 1: Experiment	
1.1 Stamp vs. Pipetting	Chang Li
1.2 Plate Configuration	Chang Li
1.3 *C. elegans* Strain Profile	Xuan He
1.4 Supplemental Experiments	Chang Li
1.5 Stamp Modification	Peter Lam
1.6 *C. elegans* Strain Cross and PCR	Chang Li
1.7 Final System Validation	Peter Lam
1.8 Blind Manual Test	Chang Li and Xuan He
Section 2: Behavioral Analysis with Navin Code	
2.1 Preliminary Data	Peter Lam
2.2 Final System Data	Peter Lam
Section 3: Analytical Program	
3.1 Processing Code	Rebecca Lee
3.2 Machine Learning	Rebecca Lee

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

We would like to express our sincere thanks to Professor Dirk Albrecht for serving as our project advisor for this MQP. We would like to thank him for helping initiating this MQP, giving us helpful guidance with his knowledge and experience along the way, as well as for providing support for lab equipment and report editing.

We would love to thank Professor Patrick Flaherty for providing valuable suggestions to resolve our issues with MATLAB data processing. Thank Professor Elizabeth Ryder for providing guidance in statistical analysis. Thank Dr. Navin Pokala from Rockefeller University for providing his worm tracking code for data analysis in this MQP. Additionally, we would like express our sincere thanks to Lisa Wall, Laura Aurilio, Ross Lagoy, Kyra Burnett, Doug Reilly, and Chris Chute for their support and assistance on this project. Our appreciation to all the aforementioned individuals, this project would not be possible without your help.

Abstract

To fully understand intracellular signaling that controls bodily processes, the identification and characterization of ligand-receptor interactions is necessary. Based on this knowledge, ligand-mimetic drugs have been developed and composed 90% of the drug market in 2013. Despite the importance of ligand-receptor interactions to drug development, many receptors do not have known activating ligands. Current systems to identify these interactions are expensive and difficult to maintain. A novel system has been proposed that transfects a receptor of interest into the aversive reflex neuron in *Caenorhabditis elegans* so that the worm would display aversive behavior when interacting with an activating ligand. This Major Qualifying Project aimed to develop the necessary components to effectively run such a system. At the conclusion of the project, the team improved the existing system components, created automated scripts to identify the chemical type and ran small scale experiments to test these components. The machine learning scripts developed by the team had superior accuracy (68.8%±6.3%) in comparison to manually identifying (50.7%±8.2%) chemical type of each spot.

Chapter 1: Introduction

Intracellular signaling is critical to maintaining homeostasis and controlling processes throughout the body. In order to understand these signaling pathways, identification and characterization of ligand-receptor interactions is necessary. Ligand-mimetic drugs exploit our knowledge of ligand-receptor interactions to create medications that compose 90% of the drug market in 2013 (Oh *et al*., 2014). A common ligand-mimetic medication is Claritin. Claritin molecules compete with histamine which trigger allergic reactions by binding to histamine receptors on immune cells (Thurmond *et al*., 2008). Despite the importance of ligand-receptor interactions for the development of novel drugs, many receptors do not have known activating ligands.

An "orphan receptor" has no known binding ligands and "deorphanization" is the discovery of a receptor-ligand interaction (Caers *et al*, 2014). Current techniques for deorphanization are expensive and inefficient. Classical deorphanization is achieved by purifying ligands from biological sources then matching the extracted substances to a receptor. Due to the exhausting and slow nature of these experiments, a technique called "reverse pharmacology" has become the dominate deorphanization approach. This strategy utilizes the genetic sequence of the receptor to find corresponding ligands. One technique for doing this is by expressing the receptor into a cell line containing a receptor binding reporting system such as a fluorescent calcium indicator (Mertens, 2004). A new approach utilizes *Caenorhabditis elegans* (*C. elegans*) and genetic engineering to create a behavior-based reporting system. In *C. elegans*, the ASH neurons is known to be linked to easy to visualize aversive behavior. By creating a transgenic animal that expresses the receptor of interest on the surface of the ASH neuron, interactions with a binding ligand would trigger observable aversive behavior. Previous research by Teng *et al.* confirmed that these components can elicit the desired behavior in the worms when they interact with a binding ligand (2006), but this system was not amenable to screening of many ligands at once.

The novel system developed in this Major Qualifying Project utilizes the aversive behavior of *C. elegans*, video analysis programs and machine learning algorithms to create a system that can eventually

be used to identify the presence of active ligand-receptor interactions. The aim is for a receptor of interest to be transfected into the aversive ASH sensory neuron in *C. elegans* such that when a worm interacts with an activating ligand, the worm would display aversive behavior. To screen for novel receptor ligands, a large population of these engineered animals would explore a test plate containing an array of potential ligands and report the position of activating ligands by the position of aversive behaviors. In the future, engineered *C. elegans* strains, each expressing a receptor of interest in ASH neuron, can be generated and then tested against any type of ligands, including chemicals and pharmaceutical compounds.

Chapter 2: Literature Review

2.1 Clinical Significance

Our system was designed to aid the discovery of new receptor-ligand pairs. Receptor-ligand interactions play a significant role in cellular signaling including the regulation of cell behavior and the transmission of cellular signals. Notably, medications that target receptor-ligand interactions were estimated to compose up to 50% of the drug market in 2002 (Hruby, 2002) growing to compose about 90% of the drug market in 2013 (Oh *et al*., 2014).

The process of designing a new drug often starts by finding the proteins that are responsible for a disease. Once the target is identified, extensive investigations can be conducted regarding the protein structure, physiochemical properties and biological functions. If the 3D conformation of the target protein is well studied, one approach of developing new drug, so-called "structure-based drug design", is to design a compound that hypothetically bind to the active site of the target protein using computer modeling. However, considering the complexity and dynamics of proteins, sufficient information about 3D structure of a target protein is rarely available. It is more common to utilize a strategy known as "ligand-based drug design", in which the hypothetical active compound is designed based on a known ligand that interacts with the target protein to produce a desired biological activity (Lee *et al.*, 2011).

The ligand-based drug design depends on identified ligands that interact with the target protein; therefore, identifying the ligand-receptor matching pairs becomes a very important step in drug development. For instance, experiments conducted by Huang *et al.* found that a type of protein called ATP synthase is expressed at an abnormally high level in human breast cancer tissues. By testing several known ATP synthase inhibitors, they discovered a significant decrease in breast cancer cell proliferation with the treatment of Aurovertin B whereas the growth of the normal cells was not affected. Further investigations on the interaction between Aurovertin B and ATP synthase can greatly benefit the drug designing process to treat breast cancer, a disease that affects more than 200,000 women each year in the

U.S (Huang, 2007). Similar mechanism can be applied to a broad range of diseases in which a drug that either mimics or opposes the function of a natural ligand.

Overall, drug development breaks down into five steps: identification of ligand-protein interaction, ligand-mimetic drug design, preclinical trial, clinical trial, and FDA approval. Our project mainly focuses on the first step of the drug development process, which is the identification of a ligand-receptor in the pathway of interest. This project will design a system and procedure with previously developed system components to create a system that allows for the rapid screening of multiple ligands simultaneously.

2.2 General Background

Our project aims to create a system for identifying ligands for orphan receptors. Considering the wide range of receptors, we have decided to use the G-protein coupled receptors (GPCR) as an example class of receptors extensively targeted in the drug market. We also discuss the *C. elegans* model system and include the background information relevant to our project goal.

2.2.1 Neurocommunication

The nervous system is the control center of the body, which consists of central nervous system (CNS) and peripheral nervous system (PNS). The CNS is made up of the brain and spinal cord. The PNS are nerves outside the CNS and compose of effector and sensory nerves. Nerves primarily contribute to the transmission of signals and communication within the body that then can control bodily processes like muscle movement. (What are the parts of the nervous system, 2014). When our sense of touch is activated by touching an object, sensory neurons will activate causing a signal to be passed along to the spinal cord and brain (CNS). Once the information in the signal is processed, another signal is passed out of the CNS to motor or efferent neurons to effect a change.

In the mammalian nervous system, neurons are the basic cellular unit of the nervous system. Neurons communicate with each other electrically and chemically. A neuron includes a cell body with nucleus, a long nerve fiber called an axon, and root-like structures called dendrites (Figure 1: Neuron and

its components (Nervous Energy, 2003)Figure 1).Axons and dendrites are critical in neuron communication. Axons conduct electrical impulses called action potential from the cell body, and transmit the signal along the axon to axon terminals. At axon terminal, axons pass the signals to dendrites or to the cell body across a space or cleft called a synapse. At the synapse, electrical signals are passed either directly through gap junctions or through chemical neurotransmitters released into the synaptic cleft and detected by the postsynaptic neuron. After the dendrites receive the signals, they continue passing the signal through the cell body and axons of another neuron again.

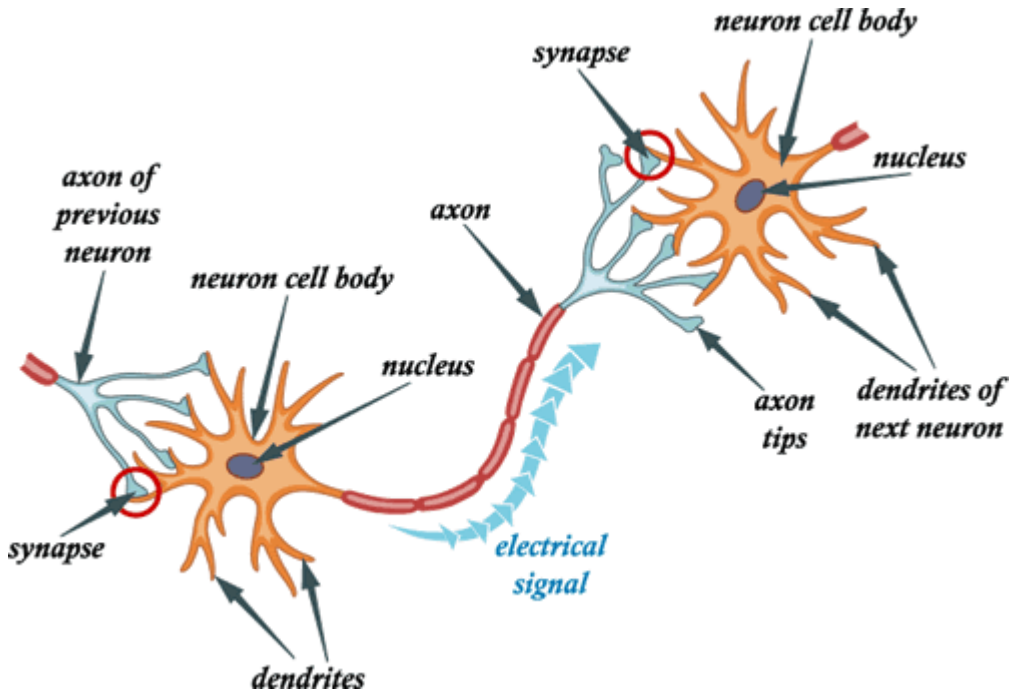

Figure 1: Neuron and its components (Nervous Energy, 2003)

At "electrical synapses", neurons can directly transmit electrical signals through gap junctions. Gap junctions form pores between presynaptic and postsynaptic cells that permit direct cell-cell transfer of small molecules and ions. The connecting cells increase the speed of synaptic transmission rapidly (Lodish *et al*, 1997). Gap junctions play an important role in electrical synapses. The signaling is instantaneous, and some are bidirectional (Figure 2).

At “chemical synapses”, indirect chemical communication occurs between cells. The chemical synapses are responsible to exchange of excitatory, inhibitory, and complex biochemical information (Hyman, 2005). When electrochemical signals are transmitted to the pre-synaptic axon terminal, neurotransmitters in membrane-enclosed synaptic vesicles are released into the synaptic cleft. The postsynaptic cell membrane contains neurotransmitter receptors and thousands of transmitter-gated ion channels. The neurotransmitters quickly bind to the transmitter-gated ion channels which open and induce changes in the electrical potential of the post-synaptic cell, triggering another action potential in neurons or contraction of skeletal muscle (Figure 2; Alberts *et al*, 2002). Since the neurotransmitters released based on the synaptic vesicles, the speed of chemical synapse through ion channels is one millisecond slower than the electrical synapses.

Chemical synapses can be excitatory or inhibitory depending on the ion channel type; for example opening Cl^- or K^+ channel would ‘hyperpolarize’ or inhibit the postsynaptic neurons, or whereas opening a Na^+ or Ca^{++} channel would ‘depolarize’ or excite the postsynaptic neuron.

Another form of chemical communication is via neuropeptides, which are released outside of synapses. Neuropeptides are short sequences of amino acids used to communicate between neurons in the brain (Li *et al*, 2008). The mammalian brain contains at least 100 different neuropeptides involved in various behaviors such as social behaviors, recognitions, and emotion (Ludwig, 2012), and any mutations in neuropeptides could cause severe disorders. For example, mutations in the neuropeptide somatostatin are associated with an increased risk for Alzheimer’s disease (Gottfries, 1995). However, not many neuropeptides and receptor pathways are identified or fully characterized.

Some neurotransmitters and neuropeptide receptors are G protein-coupled receptors. Activating G protein-coupled receptors can activate intracellular pathways, release calcium from intracellular stores, trigger phosphorylation of different ion channels, alter gene expression, and alter the functions and structures of neurons (Hyman, 2005).

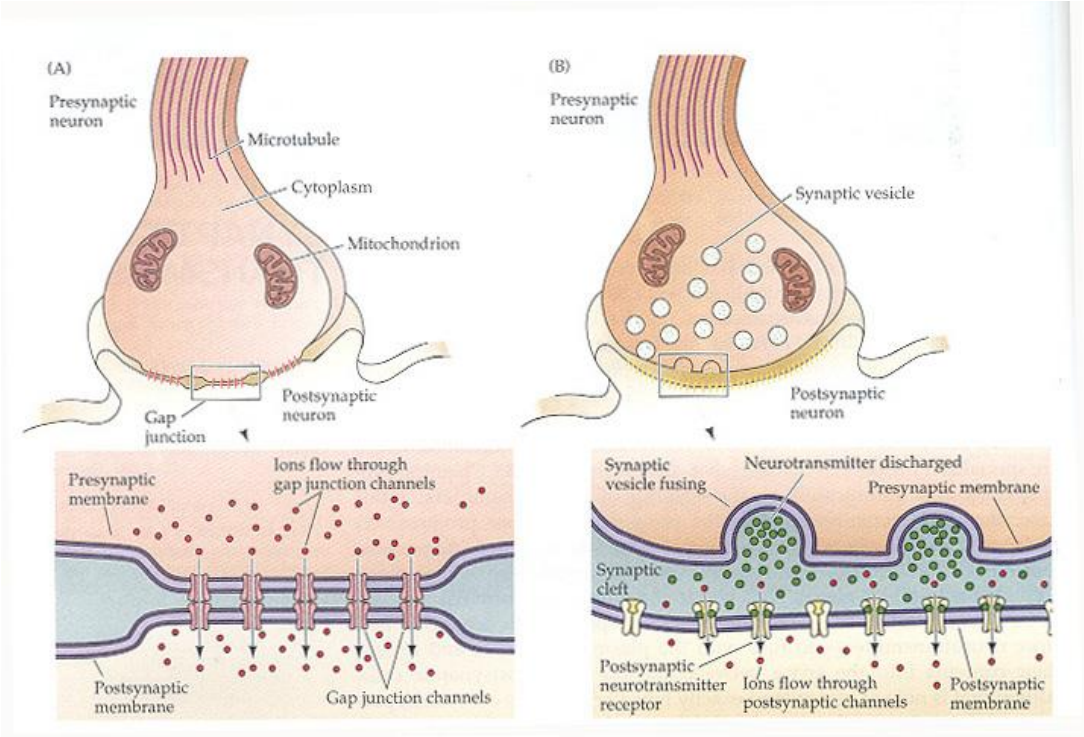

Figure 2: Gap junctions (A) versus ion channels (B) (Synaptic transmission, 2013)

2.2.2 G-protein Coupled Receptors

G protein-coupled receptors (GPCR) are a large family of membrane-bound proteins that transduce extracellular information into the cell; for example, GPCRs can interact with peptides, ions, organic molecules, photons, and various subatomic particles (Kobilka, 2007). The human genome is believed to contain over 800 unique GPCRs. However, most of these receptors have unknown physiological functions (Kobilka, 2007). Each GPCR consists of seven transmembrane segments connecting with three intracellular and three extracellular loops (Schöneberg *et al*, 2004). Figure 3shows the mechanism of action of a GPCR. These ligands binds to the extracellular portion of GPCR causing a confrontational change of GPCR that will activate the intracellular G-protein subunits (known as α, β, and γ units) (Kobilka, 2007;Schöneberg *et al*, 2004). The G-protein subunits are catalyzed by GDP-GTP exchange inside the cell, trigger further cellular signal transduction (**Error! Reference source not found.**;

Schöneberg *et al*, 2004). GPCRs are involved in a variety of transduction pathways including smell, taste, neurological, cardiovascular, endocrine, and reproductive mechanisms (Katritch *et al*, 2013).

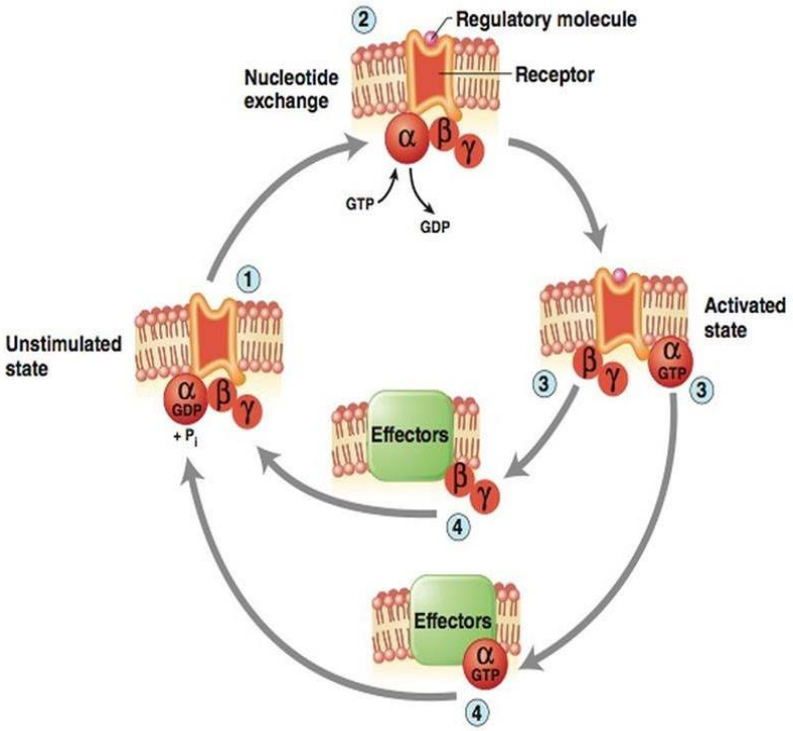

Figure 3: Activation of GPCR and associated pathways (Fox, 2011)

Due to its significant role in activating signal pathways, mutations in GPCRs are linked to more than 30 monogenetic diseases (Schöneberg *et al*, 2004). For example, a GPCR receptor called histamine H receptor is associated with allergic reactions (Dunford *et al*, 2006). Mutations in the GPR37/PaelR receptor are associated with autism spectrum disorder and Parkinson's disease (Fujita-Jimbo *et al*, 2012). Due to the significant role of GPCRs play in many ailments, these receptors are targeted by 60% of current prescribed drugs on the market (Schöneberg *et al*, 2004).

To demonstrate how GPCR targeting medications work, consider the common allergy medicine known as Claritin (loratadine). This medication works by blocking a ligand binding to a GPCR. This pathway begins when an allergen or antigen triggers mast cells which then produce histamine molecules. Normally, histamine binds to histamine GPCRs located on immune cells which leads to an allergic

reaction. Claritin molecules prevent the interaction between histamines and histamine GPCR by binding to the histamine GPCR; thus, no allergic reaction occurs (Figure 4; Thurmond *et al.* 2008).

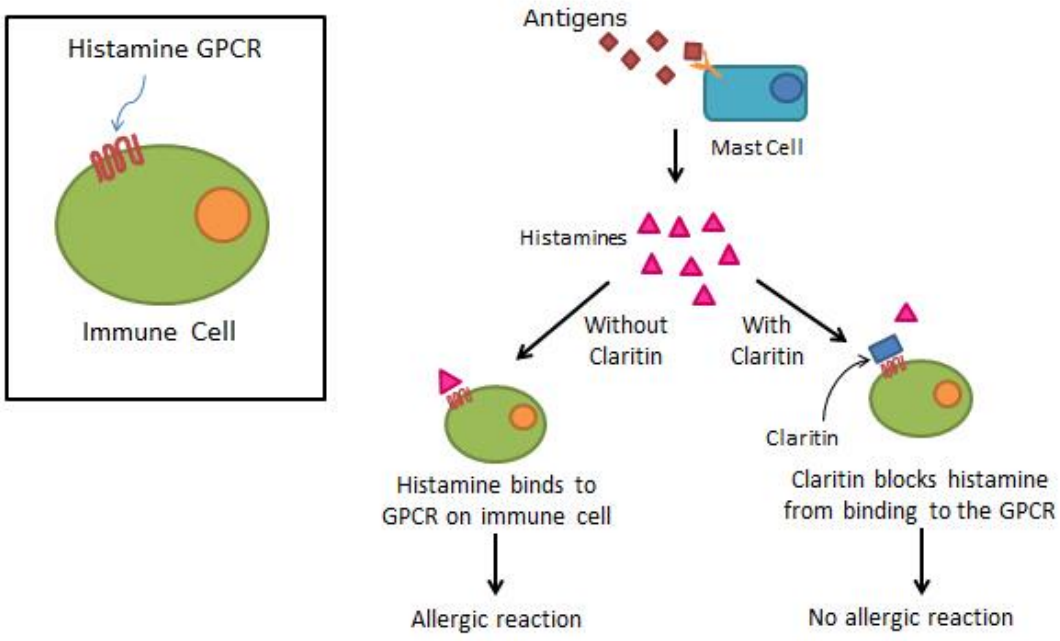

Figure 4: Allergic Reaction pathway with histamine and Claritin

2.2.3 *C. elegans*

Caenorhabditis elegans (*C. elegans*) is a multicellular nematode species found worldwide in the soil that are roughly one millimeter in length at the adult stage. *C. elegans* are easy to grow and maintain in the laboratory. They are fed with *Escherichia coli* bacteria and can grow at temperatures ranging from 12º C to 26º C (Stiernagle, 2006). Their short life cycle of 3 days at 20ºC consists of four larval stages (L1-L4) before reaching its adult stage. *C. elegans* exist as hermaphrodites and males. Hermaphrodites can self-fertilize to produce progeny that are genetically identical to the parent or they can mate with males to produce crossed progeny. This characteristic of *C. elegans* allows for high throughput studies in genetics. In 1998, *C. elegans* became the first multicellular organism with its entire genome sequenced. Comparison of the genome of the nematode with vertebrates showed similarities in the nervous systems (Frooninckx, 2012). *C. elegans* contain homologs of genes in mammalian cells that code for different

proteins, enzymes, and diseases. As a result of its characteristics, *C. elegans* is used as a model organism for research in neurobiology (McKay, 2003).

In comparison to other animal models, *C. elegans* provide a simple system for the study of behavior and human conditions (Frooninckx, 2012). Researchers have found that about 50% of human disease genes are found in *C. elegans*. The worm can model different diseases such as muscular dystrophy and kidney disease (Culetto, 2000).Because *C. elegans* is transparent, researchers can easily study the development and anatomy of the animal throughout its life cycle. The transparency of the animal also allows for fluorescent microscopy to study neuronal activity. The sensory neurons in *C. elegans* express multiple GPCRs in each neuron to respond to different stimuli in the environment. The animal provides a good *in vivo* model to study mammalian GPCR-ligand interactions as opposed to *in vitro* studies. *C. elegans* can be easily mutated to express different receptors or suppress certain behaviors for different kinds of studies (Teng, 2006).

2.3 Current Systems

GPCRs play a critical role in physiological activities making them an excellent target for therapeutic drugs. However, many GPCRs do not have known binding ligands and are designated as orphan GPCRs. Therefore, the process of identifying a ligand for a specific orphan GPCR is referred as "deorphanization" (Caers *et al.*, 2014).

Classical deorphanization is achieved by purifying ligands from biological sources then matching the extracted substances to a receptor. Due to the exhausting and slow nature of these experiments, a technique called reverse pharmacology has become the dominate deorphanization approach. Reverse pharmacology takes the advantage of the available GPCR genome database and uses the known sequences of GPCRs to ultimately detect interacting ligands. Typically, this deorphanization approach expresses the GPCR of choice in a cell-based system. The expression system is designed such that a signaling cascade and visual reporter is activated if a ligand interacts with the expressed GPCR. The triggered signaling cascade allows experimenters to visualize the molecular interaction between an active

ligand and the expressed GPCR. The schematic presentation of reverse pharmacology is shown in Figure 5 (Mertens *et al.*, 2004).

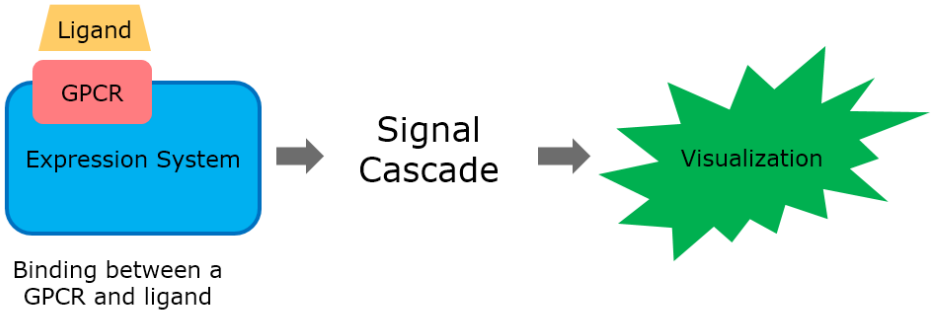

Figure 5: Illustration of the concept of reverse pharmacology

A common form of cell-based deorphanization is known as fluorescence-based calcium mobilization assay. Caers *et al.* utilized human embryonic kidney cell line (HEK293T) as the expression system and transfected the cells with the targeted GPCR sequence. The cells were also preloaded with a calcium sensitive dye (Fluo-4); consequently, the ligand-receptor interaction would be reported in terms of the level of calcium concentration. When a tested ligand interacted with the targeted GPCR, signaling events resulted in the release of calcium ions into the cytoplasm. The preloaded Fluo-4 sensed the released calcium and produced fluorescent light. The fluorescent light was measured in terms of relative fluorescent unit (RFU), which was correlated to the level of GPCR activation. To ensure the positive result was indeed resulted from interacting with the targeted GPCR instead of non-selective receptors in the cells, a specific GPCR subunit (Gα16) was also transfected before testing to ensure accurate results (2014). The general process of this assay is summarized in the schematic below (Figure 6).

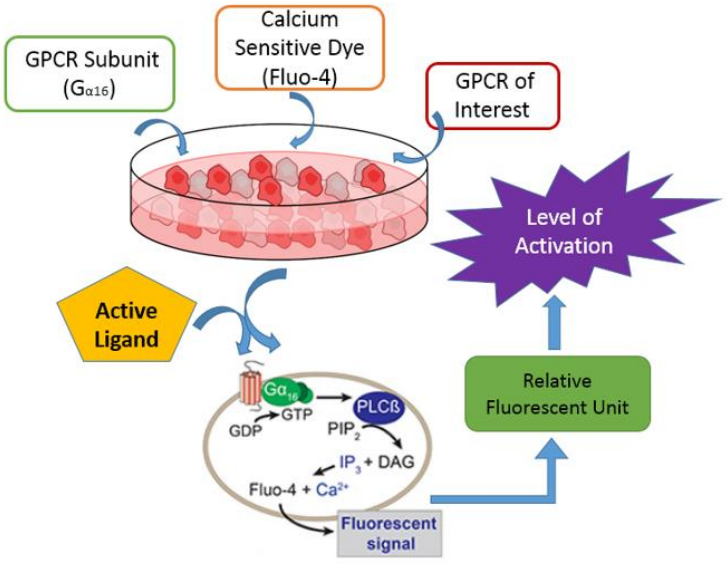

Figure 6: General process of fluorescence-based calcium mobilization assay

An alternative deorphanization method has been reported that utilizes *C. elegans* behavior to indicate positive interaction between a ligand and a receptor of interest. Teng *et al.* expressed a clinically-relevant GPCR of interest, the human cytokine receptor CCR5, in the ASH neurons of *C. elegans*. The ASH neurons was chosen because a portion of this neuron is exposed to the environment allowing interactions between given ligands and expressed receptor. In addition, being a nociceptive neuron, ASH neurons is responsible for triggering the avoidance behavior when a worm encounters a repellent. As a result, activated ASH neurons trigger an aversive response of the worm. Therefore, a transfected worm will display a clear reversal behavior when confronted with a ligand for CCR5, the cytokine MIP-1a. The experimental design in Teng *et al.* is shown below (Figure 7). About 50 transfected animals were placed behind a line of soluble ligand on the culture plate. Across from the ligand line, a line of attractant was placed to encourage the worms to move in the direction of where the ligand was located. A positive result in this case was indicated by the worms not moving across the ligand line. The level of interaction was reflected by the measure of avoidance index, which was the number of the worms behind the ligand boundary divided by the total number of worms applied (Teng, 2006).

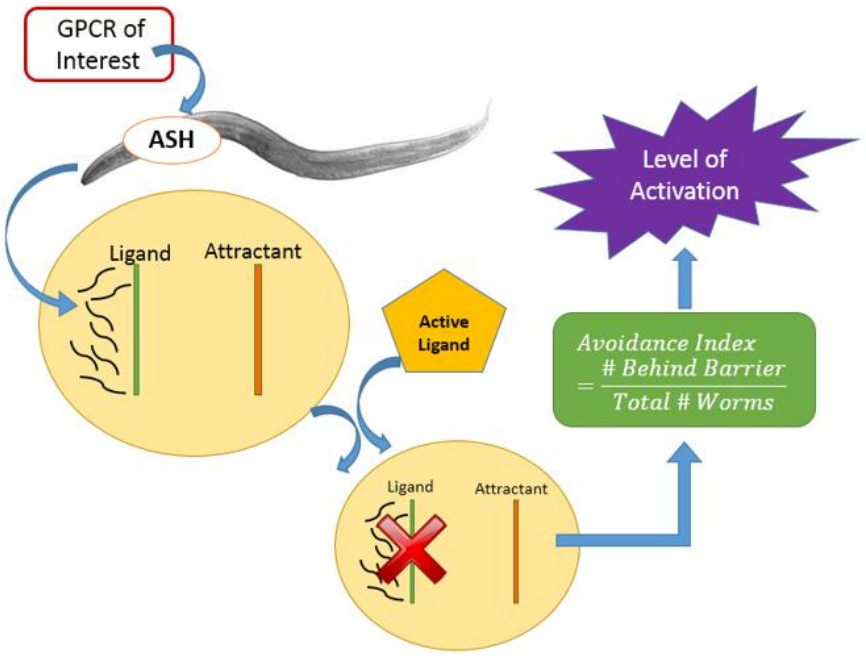

Figure 7: General Process of animal behavior assay utilizing *C. elegans*

Different deorphanization systems have their own advantages and limitations. Cell-based systems like the fluorescence-based calcium mobilization assay are easily quantifiable and different ligands or various concentrations of the same ligand can be tested in series. In contrast, the animal behavior assay results in a graded outcome utilizing confidence scores for chemical types. Despite the cell-based system being easier to quantify, the signaling pathways at cellular level are difficult to predict and control. The artificial system requires G-proteins and fluorescent reporter. It also requires very expensive instruments to detect the fluorescent dye for receptor-ligand binding. Although co-transfecting the subunit of GPCR might increase the selectivity of the results, any other non-specific interactions are hard to identify. Furthermore, not all GPCR use the same G-proteins or Ca^{++} secondary messengers for signal transmission, the system requires lot of different reporters to capture all GPCRs. Comparing to the cell-based system, the neuronal circuit of *C. elegans* is well-understood, and the behavioral system is simpler to modulate and use. In addition, the animal behavior assay allows direct visualization of a positive interaction without the need for incorporating a secondary reporting system, which also makes the system user-friendly. However, the biggest drawback of the currently reported animal behavior deorphanization

system is that only binary screening of ligand is allowed, meaning only one ligand is tested at a time. If we consider that there are estimated 800 GPCRs in the human genome and hundreds of possible binding ligands, there is a large number of possible pairs to be tested (Zhang, 2006).Ideally, the system we are designing a procedure for is based on the current animal behavior assay using *C. elegans* as the animal model and the aversive behavior as the indicator of a positive interaction but allows multiple ligand identifications at once.

2.4 Previous Major Qualifying Project

An initial investigation of a system using *C. elegans* for deorphanization was completed by a previous Major Qualifying Project, DRA AAHO, titled "Behavior Analysis for Rapid Screening of Molecular Interactions". There were two goals for the previous MQP: (1) to design a component to facilitate the transfer of chemicals and (2) to develop experimental procedures for animal behavioral screening. The team identified methods for transferring worms to the observation plate, created a PDMS stamp that could transfer 96 stimulus spots from a well plate to the observation plate, experimented with the video capture methods and identified a number of challenges working with the system (Haughn, 2014).

To develop the procedures and experiments for our project, we examined the testing and results of this previous MQP. The previous team investigated the diffusion of chemicals into the plate, because ligand chemicals may diffuse into the gel leaving less on the surface for the worms to interact with. It was determined that after an hour all of the chemicals were diffused into the gel. The team also examined the system by exposing worms to three chemicals: isoamyl alcohol (positive control), glycerol (negative control) and S. basal (neutral buffer control). They obtained clear video of the worms and found that it was most useful to analyze the frequency of different behaviors in the vicinity of the stimuli spot (Haughn, 2014).

The team identified limitations in stimulus deliver, video recording and the analysis programs. The PDMS stamp delivered inconsistent concentrations of chemical to the observation plate. Lacking

control over the concentrations meant that there could be too little of a particular chemical to stimulate a clear response in the worms. Another difficulty was that the resolutions of the video recording were not appropriate for the full plates with 96 chemical spots, preventing complete recordings of a full 96-well plate. Moreover, the video recorded generated large amounts of data resulting in long analysis times. The MATLAB program used for analysis needed to be adapted to our system and analysis needs. Finally, the previous team suggested that to get clear consistent results through each test, there needs to be a full procedure for preparing and using the system (Haughn, 2014).

The limitations found by the previous MQP will provide a guide and starting point for our experiments and procedure. We are now aware that we need to examine how the concentration of chemicals can affect the animal behavior and improve the recording and analysis programs.

2.5 System Components

Our system utilized many specific biological processes, specialized computational analysis methods and particular lineages of *C. elegans*. In this section, there is background information on each of the different components of the system. The nociceptive ASH sensory neurons in *C. elegans* is important to our system. We explain the modifications and phenotypes of *C. elegans* strains available for our system. Also, we include information on the behavior tracking system and background on machine learning that we used to analyze our data.

2.5.1 ASH Neurons

The nervous system of an adult *C. elegans* hermaphrodite consists of 302 neurons and is divided into the large somatic nervous system with 282 neurons and the small pharyngeal nervous system with 20 neurons (Sassa, 2013). Interneurons act as information processors, receiving input signals from the sensory neurons and output them to other neurons to generate different behaviors. *C. elegans* neurons communicate through chemical synapses, gap junctions and neuropeptides to control complex behaviors such as well as locomotion, feeding, foraging, mating and defecation (Altun, 2011).

The ASH neurons are a pair of bilateral sensory neurons that are found in the head. These neurons play a major role in mediating avoidance behavior in response to different stimuli in the environment, which is called nociception (Hilliard, 2005). The ASH neurons are polymodal and respond to different types of stimuli such as mechanical stress, osmotic shock, noxious heat, and chemical repellents (Esposito, 2010). The neurons' cell bodies are located in the nerve ring of *C. elegans* and extend dendrites to the mouth of the animal, where they are exposed to the outside environment. The neurons are able to detect noxious stimuli by having cilia exposed outside the body. For harsh mechanical stimuli such as a nose touch, the ASH neurons work together with the FLP neuron to sense the stimulus (Altun 2011). ASH neurons synapse onto the AVA interneurons, which will ultimately synapse to motor neurons downstream to cause an aversive behavior of the *C. elegans*. The aversive behavior is characterized by a reverse movement of the worm (Guo, 2009). Sensory behaviors mediated by the ASH neurons generate distinct glutamate release to target command interneurons that allow for separate behavioral responses in the OSM-9 and OCR-2 TRPV channels. A mutation in the glutamate vesicular transporter *eat-4* or mutations in the *osm-9* and *ocr-2*ion channels result in defects in aversive responses (Altun, 2011).

2.5.2 Genetically Modified *C. elegans*

C. elegans expresses GPCRs for chemosensation to detect food and surrounding environmental conditions, which makes it an ideal system to study GPCR-neuropeptide interaction *in vivo* (Teng, 2006). There are approximate 500-1000 GPCRs predicted to be produced by the chemosensory neurons (Bargmann, 2006), while there are only 23 identified GPCR-neuropeptide pairs (Frooninckx, 2012). The expression of GPCRs in *C. elegans* is another characteristic that makes it a good model system for our uses. Microinjection is a method to create transgenic strain of *C. elegans* by introducing DNA into worm. In general, ≥1μL of DNA injection mixture is loaded into the microinjection needle. Then, a C. *elegans* in its L4 stage is transferred and adhered to the microinjection pad with sterile technique free of bacterial food. The needle is positioned and the DNA mixture is injected to the center of the cytoplasmic core of the worm. After injection, the worm could be recovered to feeding plate. In order to create transgenic

worms more efficient, large extrachromosomal arrays is usually applied. The extrachromosomal arrays contain multiple copies of co-injected DNAs that are inheritable in worms (Evans, 2006). With this method, we could express the receptor of interest via an ASH-specific promoter like *sra-6*, and create a selectively-sensitive line. To make the *C. elegans* more useful for the aversive behavior tracking, a genetically modified lineage of *C. elegans* needs to be developed.

Since the system is based on animal behavior, it is key for the worm to selectively express aversive behavior. Although the receptor of interest will be selectively expressed in the nociceptive ASH neuron, wild type animals can also be distracted by senses other than the active ligands such as olfaction, taste, and thermosensation. Consequently, using sensory null mutant animals should increase the selectivity of the assay by eliminating possible distracting sensations. A promising candidate of a sensory null mutant *C. elegans* strain is *eat-4; egl-3; sra-6 (ASH) p::eat-4*, referred as the selectively responsive line. The gene *eat-4* encodes for the glutamate transporter which controls a variety of animal behaviors including feeding, foraging and chemotaxis (Lee *et al*., 1999). The gene *egl-3* encodes for the proteins that regulate peptide secretion (Cai *et al*., 2001). By knocking out the genes affect glutamate and neuropeptides, this rescued line is theoretically non-responsive to food or attractive stimuli. However, to ensure the normal aversive response, the gene *eat-4* is knocked out everywhere in the worm body but returned to the ASH neurons via microinjection. Besides the rescued line, another alternative *C. elegans* strain would be *him-5*. The gene *him-5* is responsible for regulating X chromosome pairing; therefore, a mutation to this gene results in an increase in males in self-fertile populations (Hodgkin *et al*., 1979). It has been suggested by our client that the male *C. elegans* might possess higher speed than the wild type but similar in animal behaviors. More testing is needed to choose the optimal animal model that meets the functional requirement of this project.

2.5.3 Behavior Recording and Worm Tracking

Behavioral assays are used alongside with video recording set-ups and tracking systems to obtain quantifiable results. Tracking systems measure different parameters from the recorded video to define

different behaviors from the *C. elegans*. Some of these behaviors include reversals, omega turns, pauses, and pirouettes (Albrecht, 2011). One type of software that exists relies on using the centroid position of the worm in the video. The centroid based trackers can track worms for long periods of time but it cannot distinguish between the different movements of the *C. elegans* such as forward movement and reversals and work at low magnifications (Geng, 2004). The more favored type of tracking system involves generating a morphological skeleton of the worm by using a skeletonizing algorithm. The basic steps to generate a skeleton of the *C. elegans* with the software involves applying a threshold to the gray-scale image from the video. Next, any holes or gaps in the image is filled. Finally, the skeletonizing algorithm is applied. This type of tracker can track worms at high magnifications and distinguish between the different types of movements by the nematodes (Huang, 2008). Advancements in tracking systems has allowed for tracking of multiple animals at once and utilizes MATLAB as the platform for video recording and worm tracking (Swierczek, 2011).

2.5.4 Machine Learning

Machine learning is a method used to predict future outcomes based on past outcomes. The process involves feeding a system with data from previous events. Once the system has the data, it will use an algorithm to analyze the data and create a model. The model can then be used to analyze new data and predict the possible outcome. Machine learning is classified into supervised and unsupervised learning. Supervised learning involves using a data set that has already been labeled and unsupervised learning is when an algorithm is used to find a hidden pattern in a data set (Bell, 2015).

Discriminant analysis is a type of machine learning algorithm that can be used to predict an outcome. This is a classification method that assumes data generated from different classes are based on different Gaussian distributions (Fisher, 1936). Discriminant analysis is similar to a regression where a linear equation is determined to predict which group the data belongs to. The equation for this method is as follows:

$$D = v_1X_1 + v_2X_2 + v_3X_3 + \cdots + v_iX_i + a$$

$$D = Discriminate\ Function$$

$$v = Discriminant\ Coefficient\ or\ Variable\ Weight$$

$$X = Respondent\ Score\ for\ Variable$$

$$a = constant$$

$$i = Number\ of\ Predictor\ Values$$

Equation 1: Discriminate Analysis Equation

Larger weights will be assigned to good predictors in the equation. Discriminant analysis require at least two or more different groups and groups should be defined before collecting data. When there is two different groups, a Gaussian distribution is generated for each group with the data collected and is compared to see whether or not they can be discriminated from each other. There is a good discrimination when there is little overlap of the two Gaussian distributions (Agresti, 1996).

Classification tree is another type of machine learning that generates a model to predict and classify a data point. This machine learning technique involves using predictors generated from a set of training data to classify different points. The method starts at the top of a classification tree, where the value of a predictor that was generated from the data is evaluated. If the value of the predictor is below a certain value the tree will classify it into the first group. However, if the predictor is greater than the value, then the tree will move down and a second predictor is evaluated. This process goes on until the data is placed into a group. This machine learning method can be compared to a true or false test, where the data is tested for each variable until it falls into a group (Rokach, 2008).

2.6 Development of a System to Rapidly Identify Molecular Interactions

For our project, a variation of the behavioral assay of *C. elegans* was utilized that can address limitations from past studies. The goal was to use the system to screen for multiple different ligands at the

same time via aversive behavior of *C. elegans*. Previous studies only tested one ligand at a time using the behavioral assay model. The use of a system that can screen for many ligands at a time will be beneficial in accelerating the progress of deorphanizing the different GPCRs and potentially other classes of receptors. Overall, the three design components of our project included selecting (1) the worm strains, (2) the chemical spotting method and (3) the analytical tool and algorithm for automatically detecting ligands that elicited aversive behavioral responses. A clear procedure was further developed to use the system that generated satisfactory results.

Chapter 3: Project Strategy

3.1 Initial Client Statement

Our client, Dirk Albrecht, has an interest in neuropeptides and the physiological pathways that they act upon. However, the process of fully characterizing neuropeptide pathways is labor intensive and time consuming. Our project aims to streamline the identification of any receptor-ligand interactions which is critical to further understanding a neuropeptide pathway. The initial client statement is as follows,

> Design, implement and characterize a system to identify multiple receptor-ligand interactions in parallel by behavioral analysis of genetically-engineered *C. elegans* nematodes.

This statement envisions a novel system that utilizes *C. elegans* to rapidly discover receptor-ligand interactions. Once fully realized, the system would test up to 96 unique ligands in each test and use computational analysis to determine whether there was a receptor-ligand interaction.

3.2 Objectives

Based on the initial client statement, the project was broken down into a series of objectives. These objectives were then categorized into three main groupings. A pairwise comparison chart was then constructed and then completed in consultation with our client, Professor Albrecht (Figure 8: Pairwise comparison of primary objectives

8). The project objectives were ranked, from highest to lowest, as (1) reliable, (2) easy to use and (3) rapid.

	Ease of Use	Rapid	Reliable	Total
Ease of Use		**1**	**0**	**1**
Rapid	**0**		**0**	**0**
Reliable	**1**	**1**		**2**

Figure 8: Pairwise comparison of primary objectives

Upon further discussion, the three primary objectives were broken down into secondary objectives. For reliability, the secondary objectives were conclusive results and reproducible results. While ease of use did not have further secondary objectives, rapid was further divided into time and whether the device can screen multiple ligands in a single test. These breakdowns are fully visualized in the following objective tree (**Error! Reference source not found.** 9).

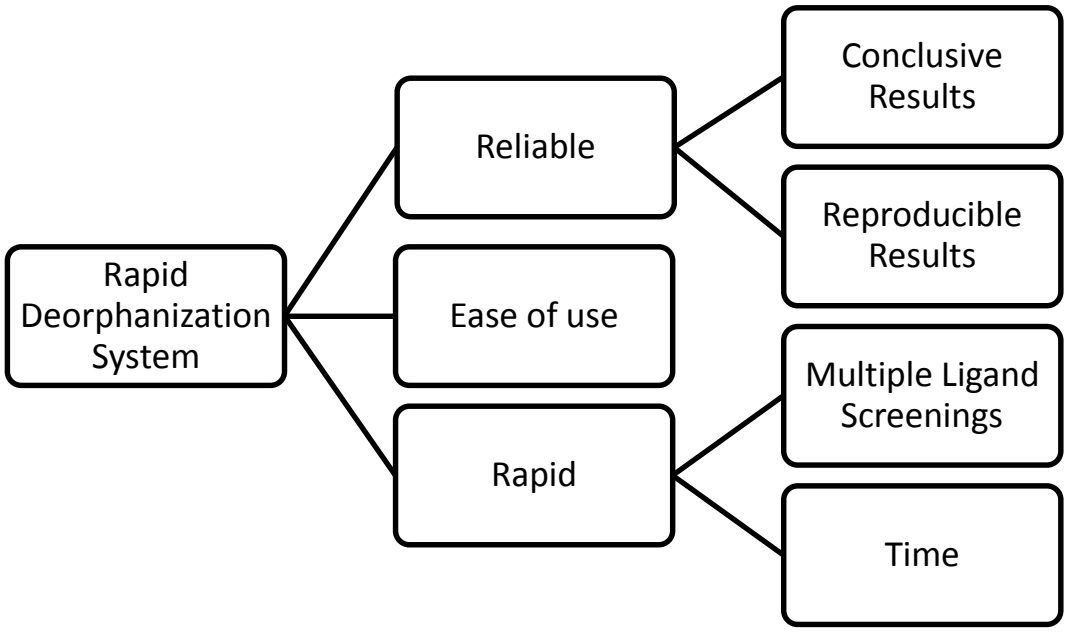

Figure 9: Objective Tree for the system being designed

Reliable is related to the results that the system produces. For the system, the results should clearly differentiate between a negative or positive interaction between a ligand and the receptor of interest. This objective is particularly important because we want data coming from the system to be easy to interpret and understand. This characteristic we wanted for the system was termed "conclusive results". Once, the system has known thresholds for different results the next concern would be whether those results are repeatable in different conditions. This secondary objective was coined "reproducible results". There is no point in being able to reproduce unclear results, which is why conclusive results are more important than reproducible results. At the conclusion of this project, the system should establish thresholds for different results and prove that these thresholds are reproducible.

Following reliability, the next objective is ease of use when compared to state of the art systems. Current cell-based systems utilize complex fluorescent reporting systems that can be unpredictable due to interference from other cellular processes. Furthermore, these cell lines need to be transfected before each test sequence and cannot be stored for later use. In contract, engineered *C. elegans* lines expressing

receptors of interest can be propagated easily and stored frozen for decades. The project should result in a system that requires less maintenance and is less complex, but produces the same quality of results.

Rapid was ranked lowest of our primary objectives. A rapid system needs to balance between testing multiple ligands while also having a short data collection period. Testing multiple ligands is important because peptide libraries are extensive; meaning that testing multiple ligands at once will allow faster results for a full library**.** In addition, there needs to be considerations for how long it takes to prepare, execute then analyze a single test. Our client is willing to have a system that is comparable in preparation and execution time in comparison to current systems, but may take longer to analyze. Basically, the final system can take a little longer than current systems as long as it makes it up by testing several ligands simultaneously.

3.3 Constraints

There are a number of constraints to guide the development of the optimal system for deorphanizing GPCRs. The first constraint is that the system has to test multiple ligands simultaneously. The current methods to deorphanize GPCRs via behavior can only test a single ligand at a time. Our system will speed up this process by being testing a large number of different ligands at once against a single GPCR. The second constraint is that the system cost. Automated systems have already been developed for deorphanizing GPCRs using the cell-based model (Caers, 2014). However, the system itself is expensive and the costs start at around $15,000. The system in development is designed to be $1000 to set up with each test costing about $100.

3.4 Revised Client Statement

With discussion with our advisor, our consideration of the project attributes and our research into the project, we created a revised client statement that better reflected the overall project aims and deliverables upon the completion of the project. The revised client statement is as follows:

> There is a need for an inexpensive, rapid and simpler to use deorphanization system. Such a system will aid research in understanding complex cellular signaling pathways that will ultimately lead to the identification of pathways for therapeutic targeting. Our project is to complete the development of a system that utilizes the behavior of *C. elegans* to find receptor-ligand interactions. At least three ligands, and up to 96, should be tested by the system in each test. At a minimum, the test and data processing should take no more than 12 hours in total to complete, and costs per test should be less than $100. The assay should report the confidence level of any determination of a positive or negative interaction between a test ligand and the receptor of interest.

This revised client statement clearly addresses the outcome of the project which is a procedure that seamlessly links together the different components that have been built. At the end of this project, the system should be fully functional for delivering clear results on receptor-ligand interactions.

3.5 Project Approach

3.5.1 Technical

For the technical approach, our project was broken up to two sections: proof of concept and system verification. The first part of the project was to generate design alternatives and make optimal design decisions for each of the system components. The second part of the project was to verify our design decisions and make further recommendations.

To compose a working system, our team decided to generate a list of design alternatives and eliminate variables by conducting a series of preliminary experiments. Based on the defined objectives and constraints, we envisioned our system to have four major components: (1) worm strain, (2) chemical spotting, (3) behavior recording and (4) data analysis. Each component suggested multiple variables that became the design alternatives. Due to the lack of literature evidence for such research, we decided to select the optimal system primarily based on preliminary experiments. Our preliminary experiments

included selecting (1) the worm strain, (2) chemical spotting device and plate configuration and (3) analytical tool.

First, the optimal worm strain should exhibit high mobility to explore all ligand spots on a test plate, and substantial response to the ASH neurons but not to other non-selective stimuli to report only positive ligand-receptor interactions. Therefore, we tested the speed and ability to activate ASH neurons by aversive stimuli for several candidate worm strains.

Second, for the chemical spotting device, we tested efficiency (the most important parameter), uniformity and volume of chemical delivery. We also looked into different plate configuration to see whether it would make a difference in the worm distribution.

Third, we needed an effective analytical tool to accurately identify and locate the targeted behaviors in a large testing array. The existing worm tracking program used in our advisor's laboratory identifies each animal's behavior and position over time, but only displays data as behavior type frequency over time (spatial information is not reported). However, we need the analytical tool to identify behaviors spatially to identify which ligand spots elicit aversive behaviors that indicate a positive receptor interaction. All of these experiments used an aversive chemical and a neutral one. For the aversive spots, we decided to use a commonly used aversive stimulus glycerol to stimulate reversals due to high osmolality. The control chemical was a salt buffer called S. basal. Overall, from the preliminary experiments, we decided on a worm strain, chemical spotting method, and analytical tool which could give us quantified results with a level of confidence.

Having the system setup composed and tested in the first part of the project, the ideal way of validating this system would be to test with a known interacting pair of receptor-neuropeptide. To do so, a candidate neuropeptide receptor needs to be expressed at ASH neurons of a selected *C. elegans* strain and the choice of testing ligand needs to be a library of purified or bacterially-derived neuropeptide. The negative control would be a ligand that is known not to interact with the expressed GPCR. Our system would be considered successful if it could separate the active ligands from the non-active ones at a tolerable error level. However, since the accessibility of the GPCR and its known ligands was very

limited and considering the time needed to develop the engineered worm strain, we decided to run our final verification experiment with glycerol as a control stimulus that would activate the ASH neurons just as a positive ligand-receptor interaction would. If our system could reliably identify glycerol spots from a large mixed chemical array, the primary goal of this project would be achieved.

Consequently, if our system is further verified with known interacting receptor-ligand pairs, we envision it can be used to identify new active neuropeptide to a specific GPCR, which would ultimately be beneficial to ligand-based drug discovery.

3.5.2 Management/Financial

In terms of task management, the project was broken up by four academic terms and the tasks to be done for each term were summarized in Figure 10. For each term, the team developed a Gantt chart to track the timeline of each task. Throughout each term, the team would have weekly meeting with our client to update project progress and discuss unresolved issues. In terms of financial part of the project, the major laboratory resources would be supplied by our client. Any additional costs would be covered by our MQP budget, which was $156 per team member.

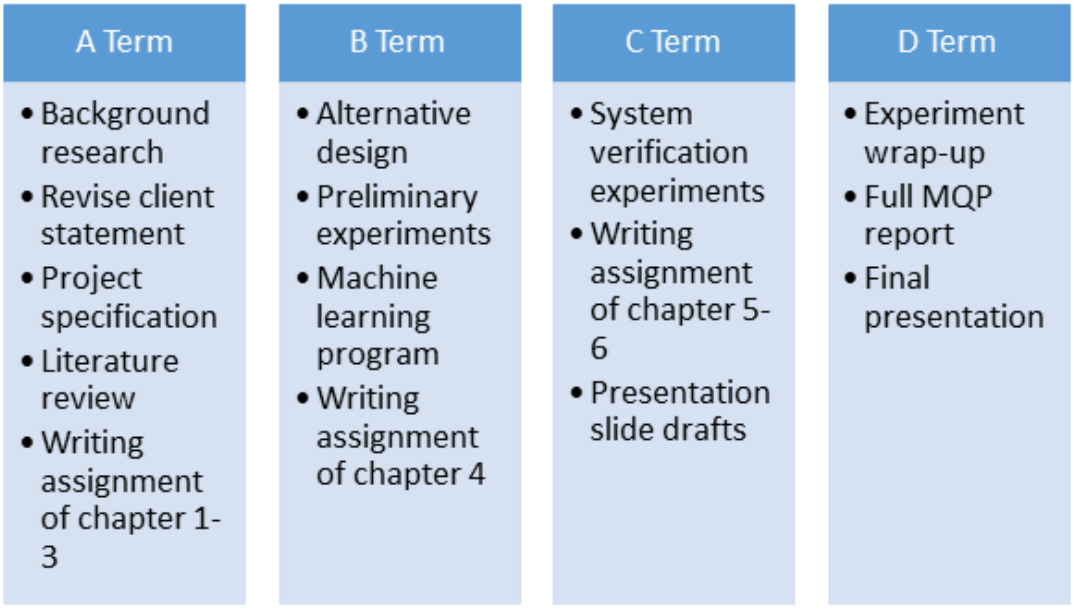

Figure 10: Overall project plan

Chapter 4: Alternative Designs

4.1 Functions

The overall goal of our project was to design a novel system that utilizes *C. elegans* to rapidly identify receptor-ligand interactions. Ultimately, this system can used to speed the development ligand-mimetic drugs. The primary objectives for the system were for the system to be reliable, easy to use and rapid. To fulfill each objective, our design must perform a number of functions listed in Table1. To be reliable, the system needed to successfully and consistently identify ligand-receptor interactions via aversive behavior of *C. elegans*. Specifically, the system needed to clearly differentiate between negative and positive interactions with reasonable level of error tolerance. In addition, the system must identify where the interactions occur in order to identify the active ligands. For the system to be easy to use and rapid, the experimental procedure of our system must be simpler to follow and less time-consuming compared to the current state-of-art system.

In order to design a system that satisfied our project goal, we generated a number of design alternatives by making a morphology chart (Table 1). We tested these alternatives within each system component and together generated an optimal system. Some variables were pre-eliminated because of difficulties with manufacturing while others were subject to experimental testing. More detailed description of the design alternatives are in the following subsections.

Functions	Requirements	Means		
Deliver worms	Need to deliver large number of worms	Picking	____	____
	Able to selectively respond to ASH neuron	Wildtype hermaphrodites	Wildtype males	Selectively responsive hermaphrodites
	Able to quickly spread out and move around			
Spot ligands	Must spot multiple ligands efficiently	Stamping soluble ligands with PDMS stamp	Pipetting soluble ligands	Spotting ligand-expressed bacterial colony
	Must uniformly distribute ligands	Rectangular plate	Circular plate	____
Identify aversive behaviors	Able to clearly record worm behaviors in an analyzable format	Video recording using *gVision*	____	____
	Must differentiate positive/negative interactions	Manual eyeballing	Machine learning algorithm	____
	Need to show spatial coordinate of identified behaviors			

Table 1: Morphological chart of function, requirements and means for our system

4.1.1 Worm Delivery

For worm delivery, we need to deliver a large number of worms to the testing arena. This functional component was broken down into two main variables: means of delivery and strains of *C. elegans*. The major requirement for means of delivery was to place a large number of worms at once to the testing array. The common laboratory practice of delivering worms is to manually pick desired worm population and transfer them to the testing plate in the form of liquid solution. We used M9 buffer to rinse highly densities of worms off the agar plates and precipitated them in a 2.5 mL microcentrifuge tube. We then pipetted desired amount of worm solution onto the testing plate.

Besides considering the means of delivery, the strain of *C. elegans* used was also a design component. To ensure the quality of the system, the desired strain needs to possess high speed and be selectively responsive to aversive stimuli. Based on recommendation from our client and previous testing, three candidates of worm strains were considered: wild type hermaphrodites (*N2*), wild type males (*him-*

5) and selectively responsive hermaphrodites (*eat-4; egl-3; sra-6 (ASH) p::eat-4)*. The genetic background of these three strains was discussed in Section 2.5.2 Genetically Modified *C. elegans* on page 24. Hypothetically, the selectively responsive hermaphrodites appeared to be advantageous over the other two strains with its ability not to be affected by non-specific stimuli. However, further testing regarding to the speed profile of these strains needed to be conducted in order to select the optimal animal model.

4.1.2 Chemical Spotting

The project component of chemical spotting required a device that can transfer multiple spots at an appropriate volume of chemicals consistently and simultaneously. Chemical spotting had two main variables: means of spotting chemicals and the configuration of spotted chemicals on the testing array. The means for spotting chemicals must be able to effectively transfer multiple ligands to the testing plates. There were three methods considered for spotting the chemicals: stamping soluble ligands, pipetting soluble ligands or placing ligand-expressing bacterial colonies. Teng *et al*. has shown in a binary system that there was a change in feeding pattern of *C. elegans* expressing human chemokine receptor 5 (CCR5) at ASH neurons encountering *E. coli* colonies expressing MIP-1alpha, which is known to interact with CCR5 (2008). However, considering the time and technical limitations to express multiple ligands in food colonies, we pre-eliminated this method because it would over-complicate our project. Therefore, we decided to use soluble ligands. Comparing stamping and pipetting, the former appeared to be superior by being less labor intensive and capable of simultaneous spotting. Since the previous MQP team has developed a PDMS stamp that could deliver up to 96 chemical spots at once, stamping method would also help simplify our project. However, due to the lack of evidence from previous MQP that showed the worm behavior was independent of the amount of the chemical spotted, we decided to further verify the stamping method in comparison to the pipetting method.

In addition to the means of chemical spotting, the configuration of the chemicals also needed to be decided. We hypothesized that the result would be more conclusive if the worms have equal chance of

accessing each ligand. Therefore, the pattern in which the ligands are placed should be unbiased. Although the stamp developed by previous MQP team only accommodated rectangular plates (i.e. 96-well plate), we proposed to test circular plate (i.e. 10cm non-tissue culture plate) considering the ligands placed along the same radius would be equally distanced to the center.

4.1.3 Behavior Identification

For the identification of animal behavior, we needed components that can record a video of the animal behavior and then analyze it to extract the appropriate data. To accurately record worm behavior, we utilized to use *gVision*. *gVision* is an image and video acquisition MATLAB application designed for physiology research. This program is customizable as well as user-friendly (Lott III, 2010). Along with *gVision*, our client has built several video tracking setups for behavioral experiments using optics, LED backlighting and machine vision cameras (Figure 11). Together with the camera setup, *gVision* can output videos in AVI format and the video can be compressed using Indeo Video 5 for further analysis (Albrecht, 2011). With *gVision* being an open source, the whole setup is fairly inexpensive (about $1000). Considering the advantages of *gVision* and the video tracking setups built by our client, we decided to use them in our system. The detailed procedure to use *gVision* was in Appendix V.

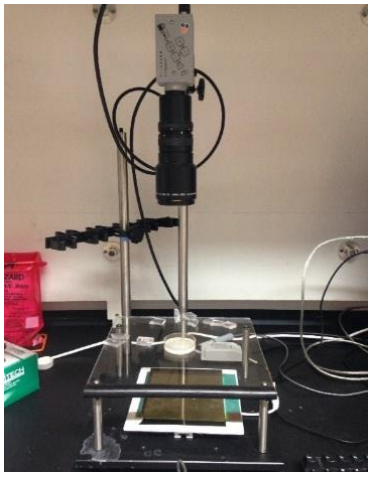

Figure 11: Video tracking setup

A working ligand-receptor screening system can successfully differentiate positive and negative interactions in a testing array. In our case, we needed the analytical tool to be able to identify the aversive behaviors among other behavior types as well as to locate where these behaviors occur. We had two options to achieve the two purposes. First of all, we can manually watch the videos and classify each spot based on the observed worm behaviors. Secondly, we can create machine learning algorithms that automatically classify each spot based on a set of training data. We hypothesized the machine learning method would yield more reliable results in an efficient fashion.

4.2 Alternative Designs

After pre-elimination process, the design alternative came down to decision about four major components: chemical spotting method, chemical configuration, worm strains and analytical tool, which are summarized in Table 2. Each component and corresponding means was described in detail in the following sections.

Functional Components	Means	Experiment plan	Decision Criteria
Chemical spotting method	Stamping vs. pipetting	Testing with rectangular plate, N2	• Chemicals can be visualized • Amount of chemicals transferred should be effective
Chemical configuration	Rectangular vs. Circular	Transferring chemicals by pipetting; Testing with N2	• Should be unbiased in inducing reversal behaviors • Should be easy to use
Worm strain	N2 vs. *him-5* vs. selectively responsive	Speed test and glycerol-avoidance test	• Aversive behaviors should be observed • Move fast and spread out • Selectively respond to aversive stimuli
Analytical method	Eyeballing vs. machine learning algorithm	Running sample tests and compare accuracy	• Accurately classify spots • Efficient

Table 2: Simplified design alternative components

4.2.1 Chemical Spotting Methods

The chemical spotting methods must efficiently delivery multiple chemical/ ligands and uniformly distribute spot volume. Two methods were considered: stamping using the PDMS stamp and

manually pipetting. Stamping the soluble chemicals involves the following steps: preparing chemicals and cleaning stamp, pipetting chemicals into 96 well plates with the desired pattern, stamping chemicals onto testing tray. The chemicals we chose to use for developing the system were 10% glycerol dissolved in deionized water (positive control) and S. Basal (negative control). Both solutions contained 1% xylene cyanol dye to visualize the spots. The schematic view of the stamping method is shown in **Error! Reference source not found.**12.

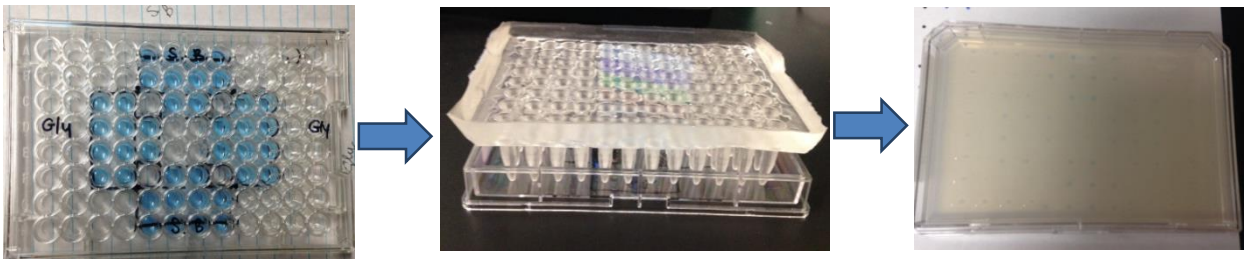

Figure 12: Images depicting major steps of chemical stamp using the PDMS stamp

To better aid pipetting the chemicals on the testing tray in the desired chemical pattern, we created a template sheet guiding the position of each chemical spot. The template sheet was placed underneath the testing tray and 1µL of the same chemicals was pipetted to each spot (Figure 13).

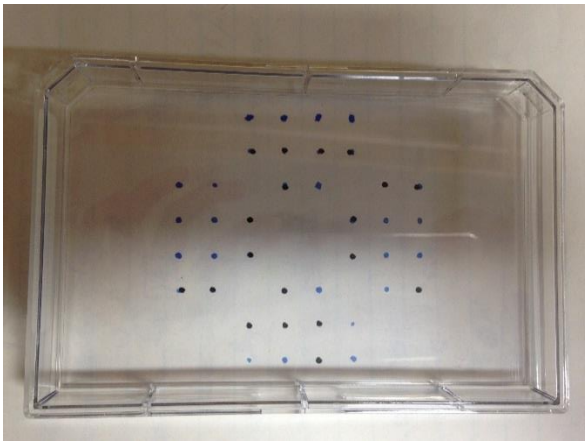

Figure 13: An empty rectangular testing plate over template sheet

To compare the two methods, we planned to run sample experiments with wild type hermaphrodites and access volume uniformity and spot sizes.

4.2.2 Chemical Configuration Options

In terms of the chemical configuration, we must uniformly distribute ligands to maximize the chance of the worms running into a spot. To achieve this criteria, we can either arrange the spots in a rectangular array or a circular array by changing the plate shape. The rectangular arrangement was the default choice of previous MQP team while we hypothesized the circular arrangement might distribute spots more evenly. The overall preparation process for rectangular plate was described in the previous section (Figure 12). Due to the lack of circular stamp, we could only use pipetting method to deliver chemicals on the circular plate. Similarly, a template sheet was made for the circular plate and the chemicals were pipetted along three different radii (Figure 14). To compare the two options, we planned to run sample experiments with wild type hermaphrodites and qualitatively assess the results.

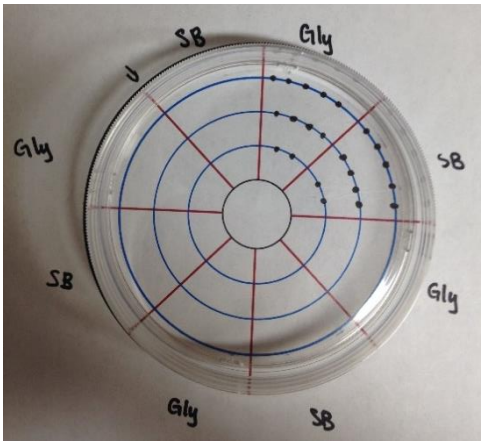

Figure 14: An empty circular plate over the template sheet

4.2.3 Worm Strain Options

The optimal worm strain for our system needed to have high mobility and the ability to selectively respond to aversive stimuli. As mentioned, we planned to test three strains: wild type hermaphrodites, wild type males and selectively responsive hermaphrodites. We planned to perform

speed test and glycerol-avoidance test to compare their mobility and ability to respond to ASH, respectively. To prepare for each test, all worms were properly maintained and relatively synchronized following standard laboratory practice. Desired number of worms was separated by picking. Rinsing prior to testing was required for all strains to avoid food stimuli.

4.2.4 Analytical Methods

The ideal analytical tool for our system needed to accurately identify chemical spot type (aversive or neutral) via worm behavior. Specifically, the analytical tool must be able to identify behaviors as well as where these behaviors occurred on the plate. As stated, we can either manually watch the recordings or generate a machine learning model to achieve the criteria. For manual test, the tester needed to eyeball each spot from each video and decided whether the spot was aversive or neutral based on observation. The obtained were compared with the actual chemical types to determine the accuracy. To avoid bias in the results, the person who spotted the chemicals was excluded from this test. The scoring system worked as the following: if most of the worms that ran into the spot showed aversive behaviors, the tester would be highly confident that this spot was aversive. Similarly, if a low number of worms showed non-aversive behaviors at a spot, the tester would confidently call this spot a neutral one. In both cases, results were given with high confidence and marked “sure”. For the spots marked “not sure”, there was either a low number of events or mixed behaviors that prevent the testers to make a confident decision.
To calculate accuracy, we divided the number of spots identified correctly by total number of spots in the assay (i.e. the results with low confidence were also included in the calculation).

For analysis using computer program, there were three main components: worm tracking, machine learning processing, generation of machine learning models and running classification. The first component of the analysis was the worm tracker code provided by Dr. Navin Pokala from The Rockefeller University. This code converts the video recordings of the worms into specific events with data on the location and behavioral state of the animal. Detailed procedure for Navin code was listed in Appendix IV. For the second component, the team created functions and scripts that sorted the data based

on the distance from the center of the chemical spot. These components also converted the data into a usable form for the machine learning algorithms. The outputs from this component were arrays and matrices containing the counts and percentage of behaviors occurring within each distance block. Once we had the processed data, we built a machine learning classification model, which was the third component, to predict the type of chemical based on the behavior data collected. To create the model, we created sets of training data using the same strain of worms that were used in the assay being analyzed (i.e. if wild type males were used in an assay, the model would be created using wild type males in the training data). The models were built with a training data set of no less than 30 chemical spots evenly split in aversive and neutral spots. The tree classification model was used to determine critical values that differentiate the neutral spots from the aversive spots. The models could then be used to test new experiments and determine the type of chemical utilizing the behavior displayed by the worms. An overview of the computer analytical method is shown in Figure 15. The accuracy of the computer analytical tool was calculated as the number of spots predicted correctly divided by the total number of spots on the assay.

We planned to run the two types of analytical methods on sample experiments with 9 spots of chemicals and compare their accuracy. The one with higher accuracy would be optimal to our system.

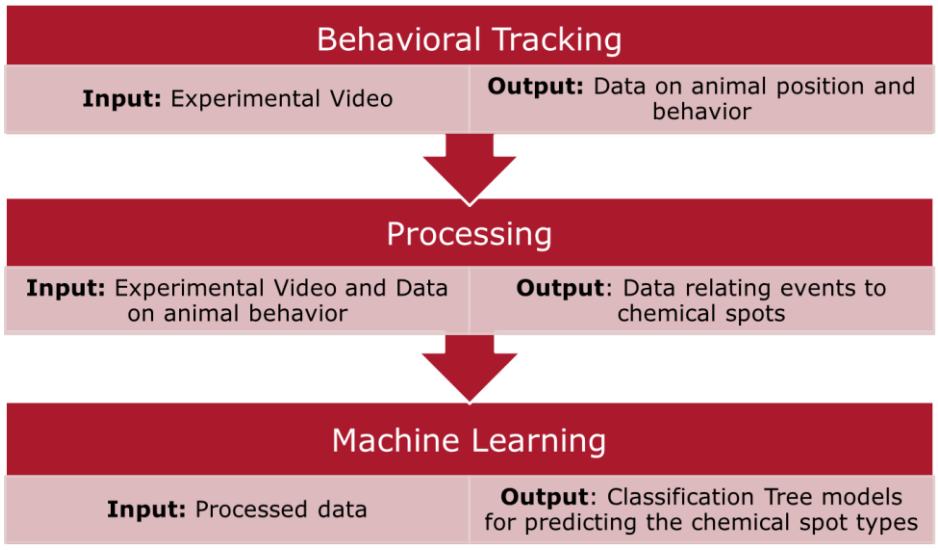

Figure 15: An overview of the computational analytical tool

4.3 Preliminary Data

In Table2, the second column lists the different means that needed to be tested. This section details the results of the various experiments conducted and some of the results. The tests included comparison of stamp versus pipetting, shape of the plate, the strain of *C. elegans*, and manual test vs. machine learning. There were additional experiments in which we examined the diffusion of glycerol in relation to worm behavior.

4.3.1 Stamp versus Pipetting

For chemical spotting methods, stamping was experimentally tested against pipetting. Concisely, S. Basal and glycerol spots were either stamped or pipetted onto the agar surface. Immediately after chemical spotting, about 30µL of pre-rinsed *N2* young adults transferred onto the center of OmniTray with agar. Once the worm liquid was dried, the behavior of the worms was recorded under the camera for 20 minutes. For more details on the procedures utilized for the stamping and pipetting experiments, refer to Appendix I.

However, based on observation, the *C. elegans* seem to aggregate in the middle and did not seem to roam around to the spots of chemicals that were placed. In addition, by the time to dry the worm liquid on the OmniTray, the glycerol diffused into the agar and became ineffective. To address the aggregation problem, the team decided to change the procedure by transferring the animals onto the agar plate first and wait 10 minutes before spotting the chemicals to allow them to move around first. The rest of the experiments were done using this method and the video shows the *C. elegans* spread out more compared to the previous procedure.

Three experiments were done with both stamping and pipetting. The same procedure was used except the method of chemical spotting was changed. Figure 16 shows one of the plates done with pipetting (left) and stamping (right). As revealed in the Figure 16, pipetting yielded better visualization and uniformity of the spots compared to stamping. The volume of glycerol for each spot was determined to be 0.2 µl for the stamp and 1.0 µl for pipetting. From the videos generated, we can see that pipetting

the chemicals yielded a longer time for the *C. elegans* to react with the glycerol because a higher volume of chemicals were spotted compared to the stamp. However, a potential problem is that since the chemicals were pipetted one by one, each spot of chemical will diffuse at different times and could throw off the results. Therefore, a decision was made to make a new stamp that can spot a higher volume of chemicals by increasing the size of the tips on the stamp.

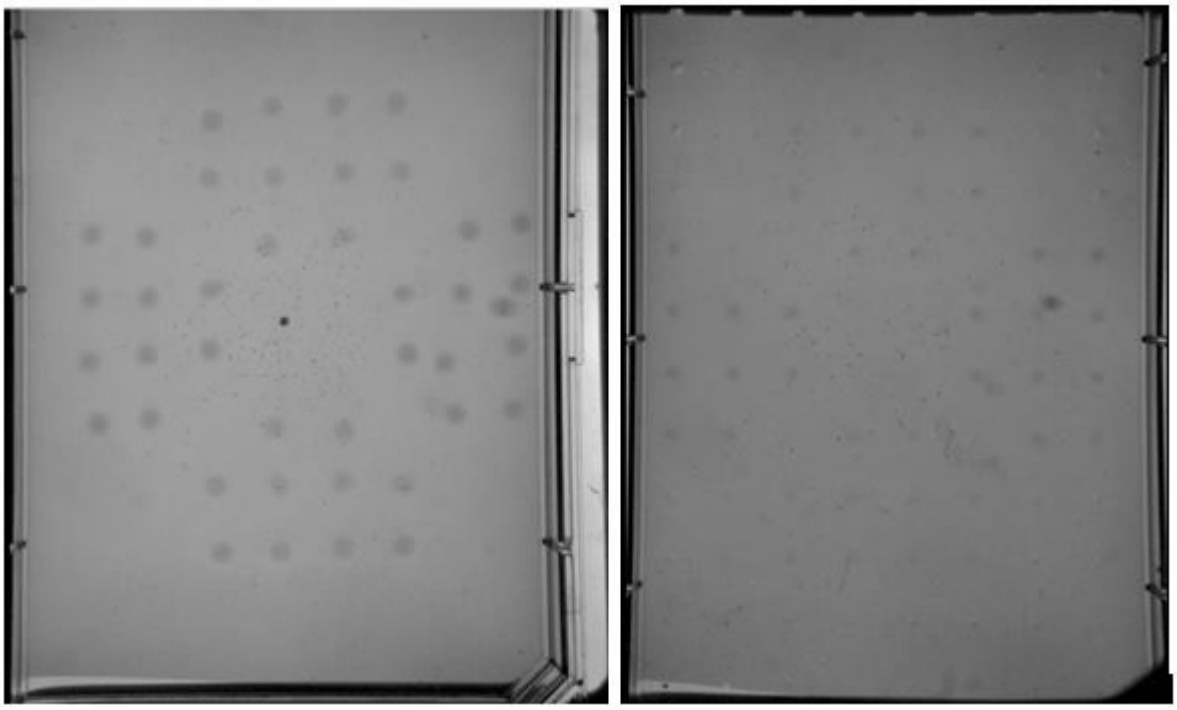

Figure 16: Video frame of the behavioral experiment. On the left, the chemicals were delivered by pipetting. On the right, the chemicals were delivered by stamping.

We made a new stamp with larger tips using a 96-well plastic plate and PDMS materials. Details on the procedures utilized for creating the new stamp are in Appendix II. The new stamp and the spotted chemicals by this new stamp is shown in Figure 17. Based on testing, the new stamp was able to yield better visualization and uniformity than the old stamp. To ensure results, we determined the volume delivered by the new stamp was about 2μL. In addition, we measured the average spot size of the two stamps by spotting 96 spots three times onto an OmniTray with agar to get a total of 288 spots for each stamp. Each plate was imaged and the diameters of each spot was determined by using ImageJ. The data for the diameter sizes can be seen in the histograms in Figures 18 and 19 for the old and new PDMS

stamp, respectively. The average spot diameter was 1.97 ± 0.36 mm for the old stamp and 3.93 ± 0.68 mm for the new stamp (Table 3). The average volume if each spot was also measured by spotting down water onto a scale to measure the weight. The volume was calculated by using the density of water (1 g/mL). Tables 4 and 5 shows the values collected for the old and new stamp, respectively. On average, the old stamp delivered 0.833 ± 0.135 µl per spot and the new stamp delivered 3.199 ± 0.472 µl per spot.

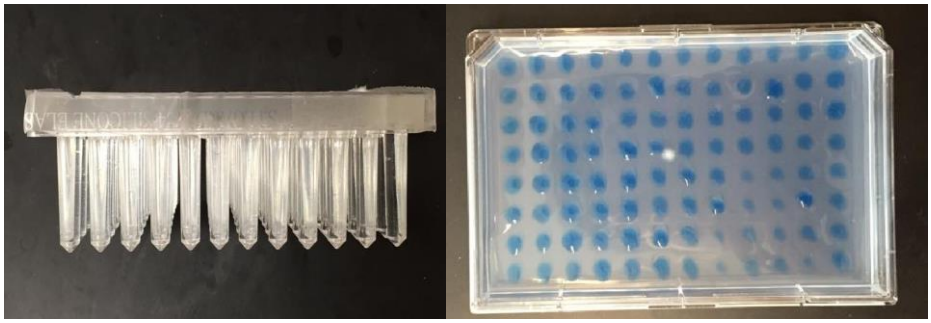

Figure 17: On the left, shown is the new stamp. On the right is the chemicals spotted on a plate.

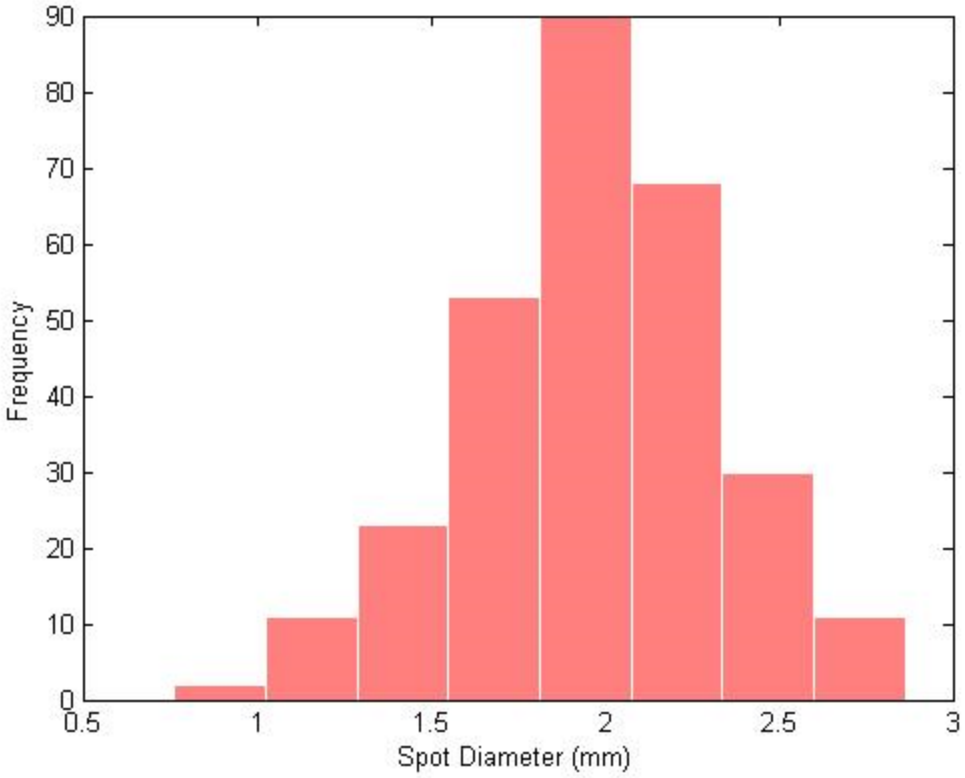

Figure 18: Histogram of the diameters of the chemical spots spotted with the original stamp.

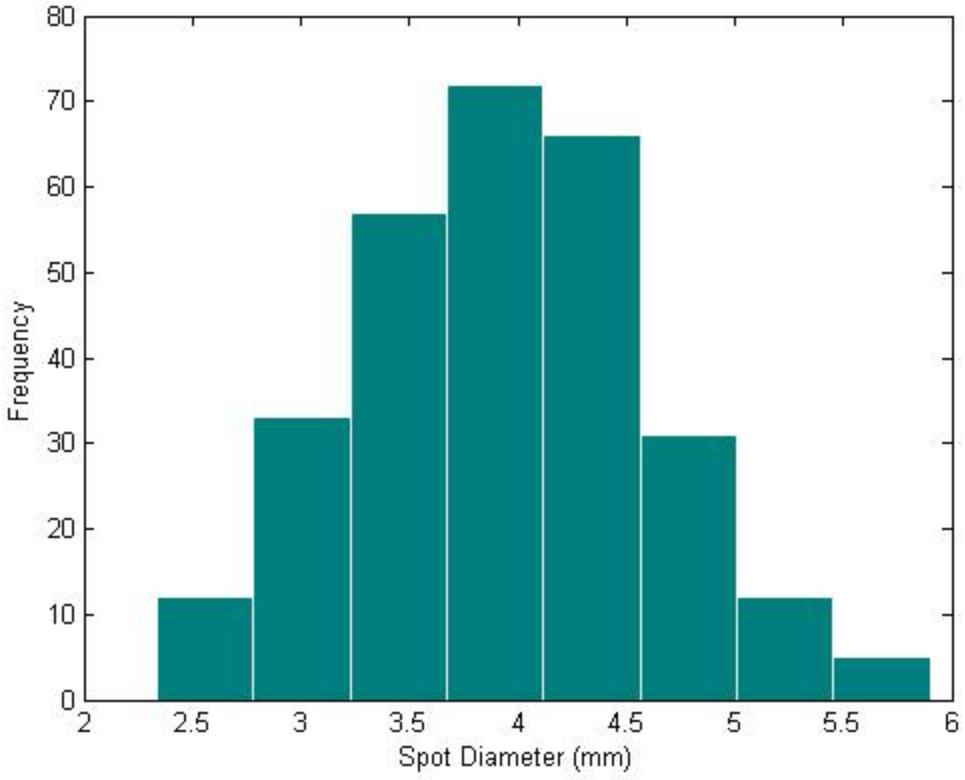

Figure 19: Histogram of the diameters of chemical spots of the new stamp

Stamp	Average Diameter (mm)	Standard Deviation (mm)
Old PDMS stamp	1.97	0.36
New PDMS stamp	3.93	0.68

Table 3: Average diameter and standard deviation of spot from the old and new stamp

Total Weight (grams)	Total Volume (microliters)	Volume Per Spot (microliters)
0.06	60	0.625
0.08	80	0.833
0.09	90	0.938
0.08	80	0.833
0.1	100	1.042
0.08	80	0.833
0.07	70	0.729
Average Volume		**0.833**
Standard Deviation		**0.135**

Table 4: Average volume delivered by the original stamp

Total Weight (grams)	Total Volume (microliters)	Volume Per Spot (microliters)
0.38	380	3.958
0.25	250	2.604
0.27	270	2.813
0.28	280	2.917
0.34	340	3.542
0.3	300	3.125
0.33	330	3.438
Average Volume		**3.199**
Standard Deviation		**0.472**

Table 5: Average volume delivered by the new stamp

4.3.2 Shape of Plate

One of the variables that was tested was the shape of the plate. The team hypothesized that using a circular plate will reduce the bias of the assay by having the spots evenly spaced out from where the worms are placed. An experiment was done similar to the experiment done with the rectangular plate (OmniTray) with a circular plate with a 100mm diameter. Three trials were done using the circular plate *N2* animals. For more details on the procedures utilized for this experiment, refer to Appendix I. Figure 20shows the circular plate setup with chemical spots placed at different radii from the center. Compared to the rectangular plate setup (Figure 21), we can see that most of the activity is in the middle, where the worms were transferred. In addition, since we modified our procedure to place the worms prior to chemical spotting, the worms were spread out regardless of the plate configuration. From these results, the team decided to continue using the rectangular plate.

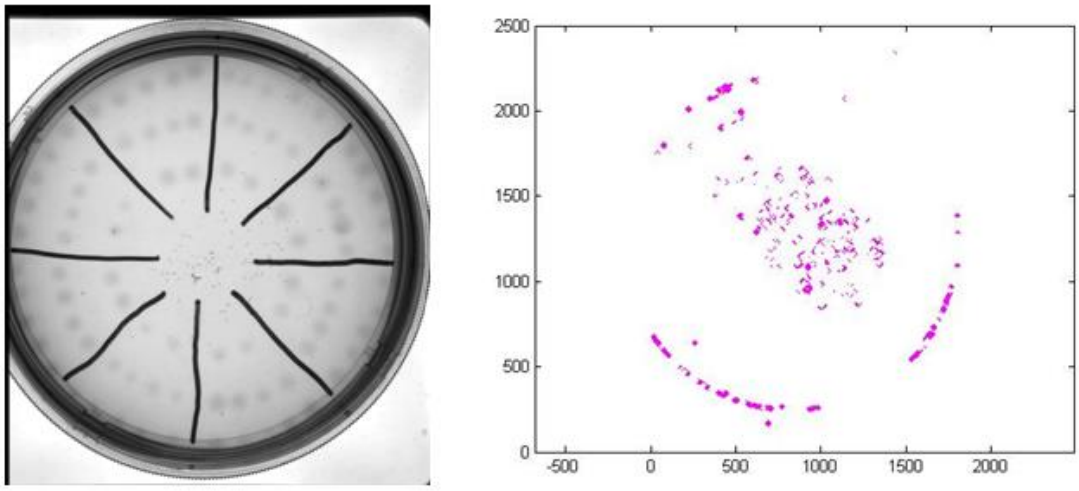

Figure 20: Images of the circular plate setup (Left) and recording with circular plate (Right)

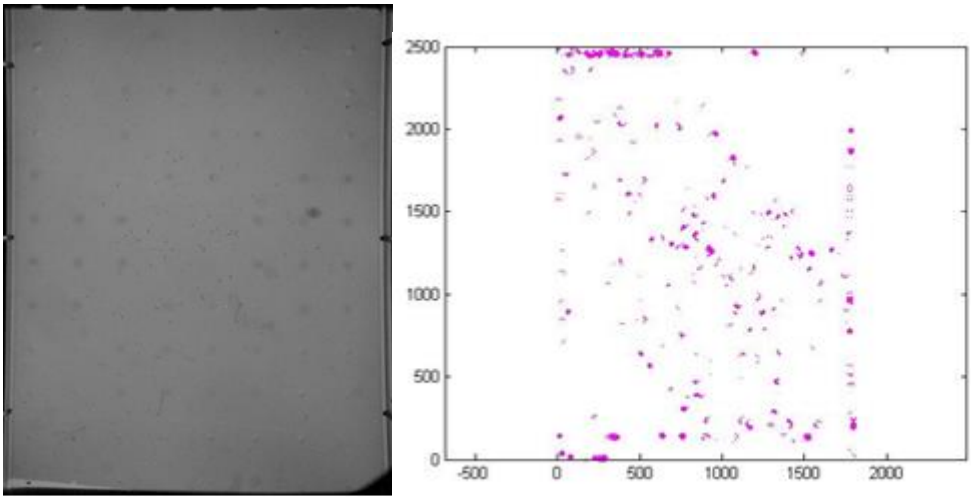

Figure 21: Images of the rectangular plate setup (Left) and recording with rectangular plate (Right)

4.3.3 *C. elegans* Strain

The strain of *C. elegans* was also a design component of our system. To determine the strain, the speed and the ability to selectively respond to ASH of the animals were examined. First of all, avoidance test was performed to test their responses. In this test, a small drop of repellent, in this case, 2M glycerol, was delivered near the tail of a worm via a capillary tube. The worm needed to be pre-rinsed and had to be moving forward in order to perform such delivery. Once the drop came in contact with the tail, the capillary action would drive the chemical to reach the anterior sensory organs and activate the ASH

neuron. The response was scored as positive if the worm stops moving forward and started moving backward within 4seconds. The result of a drop test assay was characterized by the avoidance index, which was the number of positive responses divided by the total number of trials. For comparison, the same procedure was done using the control chemical deionized water. The avoidance index for all three strains was summarized in Figure 22. This result showed all three strains were highly responsive to aversive stimuli. Noticeably, only selectively responsive hermaphrodites showed no response to deionized water suggesting its inability to respond to non-selective stimuli. From the avoidance test, using selectively responsive hermaphrodites would be optimal.

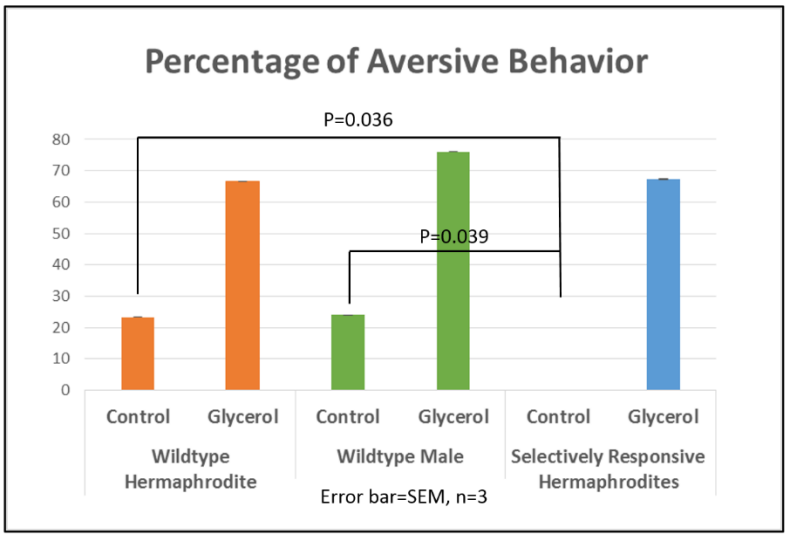

Figure 22: The avoidance index of the three strains that were tested. Mann-Whitney U-Test (non-parametric t-test) was performed on the control chemical results from the three strains (two tails). The result is significant at $p \leq 0.05$. For wildtype hermaphrodites and wildtype males, n=1. For selectively responsive hermaphrodites, n=3.

Besides their ability to selectively response to ASH, the speeds of three different strains of *C. elegans* were examined by recording a video around 20 worms for each strain. The speed was determined using the behavioral tracking program from Dr. Navin Pokala and the result can be seen in Figure 23. From the results, we can see that the speed of wild type males was comparable and faster than the other two strains.

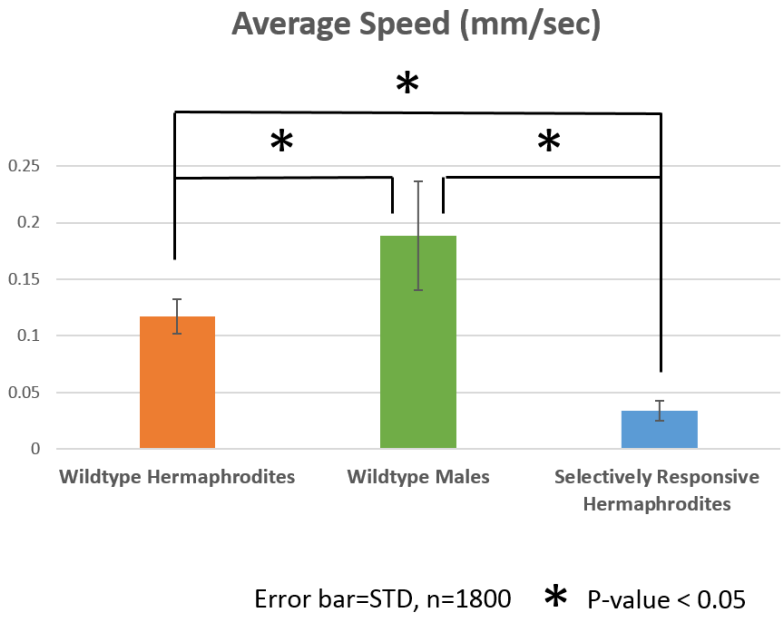

Figure 23: The speed graphs of the three strains that were tested. Statistical significance was determined using a t-test between wild type hermaphrodites and wild type males, wild type males and the selectively responsive hermaphrodites, and wild type hermaphrodites and selectively responsive hermaphrodites. The results show a significant difference at a P-value < 0.05.

Considering the results from the two tests, the ideal worm strain would be selectively responsive males. Therefore, we took the advantage of *C. elegans* being able to be genetically modified and decided to create a selectively responsive male strain as our final decision for choice of worm strain.

According to our advisor's suggestion, there two ways to create the line we wanted. We could either cross *him-5* with a selectively responsive strain or create males from the selectively responsive strain. Based on laboratory experiences, manually increasing frequency of males by altering environmental factor yields very low successful rate for any worm strain. With the limited time, we decided to pursue the crossing method. Ideally, the selectively responsive strain would be the optimal choice for this cross because it responds minimally to non-selective stimuli. However, considering the complexity of its genotype, the cross between the selectively responsive strain and *him-5* was unlikely to succeed. Therefore, we decided to cross a glutamate deficient strain (*eat-4*) and *him-5* to yield selectively responsive males as the final worm strain choice for the system. More details on the cross generation is in Appendix I.

Our team attempted to create *eat-4*; *him-5* cross as the strain for the final experiment. The *eat-4* hermaphrodites were crossed to *him-5* males, and the *eat-4; him-5* males were crossed to *eat-4* hermaphrodites again to obtain a sTable*eat-4; him-5* line (F1). Then, the F1 generation was cloned. The cloning process remained for three to four generations to stabilize the line. PCR was conducted to characterize the line after lyse the worms (Appendix III). The *eat-4* characteristic bands, approximate 600 bp, were displayed in the F3 generation PCR (Figure 24). However, the *eat-4* characteristic of the line was lost in future generation (Figure 25). This is a common issue in generating stable line with multiple mutants. The strain generation should be repeated by creating more than three plates of cross at a time to increase the possibility for obtaining the ideal *eat-4;him-5* line.

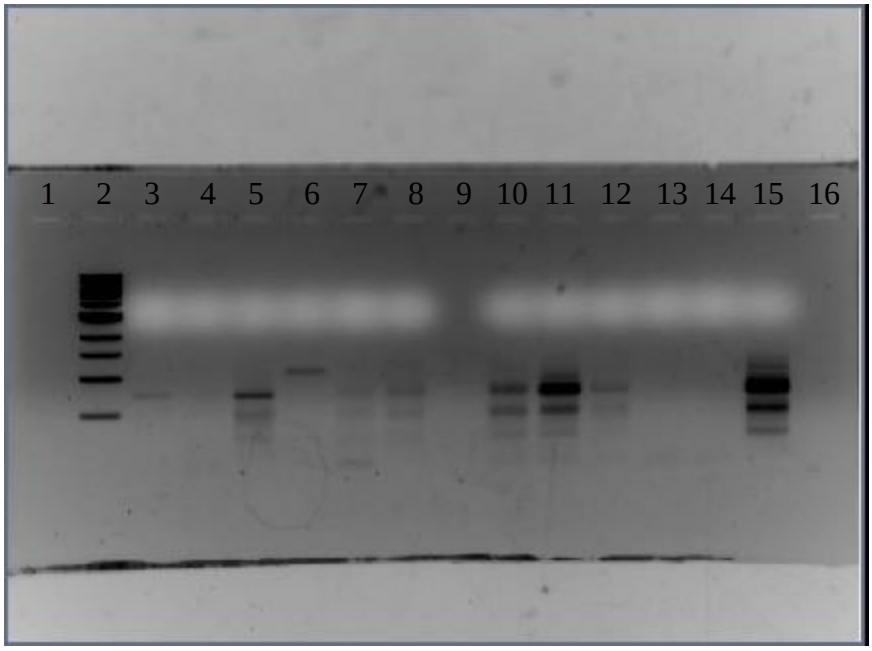

Figure 24: Electrophoresis gel of *eat-4;him-5* cross from the first PCR. Lane 2 contains a DNA ladder and the rest were the cross

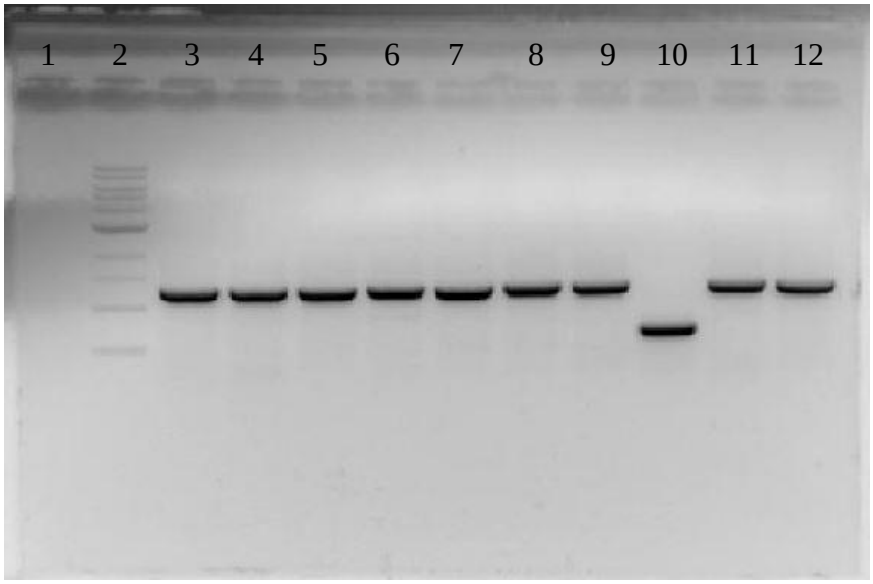

Figure 25: Electrophoresis gel of *eat-4;him-5* cross from a second PCR. Lane 2 contains a DNA ladder, lane 10 contains a regular eat-4 strain, lane 11 contains N2 strain and lane 12 contains him-5. The gel shows that there is no eat-4 strain present and the cross

Due to the time limitation, our team failed to create a selectively responsive male strain in time. Therefore, we used wild type males in our final experiments because we were able to obtain more number of events due to their high mobility.

4.3.4 Manual Analysis vs. Machine Learning Analysis

As mentioned, we needed an analytical tool that can accurately classify chemical spot types based on worm behaviors. To preliminarily compare the accuracy of the manual analysis to that of the machine learning analysis, we generated 5 different experiments each with 9 chemical spots. We did not use the optimal worm strain (wild type males) for these 9-spot assays; instead, about 50 mixed gender *him-5* worms were used per plate.

Using the scoring system described in the previous section, we conducted the manual tests on the five 9-spot videos we recorded. The results were summarized in Table 6. The average accuracy of the manual analysis was 46.7% (±14.5, n=5), which was not ideal. Based on observation, many chemical

spots did not have enough number of events to make a classification. In addition, the neutral spots were hard to be correctly identified.

Trial 1			**Trial 2**		
Spot Number	Result	Comment	Spot Number	Result	Comment
1	N	Not sure	1	A	Sure
2	N/A	Low event	2	N/A	Low event
3	A	Not sure	3	A	Sure
4	N/A	Low event	4	N/A	Low event
5	A	Sure	5	A	Not sure
6	N/A	Low event	6	N/A	Low event
7	N	Sure	7	A	Sure
8	A	Sure	8	N	Not Sure
9	N	Sure	9	N/A	Low event
# Sure	4	-	# Sure	3	
Accuracy	2/9	22.2%	Accuracy	5/9	55.6%
Trial 3			**Trial 4**		
Spot Number	Result	Comment	Spot Number	Result	Comment
1	A	Sure	1	A	Not sure
2	N/A	Low event	2	N/A	Low event
3	A	Sure	3	A	Sure
4	N/A	Not sure	4	A	Sure
5	N/A	Low event	5	N/A	Low event
6	N/A	Low event	6	N	Sure
7	A	Sure	7	A	Sure
8	N	Sure	8	N/A	Low event
9	N	Sure	9	A	Sure
# Sure	5		# Sure	5	
Accuracy	4/9	44.4%	Accuracy	5/9	55.6%
Trial 5			**Actual Spots**		
Spot Number	Result	Comment	Spot Number	Spot Type	
1	A	Sure	1	A	
2	A	Not Sure	2	N	
3	A	Sure	3	A	
4	N/A	Low event	4	N	
5	A	Sure	5	A	
6	N	Sure	6	N	
7	A	Sure	7	A	
8	N/A	Low event	8	N	
9	N/A	Low event	9	A	
# Sure	5				
Accuracy	5/9	55.6%			
Average Accuracy	**46.7%**	**STD**	**14.5**	**n=5**	

Table 6: Manual test results on 9-spot sample videos

To select an optimal analytical method, we also use machine learning analysis to identify chemical types for each spot. For better comparison, we used the same five 9-spot assays to run the

analysis. Detail procedure to run the machine learning analysis was listed in Appendix VII. However, we found there was huge variation in results between plates to plates. The accuracy ranged from 50% to 100% (Unlike manual test, machine learning always assigned a result to a spot; therefore the lowest accuracy achieved would be 50%). Therefore, we attempted to examine the reasons for this variation and the possibility to improve.

We wanted to examine if there was a noticeable difference between the percentages of behaviors occurring around the glycerol spots in comparison to the neutral chemical spots. Figure 26, shows the percentages of aversive behaviors observed for glycerol spots in five different plates with standard deviations. Figure 27, shows the percentages of aversive behaviors observed for neutral spots in five different plates with standard deviations. As shown in the two graphs, the difference in percentage of aversive behaviors around the glycerol spots in comparison to that around the neutral spots was not substantial. For both stimuli, the percentage of aversive behaviors ranged roughly from 10% to 30%. Therefore, we hypothesized because of this similarity in behaviors between the two stimuli the machine learning program was not able to differentiate the two chemical types accurately.

Besides the ambiguity in behaviors, we also observed the variation in the number of events between plates that might cause the machine learning to fail. Figure 28 showed the total number of events detected by the program for all behavior types in these five 9-spot assays. As shown, some plates had substantially lower number of events than others.

Overall, both manual analysis and machine learning analysis did not produce satisfactory results based on the preliminary data. However, the machine learning analysis was more advantageous over manual analysis because it had potential for further improvement. According to the result analysis, we hypothesized we can improve the machine learning analysis by 1) using more homogenous worm population in the test (wild type males instead of mixed gender) 2) eliminating unnecessary data sets and 3) increasing the number of events by having more spots and more animals. Both methods would be used in our final experiments and further comparison would be made.

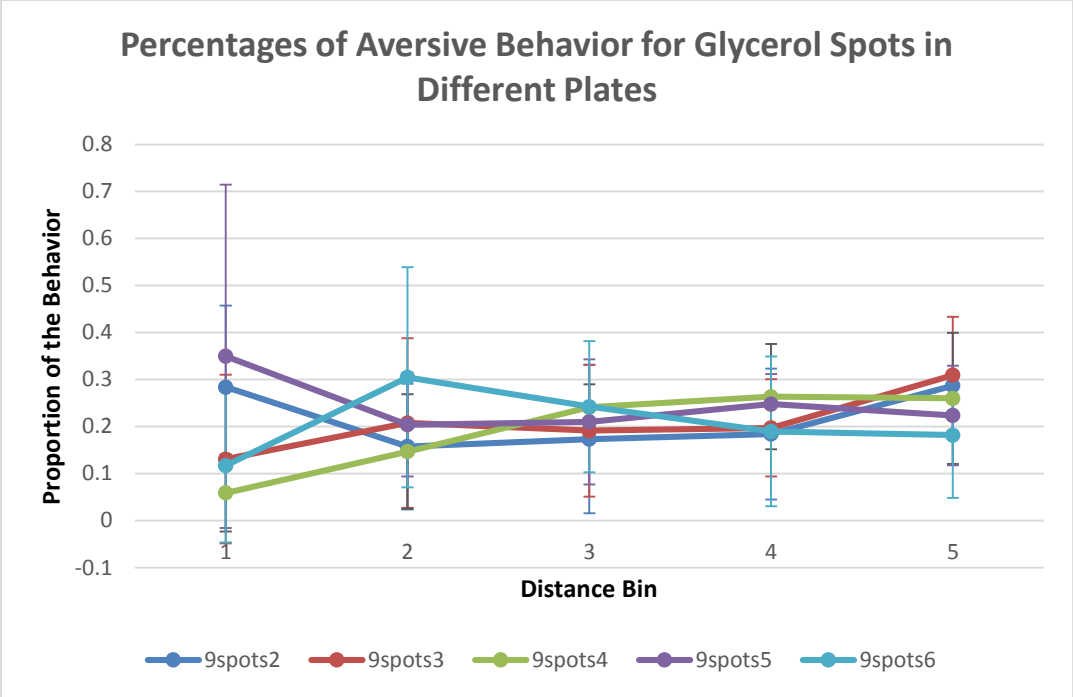

Figure 26: The plot of percentages of aversive behaviors occurring at aversive chemical spots in different distance bins for different experiments

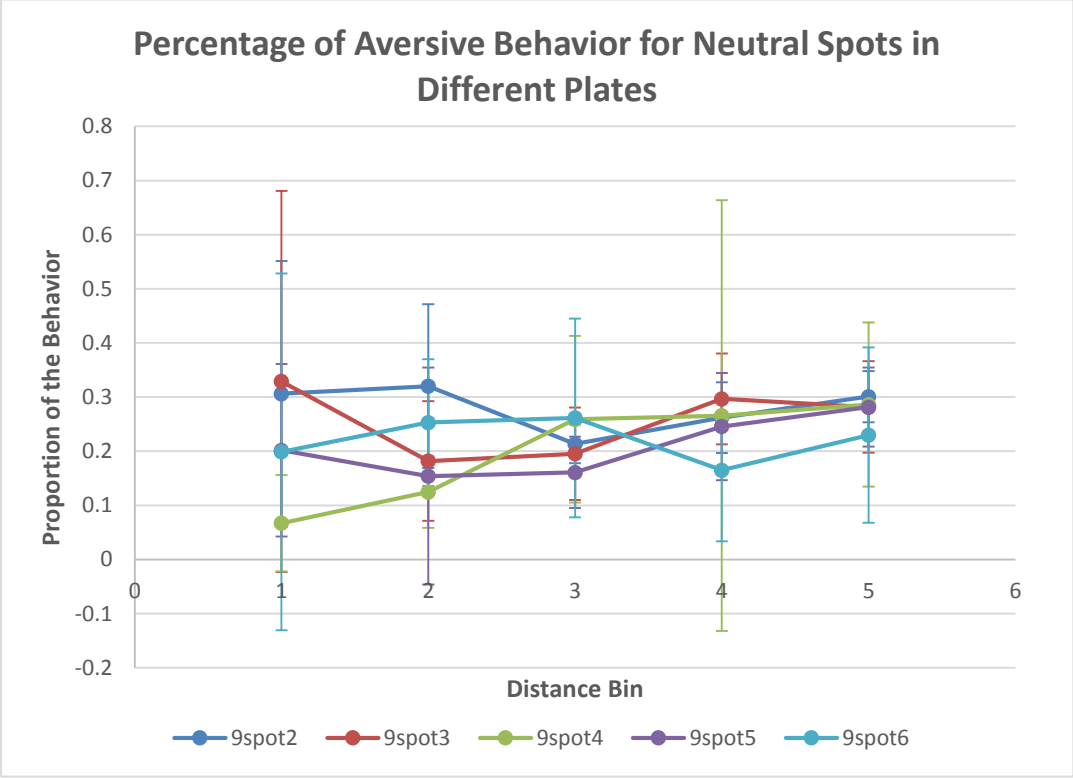

Figure 27: The plot of percentages of aversive behaviors at neutral chemical spots occurring in different distance bins for different experiments

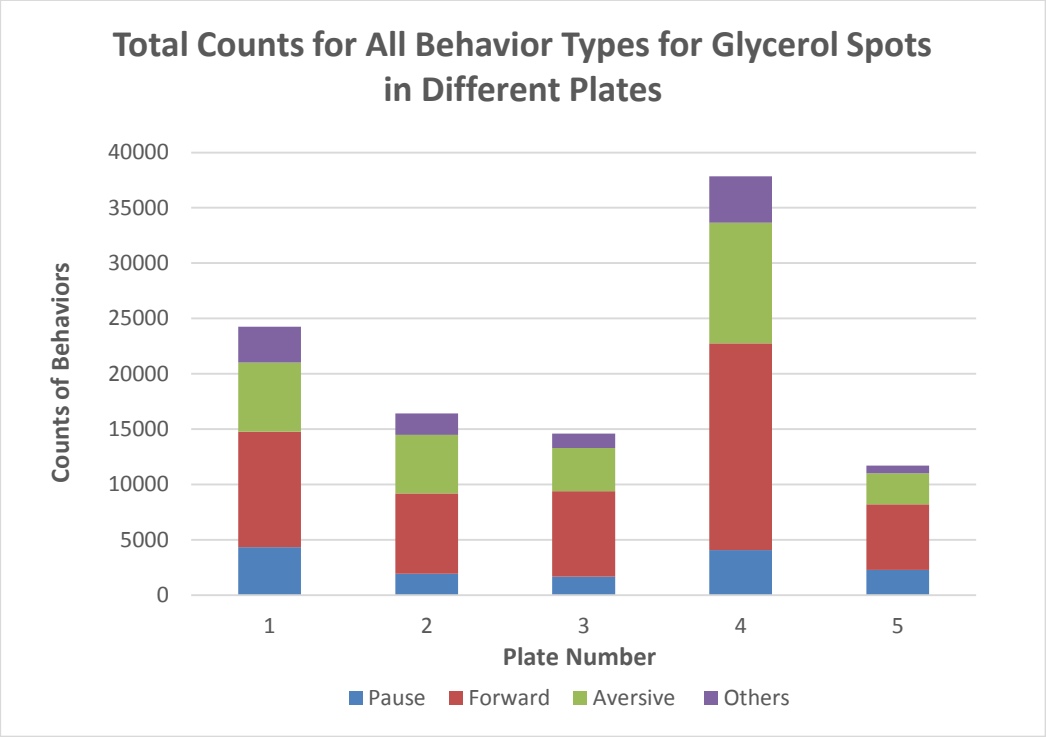

Figure 28: The plot of total counts of behaviors for all behavior types around glycerol spots in five different plates

4.3.5 Supplemental experiments

Tests were done to seen the effect of glycerol diffusion on *C. elegans* behavior. The experiment was done to see if it was viable to change the concentration of the glycerol spots to increase the time the worms will interact with the chemical. This experiment was done by placing a drop of glycerol onto an agar plate. A single *C. elegans* was placed near the drop at different time points to see if there's an aversive behavior to the glycerol. Different concentrations of glycerol were also tested. Table 7 shows the result of the test. The results show that the glycerol of all the tested concentrations lost its affect after about 10 minutes. At the higher concentrations, the worms seem to avoid the chemical spot altogether and would not approach the vicinity of the chemical spot. Another experiment was also done to see whether or not the agar concentration and the thickness (5mL of agar solution in thin plate and 10mL of agar solution in regular plate) of the gel will affect the diffusion rate. Regular thickness of agar was tested against a thinner gel ranging from 1X to 3.5X concentration of agar. Similar to the previous result, all the tests

yielded no response from the animal after 9 minutes (Table8). Appendix I and VIII contains more details on the agar concentrations and procedures followed. The concentration of agar and thinness of the agar did not significantly affect the diffusion rate. In addition, the agar in greater than 2.5X concentration of agar solution was not completely dissolve through heating the solution up. We decided to continue using the regular concentration and thinness of the agar plate for further experiment. Figure 29shows the plate setup done for this experiment and the state plot. Five spots of glycerol were spotted onto an unseeded agar plate. The visualization of the reversals is more prominent here because there were less worms on the plate compared to the experiments done on the OmniTray. In the state plot, five distinct areas can be seen from the behaviors that correspond with the 5 spots of chemicals that were spotted.

Glycerol Concentration	0 Minutes	5 Minutes	10 Minutes
10%	Response	Response	No Response
20%	Response	Response	Response
30%	Response	No Response	No Response
40%	Response	No Response	No Response
50%	Response	Response	No Response

Table 7: C. elegans response to different glycerol concentration at different time points

Agar Concentration and thickness	0-2 Minutes	2-4 Minutes	4-6 Minutes	6-8 Minutes	After 8 Minutes
1X, regular	Response	Response	Response	No Response	No Response
1X, thin	Response	Response	Response	No Response	No Response
1.5X, regular	Response	Response	Response	No Response	No Response
1.5X, thin	Response	Response	Response	No Response	No Response
2X, regular	Response	Response	Response	No Response	No Response
2X, thin	Response	Response	Response	No Response	No Response
2.5X, regular	Response	Response	Response	Response	No Response
2.5X, thin	Response	Response	Response	Response	No Response
3X, regular	Response	Response	Response	Response	No Response
3X, thin	Response	Response	Response	Response	No Response
3.5X, regular	Response	Response	Response	Response	No Response
3.5X, thin	Response	Response	Response	Response	No Response

Table 8: C. elegans response to glycerol in different agar concentration and thickness

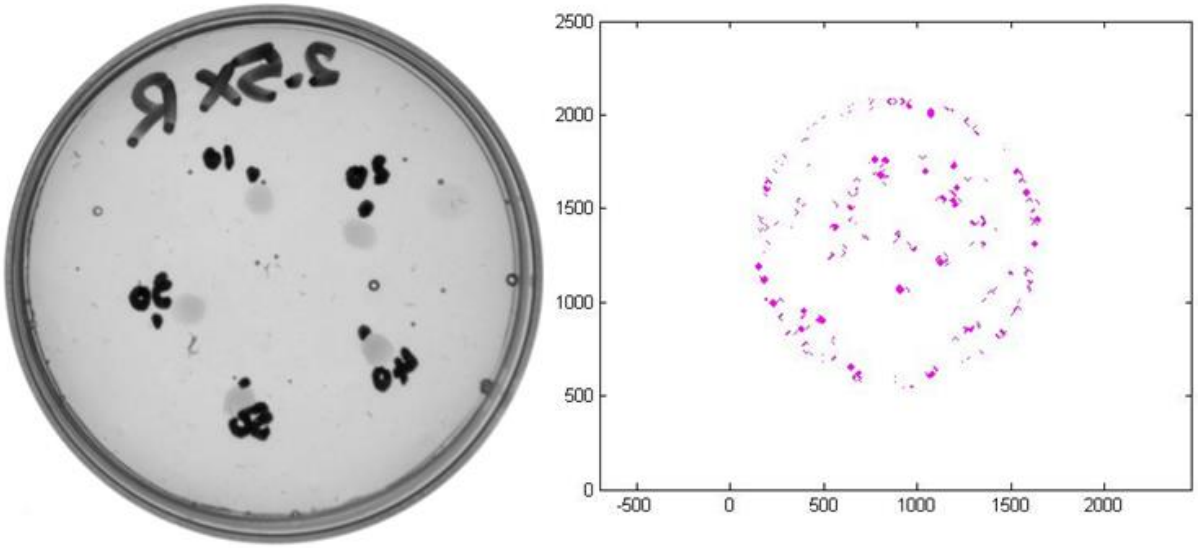

Figure 29: Setup of the experiment done with 3.5x agar on regular thickness along with state plot

Chapter 5:Design Verification

5.1 Final Design Decisions

In general, the goal of our project was to design a rapid ligand-screening system that will expedite ligand-based drug discovery process. The system utilized the aversive behavior of *C. elegans* as animal model to indicate ligand-receptor interaction. In addition, the system needed to accurately detect multiple active ligands that interact with a specific receptor of interest in one testing array. In order to achieve these requirements, our system was designed to have four major components: chemical screening array, worm strain, behavior recording, and data analysis programs. To create each of these components, the team designed a chemical spotting tool, chose a worm strain and created an analytical tool. Based on the preliminary experiments, we decided final design decisions for each of these components. The worm strain that was used for our final testing was him-5 males. The modified PDMS stamp was chosen to deliver the chemicals that was able to deliver an average of 3.199 ± 0.472 µl per spot. For analytical method, both manual test and machine learning were used to analyze the final experiment results.

5.1.1 Overview of System Process

Overall, we would prepare the worms and delivery them at five spots on the testing plate (four corners and center spot). This way provided the most efficient worm spreading. Following worm distribution, chemicals were spotted onto the rectangular testing plate using modified PDMS stamp. We would start recording immediately using *gVision*. Each assay would be recorded for 20 minutes. Lastly, we would analyze the video using modified analysis package. The overall process of the system is summarized in Figure 30.

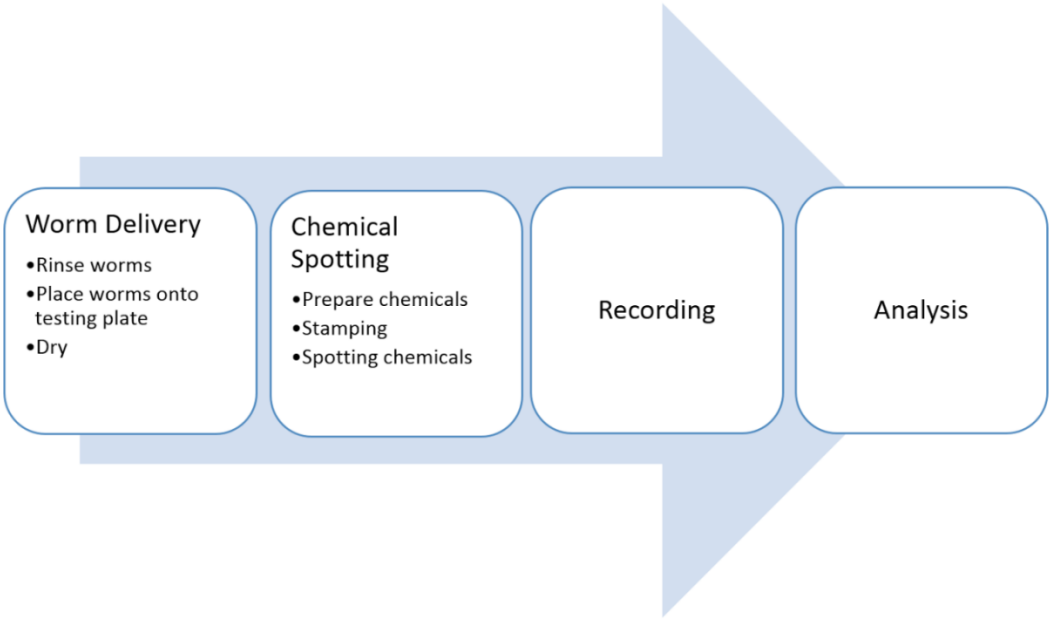

Figure 30: Finalized overall process of the system

5.2 Experimental Results

5.2.1 Final System Experiment

By following the final design decisions, experiments with 4, 9 and 16 chemical spots in regular 60 mm Petri dishes and a full plate of 96 chemical spots were performed. In order to obtain enough him-5 males in the experiment day, the *him-5* male worms were prepared by crossing 3 him-5 hermaphrodites to 8 him-5 males three days ahead. On the experiment day, approximate 30, 30, 60 and 300 him-5 male worms were isolated, rinsed by M9 solution twice and transferred to a rectangular Omni Tray plate. At the time the worms were spreading out on the plate, 10% glycerol and S. basal solutions with 1% of dye were prepared. By the time the worms were evenly distributed, the corresponding numbers of chemicals were pipetted to the 60 mm Petri dishes or stamped to the Omni Tray plate with the modified PDMS stamp from 96 well plate that containing the chemical in specific patterns (Figure31). The plates were recorded and analyzed through the MATLAB program.

The final system was easy to use by simply following the instruction on the protocol. Since *him-5* male was selected for the final experiment design due to its quick aversive response to the specific chemical associated to the ligand-receptor interaction, enough *him-5* males need to be generated before the experiment. Crossing was set up 3 days ahead as mentioned in the previous section. The procedure of crossing was straight forward by placing 3 and 8 *him-5* hermaphrodites and males, respectively, to an OP50 seeded NGM plate. Multiple crosses could be set up if needed. Besides ease of use, the system was reusable as well, which would be economic. Once the strain was generated, the worms only required to pick and maintain twice a week. In addition, since the recording only took 10 minutes, the plates could be reused by sanitizing the plates properly with bleach, 70% ethanol and distilled water after discarding the agar containing worms and chemicals to the biohazard trash. In the same manner, the PDMS stamp as well as the 96 well plate could be washed and rinsed for future use.

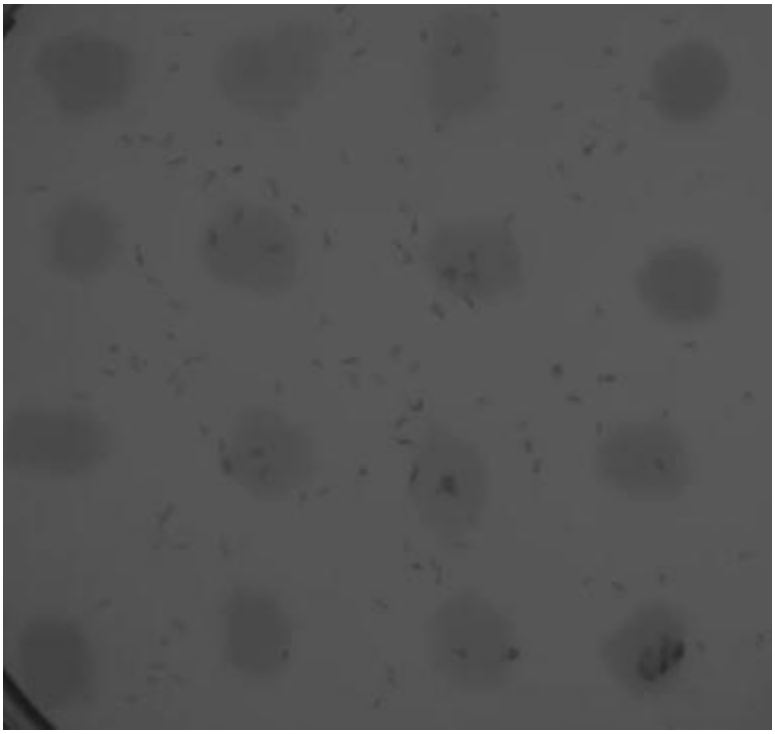

Figure 31: Picture of a 4x4 chemical spot array with C. elegans exploring the array

5.2.2 Machine Learning

5.2.2.1 Lasso Analysis

To determine the importance of different observations in the experiment, we utilized a technique called Lasso. Lasso is a machine learning technique that assigns penalties and weights to determine the importance of different observations to the determination of different classes of the data being examined. The results of a Lasso analysis can be graphical shown. The most significant observations break from the penalty line early in the plot and also have largest end magnitude. If the observation remains on the penalty line (at 0), it is not significant in distinguishing between the different classes in the data.

For our data analysis, we processed the final experiment data with 16 chemical spots into five distance bins with data from all of the frames. The observations utilized for the Lasso was the percentage of behavior occurrence within each of the distance bins. This resulted in 20 different observations being

considered in the data analysis. Figure 32 shows the plot of the lambda values associated with each of the observations.

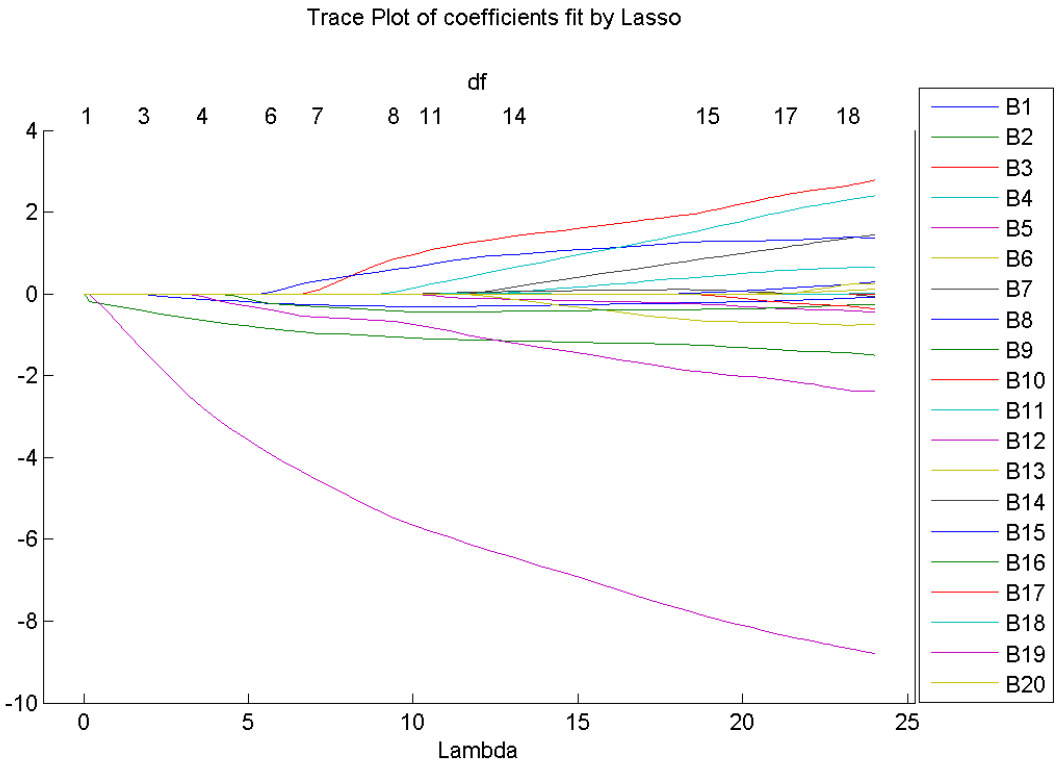

Figure 32: Plot of lambda values of each observation run through Lasso analysis

5.2.2.2 Optimizing the Machine Learning models

While preparing for the final experiments, the team examined how to optimize the machine learning models to get the most accurate results. Within MATLAB, it is possible to modify the weighs of observation – this modifies how the classification model considers the different observations. The team opted to modify the training data sets that were being considered by the program. Based on the observations from manual testing and from the Lasso analysis, the team determined several likely data configurations that may be more important to determining the chemical type.We used 5 distance bins and all data sets had 1801 frames (10 minutes of data). We tested two different configurations of frame sets

and 4 different configurations of distance bins. The frame sets tested were all frames (0-1801) and trimmed frames (250-1000 frames). Based on our observations the behavior towards the beginning and end of the video were not very helpful to determining the chemical spot type. For the configuration of distance bins, we excluded the outer distance bins and innermost distance bins in different configurations because we observed that the area close to the center of the chemical spot and further away from the chemical spot center were not useful for determining the spot type.

For these tests, the analysis used the data collected from the 9 chemical spots with mixed gender plates of *him-5* worms were utilized. The models were tested with the built in MATLAB functions to determine re-substitution error and k-fold loss with a 4 fold cross validation. A 4 fold cross validation was conducted because we had such a small data set and a 4 fold cross validation allow for a reasonable number of data points to be integrated into each model. The results are shown for the different models tested. The best result was the model that had the lowest values for re-substitution error and k-fold loss (Table 9).

Frame Number Selection	Distance Bins utilized	Re-substitution Error	K-fold Loss
0-1800	1 to 5	0.0857	0.4286
0-1800	2 to 5	0.0286	0.4286
0-1800	3 to 5	0.0857	0.4857
0-1801	2 to 4	0.0857	0.4
250-1000	1 to 5	0.0286	0.5143
250-1000	2 to 5	0.1429	0.6857
250-1000	3 to 5	0.1143	0.5714
250-1000	2 to 4	0.1429	0.6857

Table 9: Model characteristics including data composition, re-substitution error and k-fold loss

5.2.2.3 Machine Learning Results for Final Experiments

The Machine learning model that was generated from training data was used to determine the type of spots there are on a 4 x 4 spot array on three different plates. The confidence scores for each trial can be seen in Figures 33, 35 and 37. The actual spot array and the spot array determined by the model can be seen in Figures 34, 36, and 38. Any spots with a confidence score of 0 to 0.5 were considered to be neutral and any scores above that were considered aversive. For the first trial (Figures 33 and 34), the

model was able to determine 10/16 spots correctly. For the second trial (Figures 35 and 36), the model was able to determine 11/16 spots correctly. For the third trial (Figures 37 and 38), the model was able to determine 12/16 spots correctly. The average number of spots that the machine learning model got correct was around 68.8% (±0.063, n=3). Details on the code used to generate these plots, create the models and test the data are in Appendix VI. Code and Appendix VII. Code Procedure.

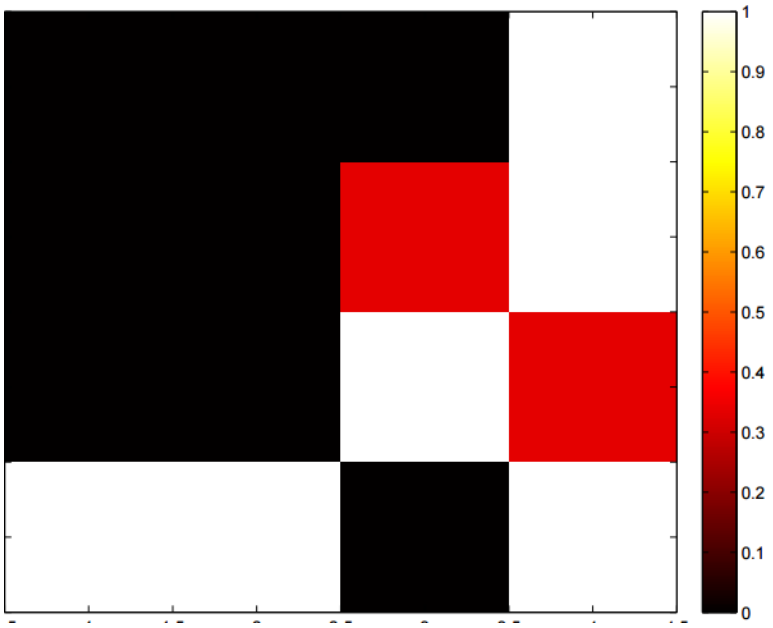

Figure 33: Confidence score for the 4 x 4 spot array for the first trial. A confidence score of 1 represents 100% confidence that the spot was aversive and a score of 0 represents a 100% confidence that the spot was neutral

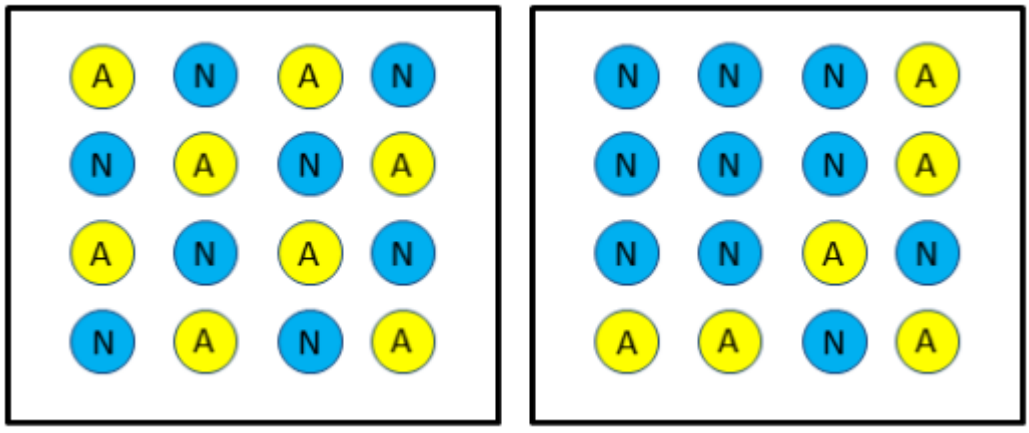

Figure 34: The actual spot array can be seen in the Figure to the left and the spot array generated by the model can be seen on the right for the first trial. The model determined 10/16 spots correctly

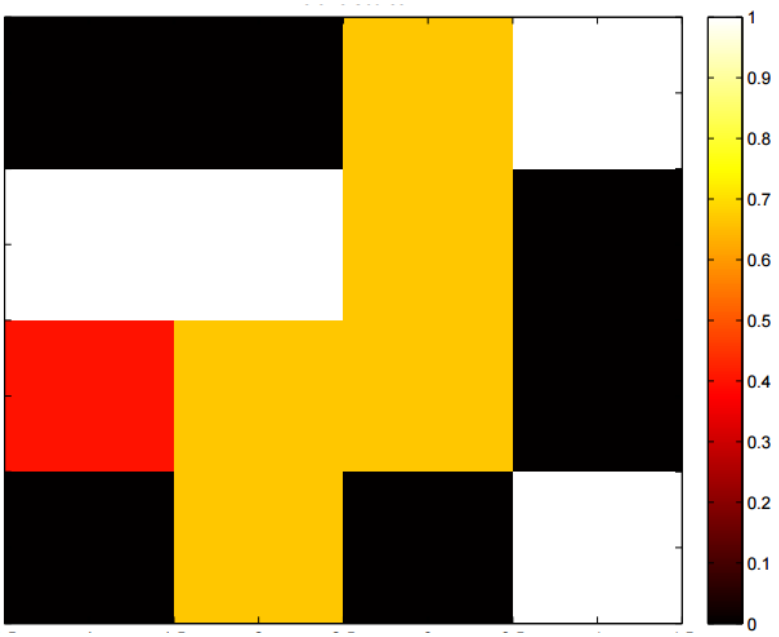

Figure 35: Confidence score for the 4 x 4 spot array for the second trial. A confidence score of 1 represents 100% confidence that the spot was aversive and a score of 0 represents a 100% confidence that the spot was neutral

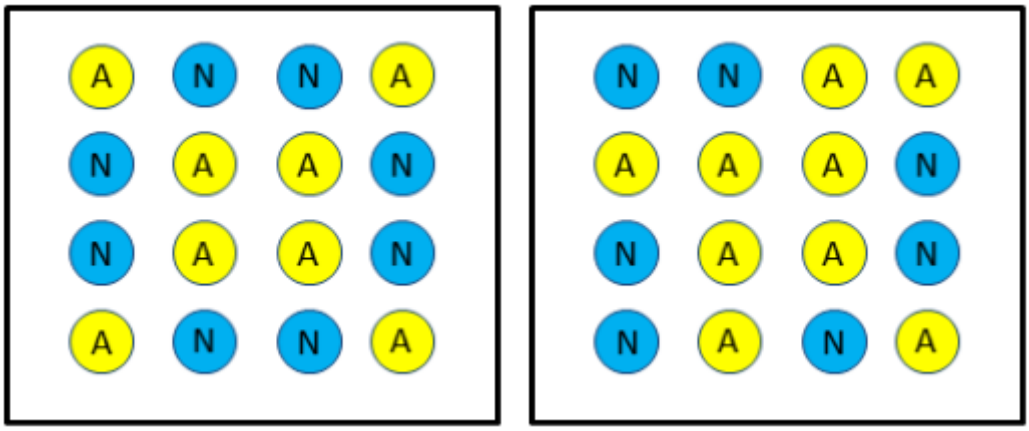

Figure 36: The actual spot array can be seen in the Figure to the left and the spot array generated by the model can be seen on the right for the second trial. The model determined 11/16 spots correctly

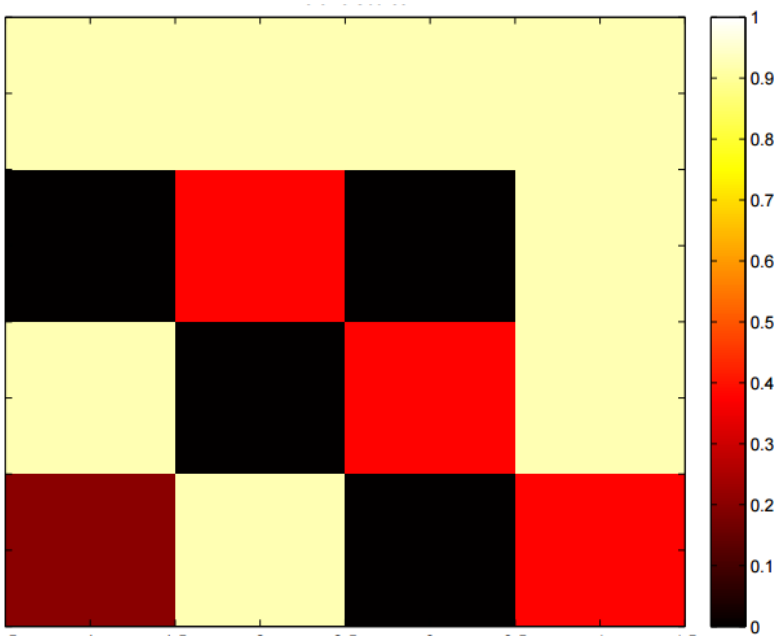

Figure 37: Confidence score for the 4 x 4 spot array for the third trial. A confidence score of 1 represents 100% confidence that the spot was aversive and a score of 0 represents a 100% confidence that the spot was neutral

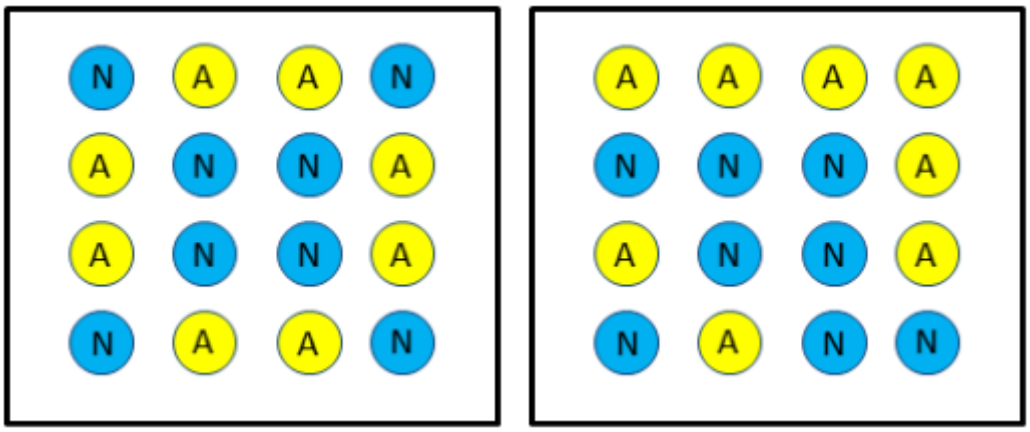

Figure 38: The actual spot array can be seen in the Figure to the left and the spot array generated by the model can be seen on the right for the third trial. The model was able to determine 12/16 spots correctly

5.2.3 Manual Test Results

We performed manual tests on the data we collected from three 4 x 4 spot array videos. The instruction of the manual test was described in Chapter 4.2.4. We listed the results for three trials in Table 10, 11, 12, respectively.

Overall, the manual tests were able to determine 47.4% (±11.6, n=6) of the chemical spots correctly. The spots that the identifier were not sure on (low confidence) were included in the calculation

of average percentage of spots identified correctly. Since the machine learning algorithm determines the spot types even for low confidence spots, the low confidence spots in the manual identification test were included for a better comparison. During the manual tests, we observed that the accuracy of classifying neutral spots was generally lower than that of the aversive ones. In other words, the aversive spots were much easier to recognize than the neutral ones. This observation can be caused by two possible reasons. First of all, there was a high frequency of random aversive behaviors suggesting the limitation of using wild type males (*him-5*) as the animal model. This problem can potentially be overcame by using selectively responsive males. Secondly, using S. Basil as the neutral spots might not be sufficient to produce clear comparison to the aversive spots. We suggested to use other neutral chemicals or attractants to produce a strong contrast with the aversive chemicals.

Trial 1					
Spot Number	Person 1	Comment 1	Person 2	Comment 2	Actual
1	A	SURE	A	LOW EVENT	A
2	A	SURE	N	NOT SURE	N
3	N	SURE	A	SURE	A
4	A	SURE	A	NOT SURE	N
5	N	SURE	N	SURE	A
6	N	SURE	N	NOT SURE	N
7	A	SURE	A	SURE	A
8	N	SURE	A	SURE	N
9	A	SURE	N	SURE	A
10	A	SURE	N	SURE	N
11	N	SURE	A	NOT SURE	A
12	A	SURE	A	NOT SURE	N
13	A	SURE	A	SURE	A
14	N	SURE	N	SURE	N
15	A	SURE	N	SURE	A
16	N	SURE	N	NOT SURE	N
# SURE	16		16		
Confidence	9/16	56.3%	10/16	62.5%	

Table 10: Manual identification data of the spots for the first trial video

Trial 2					
Spot Number	Person 1	Comment 1	Person 2	Comment 2	Actual
1	A	SURE	A	SURE	A
2	N/A	NOT SURE	N	SURE	N
3	N/A	NOT SURE	N	SURE	N
4	A	SURE	A	SURE	A
5	N/A	LOW EVENT	N/A	LOW EVENT	N
6	N	SURE	N/A	NOT SURE	A
7	N	SURE	N/A	NOT SURE	A
8	N/A	LOW EVENT	N/A	NOT SURE	N
9	N/A	LOW EVENT	N/A	NOT SURE	N
10	A	SURE	N/A	NOT SURE	A
11	N	SURE	N/A	NOT SURE	A
12	A	SURE	N	SURE	N
13	A	SURE	N/A	NOT SURE	A
14	A	SURE	N	SURE	N
15	N	SURE	N	SURE	N
16	N/A	NOT SURE	A	SURE	A
# SURE	10		8		
Confidence	5/16	31.3%	8/16	50.0%	

Table 11: Manual identification data of the spots for the second trial video

Trial 3					
Spot Number	Person 1	Comment 1	Person 2	Comment 2	Actual
1	N/A	LOW EVENT	N/A	LOW EVENT	N
2	N	SURE	N/A	NOT SURE	A
3	A	SURE	A	SURE	A
4	N/A	NOT SURE	N/A	NOT SURE	N
5	A	SURE	A	SURE	A
6	N/A	NOT SURE	N/A	NOT SURE	N
7	N	SURE	N	SURE	N
8	N/A	NOT SURE	N	SURE	A
9	A	SURE	A	SURE	A
10	N/A	NOT SURE	N	SURE	N
11	A	SURE	N/A	NOT SURE	N
12	N/A	NOT SURE	N/A	NOT SURE	A
13	N/A	NOT SURE	N	SURE	N
14	A	SURE	A	SURE	A
15	A	SURE	N	SURE	A
16	N/A	LOW EVENT	N/A	NOT SURE	N
# SURE	8		9		
Confidence	6/16	37.5%	7/16	46.7%	

Table 12: Manual identification data of the spots for the third trial video

Chapter 6: Discussion

6.1 Reviewing Objectives and Constraints

The primary goal of this project were to create an easy to use behavior-based system that could rapidly and reliably determine the occurrence of a molecular interaction. Additionally, the final system had to be relatively inexpensive in comparison to the existing systems and had to test multiple ligands simultaneously. Reliability was determined to be the most important objective. The final system needed to provide conclusive and reproducible results on the type of each chemical spot. Next, the system needed to be easy to use, not requiring specialized training with a simple to understand and execute procedure. Finally, a rapid system will be both timely and efficient. Such a system would be able to test as many ligands as possible in little time. Our final design was a system that had all the necessary components to test up to 96 ligands simultaneously. In the previous sections, we described the experiments conducted to verify these components and procedures. This chapter contains the discussion of our findings in terms of our objectives and constraints.

In terms of reliability, the machine learning program had an overall accuracy of 68.8% with the best test delivering an accuracy of 89%. To create the models for these tests, we utilized a data set of *him-5* males containing the behaviors occurring at both neutral and aversive chemical spots. Based on the optimization we conducted in Section 5.2.2.2, we used modified training data sets. In comparison to the manual tests accuracy of 47.4%, the machine learning had an improved accuracy rate. A closer examination of the confidence scores of the machine learning revealed some interesting patterns. The machine learning program was usually able to confidently determine that a chemical spot was aversive. Meaning, if the program was very confident that the spot was aversive it was likely aversive. However, with the neutral chemical spots the program was more likely to be unsure or to decide the wrong type of chemical spot. This suggests that it was easier to distinguish an aversive spot than it was to determine a neutral spot. This result is desired because it is better to get more false positive (aversive) results than to get more false negative (neutral) results. Our system is supposed to be used for initial discovery of ligand-

receptor pairs. For example, if 7 out of 16 spots were identified to be aversive spots, we would want that data set to include all the actual aversive spots then to lose them in the first test. By having more false positive results from our system, we limit the possibility of identifying right away an actual aversive spot as neutral. It is better to include all potential positive spots than to lose them initially because later testing can be done to sort through the actual positive results. When conducting our experiments, we minimized the variability between the preparations of different experiments by carefully following our procedures. However, we observed that there was still variability due to ambient conditions and variation in the worm growth conditions. This variability could also be due to the fact that the worm were also responding to other stimuli that made our analysis more difficult.

The final system and final procedure is easy to use. To start the experiment, worms were rinsed and then plated into the testing plate, chemicals were transferred with the PDMS stamp after the worms were evenly distributed. The behavioral responses was recorded with a simple camera setup and analyzed the recordings using *gVision*, a MATLAB program. In addition, the analytical processes built by the teamwas automated. The user only needs to upload the files to the correct folder, run the scripts and then input the information prompted from the scripts. This is a huge advance over the binary behavior-based system developed by Teng*et al* in which manual counting of aversive behaviors was required (2006). In addition, in comparison to the cell-based system, using worm behaviors had the following advantages. Firstly, it is possible to create a stable worm strain that expresses the receptor of interest that can be preserved and used repeatedly. The cell-based system, on the other hand, requires each-time transfection of receptor. Secondly, the neuronal circuit that controls the aversive behavior of *C. elegans* is well-characterized whereas the intracellular signaling pathways activated by the transfected receptor are not always consistent. Lastly, *C. elegans* required minimal maintenance with very basic techniques and skills.

The system was also able to identify spots in a rapid manner. During the course of this project, the team initial conducted experiments with 4 chemical spots, 9 chemical spots and 16 chemical spots. At most, it took 8 hours to collect, process and then analyze the data. The processing to determine worm behavior with the code provided by Navin Pokala took the most time. Oftentimes, the team ran the worm

behavior code overnight and then completed the remaining analysis the next day. This time can be reduced by having better computing power.

Our final system stayed within the original constraints creating a system that can screen multiple ligands in a single test while be significantly less expensive that the competitive cell-based systems. As previously stated, we created all the appropriate components to run a 96 chemical experiment and determine the optimal conditions to generate meaningful data for a 96 chemical experiment. Additionally, our experiments demonstrated that 16 chemical spot experiments could be tested and the chemical types could be determined with reasonable accuracy. In the end, the expense of the final system was estimated to cost around $900 for all the system components and the cost of experiments will cost around $50 per test. The experimental cost includes the cost of agar plates, chemicals, and animals. This is significantly cheaper than the cell-based assays available on the market.

6.2 Discussion of Results

6.2.1. *C. elegans* Cross

The results of the *eat-4;him-5* cross was not successful based on the PCR conducted. Initially, the *eat-4* strain was crossed with *him-5* males. The male progeny from that cross was isolated and crossed back to the eat-4 strain again. The hermaphrodites of the second cross were isolated to check for males in their progeny. The resulting plates with males were believed to have both the*eat-4* and *him-5* strains and PCR was performed to confirm the assumption. Wild type creatures would have a band occurring at 1235 base pairs (bp) while the mutant *eat-4*would be indicated with a band at 621bp. The first PCR gel in Figure 24 shows some indication of the *eat-4* allele present. However, a subsequent PCR did not show the *eat-4* allele and only had the wild-type. This suggests that the initial cross was unstable and after a few generations, the animals in the plate reverted back to wild-type and lost the eat-4 allele.

6.2.2. Chemical Spot Size

The average spot sizes based on diameter were calculated from measured values for the old and new PDMS stamp. The average spot diameter was 1.97 ± 0.36 mm for the old stamp and 3.93 ± 0.68 mm for the new stamp. Based on the results, the new stamp was able to consistently deliver a controlled amount of chemicals to each spot. The standard deviation for the new and old stamp were both around 18% of the average, which is relatively low. The purpose of modifying the old PDMS stamp was to deliver a greater volume of chemicals because once a significant portion of the glycerol diffuses into the agar the worms would no longer react to the chemical. Once modified the new stamp delivered improved volumes at a consistent rate.

6.2.3. Data Analysis of Animal Behavior

The model generated from the machine learning was used to determine the type of chemical for each spot in three trials. From the results, the machine learning model was able to correctly identify 67% of the chemical spots. These results were compared to the manual identification, where a person manually determined the chemical spot type by watching the videos that were recorded. The manual identification was able to identify 43.5% of the chemical spots correct. The results suggests that the algorithm is better compared to manual identification of the chemical spots, but it is not perfect. 67% success is relatively low compared to the cell-based assays, which reported an accuracy of about 95% (Caers *et al*., 2014). The problem with manual identification is that it is hard to determine some of the spots when there are a lot of animals in on plate and some of the spots were identified with low confidence. The results from the machine learning model also shows results that were not all 100% confident for aversive and neutral spots. However, the machine learning model makes a decision based on data that a person might not have in determining the spot type from the manual identification. Overall, the machine learning model needs to be improved upon to generate more accurate and reliable results.

6.3 Economics

Our animal behavior-based ligand screening system can be utilized to the drug discovery process, specifically for the creation of ligand-based drugs. There are several steps in ligand-based drug discovery process: identification of a receptor-ligand interaction, design of a ligand-mimetic drug, drug optimization and clinical testing. The ligand-based drug discovery process starts with identifying specific ligand-receptor interaction that is related to the biological phenomenon or a disease of interest. Once this interaction is identified and characterized, a drug can be designed to mimic the known ligand to either block or enhance the receptor –ligand pathway. After these initial steps, there is further drug optimization, preclinical trials and finally, clinical trials. Figure 39 illustrates this process with more details (Castanho, 2011). Our project addresses the initial step of this drug discovery process. We want to create a system that expedites the process of identifying receptor-ligand pairs that have the potential to lead to the development of new drugs.

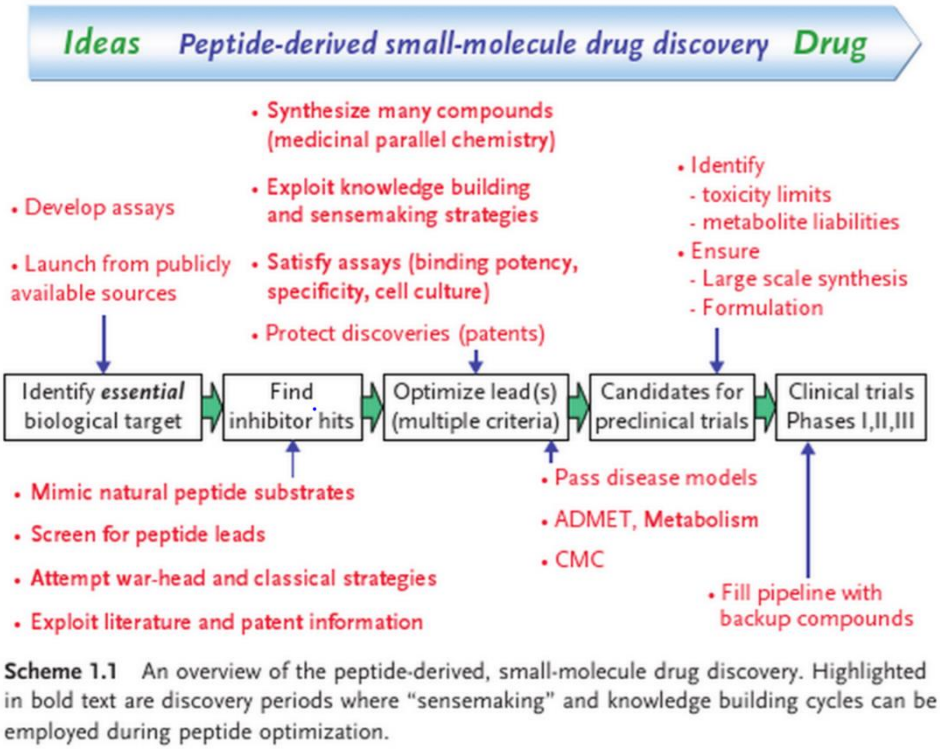

Figure 39: An overview of a ligand-based drug discovery process (Castanho, 2011)

Our system is very cost efficient compared to other available ligand-screening systems. The components of our system include stable *C.elegans* strains, 96-well plates, a PDMS stamp, OmniTray plates, a camera and a recording station. Excluding the cost of chemicals, ligands and receptor, the most expensive component of our system would be the camera and recording station, which roughly cost $1000. Other lab equipment expenses were trivial. The inexpensive and rapid receptor-ligand identification system being developed in this project has the potential to reduce the enormous costs incurred during the drug development process. Based on a 2003 study, the average cost of a successfully developed drug was estimated to be $802 million in 2000 (Austin, 2006). A significant portion of the new drug discovery expense goes to projects in research and development phase. This phase could utilize our system to accelerate the drug development process. Figure 40 is an estimated cost break-down for drug discovery and development. Despite of the enormous cost to design and market a new drug, the pharmaceutical industry remains one of the most profitable industries in the United States. When the average return on assets for all other industries is 4.7%, the pharmaceutical industry has a return as much as 10.3% in 2005 (Austin, 2006). Consequently, the success of our system could aid in the discovery of a plethora of new drugs and potentially reduce the costs for drug development.

DiMasi and Others' Estimate of Average Research Costs and Times for Successfully Developed New Molecular Entities

	Average Length of Research Phase		
	Preclinical Phase (4.3 years)[a]	Clinical Trials and FDA Approval (7.5 years)	Total (11.8 years)
Research and Development Costs (Millions of 2000 dollars)			
Direct costs	121	282	403
Opportunity costs[b]	214	185	399
Total Costs	**335**	**467**	**802**

Source: Joseph A. DiMasi, Ronald W. Hansen, and Henry G. Grabowski, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics,* vol. 22, no. 2 (March 2003), pp. 151-185.

Note: FDA = Food and Drug Administration.

a. The estimate for the duration of the preclinical phase is based on the comprehensive drug database maintained by the Tufts University Center for the Study of Drug Development.

b. Opportunity costs are the costs associated with keeping capital tied up in a specific drug-development project for a given period (that is, the forgone interest or earnings that a company might have gained from investing its capital in other ways). DiMasi and others assumed the forgone rate of return to be 11 percent per year.

Figure 40: Estimated costs for the development of a new drug (Austin, 2006).

6.4 Environmental Impact

In general, none of the materials used in our experiments or our equipment had any harmful effect to the environment. The chemicals utilized in our tests are not known to cause significant environmental detriment. To maintain our animal model *C. elegans*, they were fed a bacterial food source and grown on Nematode Growth Medium (NGM) agar. The food source was *E. coli* OP50 (Stiernagle, 2006). All ingredients involved in bacterial growth were properly stored and used following standard laboratory practices to prevent contamination and infection. Preparation of NGM plates required NaCl, agar, peptone, cholesterol in ethanol, KPO_4 buffer and $MgSO_4$ (Stiernagle, 2006). Unwanted worm plates were disposed of in biohazard waste bins and all other excess chemicals were properly disposed of. Additionally, the team fabricated a PDMS stamp. PDMA is known as a safe biomaterial and is a component of many everyday products. Other chemicals used for experiments included small amount of glycerol, S. basal and dye. They are all water soluble were disposed of following standard laboratory procedures.

6.5 Societal Influence

Our system was designed for the initial stage of ligand-based drug design; this technology can be applied to a broad range of disease treatment. Ligand-receptor interactions throughout the body regulate homeostasis, neuronal communication, and cell proliferation. As a result of their extensive involvement in many physiological activities, mutations or miscommunication of a ligand-receptor pair can lead to acute or chronic diseases. Once the effect of a receptor-ligand interaction has been characterized, successful drug treatment can then developed. A simple example of a successful drug based on known a ligand-receptor interaction is the anti-histamine drug, Claritin. Claritin was developed by taking advantage of the receptor-ligand interactions of histamine and the importance of these interactions to the propagation of allergic reactions. The drug binds competitively and with greater affinity than histamine curbing the allergic reaction. For every successful ligand-mimetic drug, there are numerous

metabolic and neurological diseases that can possibly be treated by exploiting yet unknown ligand-receptor interactions. According to a global market report, high throughput screening assays to discovery these receptor-ligand interactions will account for 30% of the global drug discovery technology market in 2015 (estimated to be $18 billion) (“Drug Discovery Technologies”, 2014). Our system can aid in the creation of libraries of receptor-ligand interactions than can later be utilized to create new ligand-mimetic drugs.

6.6 Political Ramifications

In recent years, there has been many discussions regarding the accessibility of health care products. With more government regulation and intervention in place, the cost and time to get a new drug approved has significantly increased. Simultaneously, the United States has an aging and growing population. To counter these issues, the government has the obligation to make advanced health care products affordable. From the technical aspect, one way to reduce the cost of a new drug is to advance the drug discovery technologies so that the drug discovery process can be well refined and efficiently accelerated.

Our project can be applied to advance the first steps of drug discovery. However, a fully refined drug discovery process requires a profound understanding of the disease mechanism prior to drug design. This implies that technology advancement is needed in every stage of the drug discovery process. For example, a sophisticated computer program will be needed for molecular modeling of the identified ligand-receptor pairs in order to study their structures and functions. Other innovative technologies may also be needed to understand signaling pathways that lead to a condition of interest (Gwynne and Heebner, 2001). However, research is a low profit venture even if it fundamental to drug development. Private companies are more likely to be attracted to the actual drug design process if the basic research was previously completed. Therefore, changes to funding policies are needed to allow for this research and will help achieve more affordable and accessible health care in the United States.

6.7 Ethical Concerns

This project created and used mutant strains of *C. elegans*. Among the stains utilized were a stain that had deletions for glutamate gene and another stain that had increased male population in the offspring. Additionally, the goal of the system was to inject worms with genes to express a gene of interest the ASH neuron. This would elicit an aversive behavioral response in the worms. The team utilized standard techniques for genetic modification of these worms.

In general, genetic modification of animals in the interest of scientific research has received considerable ethical debate. The objections toward these practices are generally based on the concepts of the intrinsic value of animals and of animal integrity. The intrinsic value of animals believes that animal should not be treated in the same way people treat laboratory equipment. Animal integrity mostly refers to the wholeness and completeness of animals. Genetic modification will violate this wholeness and completeness of animals by changing their genome (De Vries, 2006). However, these moral objections concerning genetic engineering is mostly focused on high level organisms such as mice, dogs, and cats. Considering the simplicity of *C. elegans* as an organism, there are few ethical concerns involving the genetic modification of *C. elegans*.

6.8 Health and Safety Issue

As stated previously, our project belongs to the initial stage of drug development; therefore, there has no experiment involved with human as study model. In terms of the safety of the team, all team members have been certified to work in Worcester Polytechnic Institute's Gateway Park Laboratories by passing Lab Safety Exams. Therefore, the health and safety of the lab members were ensured by strictly following lab protocols.

6.9 Manufacturability

In general, there have been two approaches to improve drug discovery technologies: automation and assay miniaturization. Laboratory automation is an effective but fairly costly mean of

increasing productivity. Many researchers have developed or utilized automated equipment or systems to screen a large quantity of samples in much shorter time. Our project is one example of assay miniaturization, where the cost was lowered by reducing volume. This approach also allowed us to screen a large array of chemicals in a single experiment for a low cost. The use of basic laboratory equipment makes our system easy to reproduce. The genetic modification of *C. elegans* was genetically stable; therefore, only regular strain maintenance will be required. The components of chemical assays were all standard laboratory plates which are widely available in the market. One drawback of our system is low scale-up manufacturability. Maintaining and preparation of *C. elegans* for a large array screening test with more than hundreds of ligands will be labor-intensive.

6.10 Sustainability

To better sustain our ecology and reduce laboratory disposal expense, we consistently recycled non-biohazard or non-contaminated lab plastic ware. Since our system has low requirements for sterilization, agar was disposed into biohazard waste and plates were cleaned for future use. Similarly, since all chemicals we used were water-soluble, we would properly clean and reused the 96-well plates. In addition, we utilized small amount of chemicals for each experiment; therefore, there was no heavy chemical contamination or additional waste processing created by our project.

Chapter 7: Final Design and Validation

7.1 Review of the Project

Ligand-receptor interactions are widely involved in physiological activities such as neuronal communications, homeostasis regulation and cell proliferation; therefore, these interactions in involved in many disease pathologies and have be utilized to developed cures for these diseases. Once the binding ligand of a particular receptor is fully characterized, drugs can be designed either to mimic the function of the ligand or to antagonize it. The currently existing ligand-receptor screening systems can be very time-consuming and expensive. Knowing the tremendous value of ligand-receptor identification in drug market, we were asked to design, implement and characterize a system that utilizes behaviors of *C. elegans* to identify multiple receptor-ligand interactions simultaneously. Such system aimed to successfully differentiate active and non-active ligands in an efficiently with a simple to use procedure.

In order to achieve the project goal, aversive behavior of *C. elegans* was used to indicate a positive interaction. Since aversive behavior can be easily stimulated by aversive chemicals, for the convenience of our project, we decided to use chemical interactions to simulate ligand-receptor interactions. Based on the client statement, our system were composed of four major components: worm transfer, chemical spotting, behavior recording and data analysis. For each system component, we generated a number of alternative designs and performed a series of preliminary experiments to choose the optimal final design. For the final design, we decided to use wildtype males (*him-5* strain) as the animal model due to its high mobility and good response to aversive chemicals. We modified the PDMS stamp designed by last-year MQP team to transfer larger volume of chemicals to each spot. Finally, the team used classification machine learning algorithms to classify spot types based on the observed behaviors. The details regarding to our final design are listed in the following sections.

7.2 Description of Final Experiment

According to our preliminary experiments and final design decision, our final system used the*him-5* strain to explore to up to 96 different ligand chemicals for the ligand-receptor identification screening test. Testing plates (NGM plates) are prepared prior to the test day. Three to six plates of *him-5* male strains were generated by crossing three*him-5* hermaphrodites to eight*him-5* males in each plate four to five days ahead the recording experimental day. The plates are stored in 15˚C. During the recording day, *him*-5 males are isolated to new clean seeded or unseeded plates. The amount of *him-5* males depended on the size of the testing plate. For example, approximate 100 worms are needed for 60 mm petri dish; at least 1000 worms are needed for a full test on an Omni Tray plate. The isolated worms are rinsed off from the plates with M9 solution and transferred to a 2.5 mL Eppendorf tube. The worms are washed two more times with M9 solution and left to settle. Once the worms are settled to the bottom of the tube, the worms could be distributed to the test plate. 25 µL of worms in M9 solution are then pipetted to the four corners and the center of the plate. Larger volumes may require for larger plates or in the case that not enough worms are transferred into the plate. The plate with worms are placed in hood for the M9 solution to evaporate the solution and to allow the worms to spread throughout the plate. The plate remains in the hood for approximately 15 minutes. The test chemicals are loaded into the well plate. Once the worms are evenly distribute, the test ligand chemical spots can be transferred to the test plate with the PDMS stamp.

Once the plate is set-up, the plate can be placed under the camera to record a video using a MATLAB program called *gVision* to record the animal behavioral. A detailed instruction on how to record a video using *gVision* can be seen in Appendix V. For our final experiment, the resolution used was 1280x1024, the frame rate was 3 frames per second and the video was recorded for 10 minutes.

7.3 Description for Final Data Analysis

Once the experiment has been completed and the video has been recorded, the video can be analyzed for the different types of animal behavior. The process involves opening the tracking program

developed by Navin Pokala. Specific instructions on how to set-up the program can be found in Appendix IV.Navin Code Procedure. Once the program has been set up, the directory in MATLAB has to be opened to where the video file was stored. The tracking script can be inputted to start the tracking program. A few inputs will be asked of the user and more information on that can be found in Appendix V. *gVisionProcedure*. After all the inputs are finished, the software will analyze the video for the animal behavior. Once the program is finished, it will generate a number of files with the data that was processed. The linked tracks file is the one that will be used in the next step to identify the different spots. A MATLAB script that was developed by the team will be executed on the linked tracks to process the data that was generated from the tracker. Once the script has been started, a prompt will appear to ask the user to click on the spots of interest from the experiment. Details on these scripts are located in Appendix VI. Code and Appendix VII. Code Procedure.

Chapter 8: Conclusions and Recommendations

8.1 Overview of Project and Objectives

With the need of a rapid ligand-screening system for drug market, the overall goal of our project was to design a system that utilizes the behaviors of *C. elegans* to identify multiple ligand-receptor interactions simultaneously.

The goal of our project was broken down to three primary objectives. First of all, this system must be reliable meaning it can not only generate conclusive results but also do so repeatedly. Having conclusive results referred to the ability of the system to accurately differentiate positive and negative interactions in a large testing array. Also, this system needed to consistently produce such results. Secondly, the system we designed must be easy to use. One of the limitations of the current cell-based ligand screening system is painstaking assay preparation process. With the expensive automated machine, the cell-based assay and data analysis only take about half hour; however, each assay needs days to prepare. We aimed to make our system less complex and labor-intensive. Thirdly, the system should be rapid. Since our ultimate goal was to accelerate the drug design process by generating an efficient ligand-receptor screening system, each experiment run should be able to test for multiple ligands in a short period of time.

To achieve our objectives, we have went through a series of design process and generated a final design. Based on the result of the final validation experiment, we concluded our system capable of correctly identifying behavioral response of *C. elegans* to the chemical spots in an array of different chemicals. Overall, our system resulted in an average confidence level of 68.8%. The relatively low accuracy of our system was due to the false positive results from the system due to the spontaneous behavior of the wild type strain. The spontaneous reversals gave a false positive result to neutral stimuli. We believe that these results can be improved with the use of a selectively responsive strain of *C. elegans.* The false positive results would be mitigated with the use of this strain because they do not show the spontaneous reversals.

8.2 Limitations

In the goal of accelerating ligand-receptor screening process, we have designed a system that can successfully differentiate aversive and neutral stimuli68.8% of time using behaviors of *C. elegans.* Despite our system have satisfy most of the objectives, there were numbers of limitations that can be further improved.

First, it was difficult to generate accurate results with low number of events in an experiment run that involved large number of ligands (96 spots). To testify our system, we have performed experiments on 9, 16, and 96 chemical spots. For 9 spots arrays, we used about 50-60 worms each plate which resulted in a number of events ranging from 8,000 to 45,000. This range allowed us to find patterns in the behaviors and generate machine learning models. For 16 spots arrays, we used about 100-200 worms per plate which yielded about 100,000 events which allowed us to successfully classify spot types with an average confidence level of 68.8%. For the 96 spots array we have run, we used 300 worms; however, there were only two spots had enough number of events to run data analysis. Therefore, we hypothesized we needed at least a 1000 worms in order to have enough number of events to yield accurate results. Potentially, we can overcome this limitation by using liquid culture of worms instead of picking worms from agar culture. However, the use of liquid culture can result in variation in animal behavior due to different cultivation conditions (Hart, 2006). In this case, we will have to generate a new series of training data using liquid culture to create a new machine learning models.

Second, manual picking wildtype males was very labor-intensive and time-consuming. As mentioned, we needed about thousands of worms for the 96 spot array but manual picking such a large animal size is implausible. Although mutation at the gene *him-5* increases the frequency of males in the offspring, the produced males are always mixed with substantial amount of hermaphrodites in the same population. Since there is no efficient way of separating the hermaphrodites from the males, we had no choice but to manually picked males onto a separate plate. Therefore, better male sorting methods needed to be created to make our system more efficient.

Third, our system had limited ability to produce reproducible results. The difficultly to create reproducible results might be caused by the wildtype males utilized in our experiments. Wildtype males exhibited high mobility and good response to aversive stimuli; however, these worms still demonstrated a response to non-specific stimuli. Because of this, there might be substantial counts of random behaviors and spontaneous reversals that were not caused by glycerol spots. These non-selective behaviors compromised the accuracy of the machine leaning program, which was lower than the average confidence level of the cell-based system (Caers *et al*, 2014). The cross of *eat-4* and *him-5* was ideal to use because it would be both highly mobile and selectively responsive. However, due to the complications with the crossing of the two strains, this option did not come available with the time we had. In addition to the problem with animal models, research indicated behavioral assays are difficult to control. Behaviors of *C. elegans* can be affected by different feeding status, cultivation conditions and environmental changes. Specifically, little factor can cause behavioral variations such as the size of the bacterial lawn, the type of bacteria, the agar, and ambient conditions can affect behavioral response (Hart, 2006). Lastly, the inconsistency in our results can also be caused by the variation in the number of events collected in a single experiment. Although we placed the same number of worms to every testing plate and we allowed them to evenly distribute, the actual times these worms ran into a spot were very different and hard to control. One way to increase the number of events would be to prolong the recording time. However, the supplemental experiments indicated that glycerol completely diffused into the agar after 10 minutes independent of its concentration and agar stiffness. This meant worms were no longer stimulated by glycerol after 10 minutes. We predicted this diffusion issue can be solved using peptide-receptor pairs because the average large molecular weight of peptides.

8.3 Future Recommendations

There are some areas that can be considered in the future to improve the results. First, a sensory-deficient strain of *C. elegans* needs to be tested. The initial idea was to use a selectively responsive strain of *C. elegans* to test the screening system. However, the strain was not used due to the time constraint on

the project and the unsuccessful attempt to create a selectively responsive strain with an increased probability of producing males by crossing the *eat-4* strain to *him-5* strain. The final system was tested with *him-5* males and their behavioral response was not desirable because they are similar to the wild type behavior. The *him-5* males display spontaneous reversals that occur without encountering any aversive stimuli and these behaviors can impact our results. Second, the system needs to be tested with a known ligand-receptor pair to test if the system is successful in identifying the correct ligand. The ligand-receptor pair was not obtained due to limited resources and time constraint on the project. Also, it was pointless to get the receptor unless the selectively responsive strain was created successfully. Another recommendation for the future is to have ligands expressed in a bacterial lawn instead of a liquid solution. The use of the bacterial lawn eliminates the problem of diffusion in the agar and could potentially generate much better results. Finally, the machine learning model can be improved to provide more accurate and reliable results. This can be done by collecting more training data and generating models with more training data. Due to time constraints, only three trials using the 4 x 4 spot array were tested with the model. The data sample size needs to be increased dramatically and the 96 spot array should also be tested. Additionally, there can be a deeper examination into the data sets best for training data. There can be more distance bins, different sets of frames and different combinations of these variables. The original Navin code distinguished between about 21 different behaviors and it is possible to examine each specific behavior rather than classifying these behaviors into four general classes. Built into the model generation function in MATLAB it is possible to control different factors in the model generation like the weights attributed to the different observations which could further improve the code.

References

Agresti, A. An Introduction to Categorical Data Analysis. John Wiley and Sons, 1996.

Albrecht, D. R., &Bargmann, C. I. High-content behavioral analysis of caenorhabditis elegans in precise spatiotemporal chemical environments. *Nature Methods*,2011; *8*(7), 599-605.

Albrecht, D.R. (2011). Arena Worm Tracker. Retrieved from http://arenawormtracker.blogspot.com/

Alberts B, Johnson A, Lewis J, Raff M, Roberts K et al. (2002) Molecular Biology of the Cell, 4th edition. *Garland Science*. Obtained from http://www.ncbi.nlm.nih.gov/books/NBK26910/.

Altun Z & Hall D. Nervous system, general description. *WormAtlas,* 2011.

Bargmann C. Chemosensation in *C. elegans*.WormBook, 2006.

Bell, J., & Safari Books Online. Machine learning: Hands-on for developers and technical professionals. *Wiley*, 2015.

Caers J, Perymen K, Suetens N, Temmerman L, Janssen T, Schoofs L, Beets I. Characterization of G Protein-coupled Receptors by a Fluorescence-based Calcium Mobilization Assay. *Journal of Visualized Experiments,* 2014; 89, 1-10.

Cai, T., Krause, M. W., Odenwald, W. F., Toyama, R., &Notkins, A. L. The IA-2 gene family: homologs in Caenorhabditiselegans, Drosophila and zebrafish. *Diabetologia*, 2001; 44, 81-8.

Culetto E, &Sattelle D. A role for caenorhabditiselegans in understanding the function and interactions of human disease genes. *Human Molecular Genetics,* 2010; *9*(6), 869-877.

Dunford P, Williams K, Desai P, Karlsson L, McQueen D *et al*. Histamine H4 Receptor Antagonists are Superior to Traditional Antihistamnes in the Attenuation of Experimental Pruritus. *Journal of Allergy and Clinical Immunology*, 2007; 1: 176-183.

Esposito G, Amoroso M, Bergamasco C, Di Schiavi E, &Bazzicalupo P. The G protein regulators EGL-10 and EAT-16, the G.sub.i [alpha] GOA-1 and the G.sub.q [alpha] EGL-30 modulate the response of the *C. elegans* ASH polymodal nociceptive sensory neurons to repellents. *BMC Biology, 2010; 8,* 138.

Evans, T. Transformation and Microinjection. *Wormbook, The C. elegans Research Community*, 2006.

Fisher, R. A. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, Vol. 7, pp. 179–188, 1936.

Fox S, Human Physiology, 12th ed. New York, NY: McGraw Hill, 2011.

Frooninckx L, Rompay L, Temmerman L, Sinay E, Beets I *et al*. Neuropeptide GPCRs in *C. elegans*. *Frontiers in Endocrinology*, 2012; 3: 1-18.

Fujita-Jimbo E, Yu Z, Li H, Yamagata T, Mori M *et al*. Mutations in Parkinson Disease-Associated, G-Protein-Coupled Receptor 37 (GPR37/PaelR) Is Related to Autism Spectrum Disorder. *PLoS ONE*, 2012; 7: 1-6.

Geng W, Cosman P, Berry C, Feng Z, & Schafer W. Automatic tracking, feature extraction and classification of C. *elegans* phenotypes. *IEEE Transactions on Biomedical Engineering,* 2004; *51*(10), 1811.

Gottfries C, Frederiksen and Heilig M. Neuropeptides and Alzheimers' Disease. *European Neuropsychopharmacology*, 1995; 5: 491-500.

Guo Z, Hart A, &Ramanathan S. Optical interrogation of neural circuits in caenorhabditiselegans. *Nature Methods*, 2009; *6*(12), 891-896.

Hart, Anne C., ed. Behavior (July 3, 2006), WormBook, ed. The*C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.87.1, http://www.wormbook.org.

Haughn M, Masone D, &Yalcin.Behavior Analysis for Rapid Screening of Molecular Interactions. *Worcester Polytechnic Institute Major Qualifying Project Report*, 2014.

Hilliard M, Apicella A, Kerr R, Suzuki H, Bazzicalupo P, & Schafer W In vivo imaging of C. elegans ASH neurons: Cellular response and adaptation to chemical repellents. *The EMBO Journal, 2005; 24*(1), 63-72.

Hruby V. Designing Peptide Receptor Agonists and Antagonists. *Nature Reviews*, 2002; 1, 847-858.

Huang T, Chang H, Hsu C, Kuo W, Chang K, & Hsueh F. Targeting Therapy for Breast Carcinoma by ATP Synthase Inhibitor Aurovertin B. *Journal of Proteome Research*, 2007; 7, 1433-1444.

Huang, K. M. *Tracking and analysis of C. elegans behavior using machine vision.* University of California, 2008.

Hyman S. (2005). Neurotransmitters. Current Biology, 2005; 15: R154-R158.

Katritch V, Cherezov V, and Stevens R. Structure-Function of the G-protein-Coupled Receptor Superfamily. *Annu. Rev. PharmacolToxicol.*, 2013; 53: 531-556.

Kobilka B. G protein coupled receptor structure and activation. *BiochimicaetBiophysicaActa*, 2007; 1768: 794-807.

Lee C, Huang H, Juan H. Reviewing Ligand-Based Rational Drug Design: The Search for an ATP Synthase Inhibitor. *International Journal of Molecular Science*, 2011; 12, 5304-5318.

Lee R, Sawin E, Chalfie M, Horvitz H, & Avery L. EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in *caenorhabditiselegans*. *Journal of Neuroscience*, 1999; 19(1), 159-167.

Lodish H, Baltimore D, Berk A, Zipursky S, Matsudaira P et al. (1997) Molecular Cell Biology, 3rd edition. Chapter 24. *Scientific American*. ISBN 0-7167-2380-8.

Lott III, G. (2010). GVision-Scientific Video & Image Acquisition. Retrieved from http://gvision-hhmi.sourceforge.net/

Ludwing M. Are Neuropeptides Brain Hormones?. 2012. Retrieved from http://sbmsintranet.bms.ed.ac.uk/sources/Briefings/Ludwig%20-%20Are%20Neuropeptides%20Brain%20Hormones.pdf.

McKay R, McKay J, Avery L, & Graff J. C elegans: A model for exploring the genetics of fat storage. *Developmental Cell*, 2003; *4*(1), 131.

Mertens I, Vandingenen A, Meeusen T, De Loof A, Schoofs L. Postgenomic Characterization of G-Protein-coupled Receptors. *Pharmacogenomics*, 2004; 5(6), 657-672.

Rokach, L &Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc.

Salon J, Lodowski D, Palczewski K. The significance of G protein-coupled receptor crystallography for drug discovery. Pharmacological Reviews, 2011; 63:901-937.

Sassa T, Murayama T, & Maruyama I. Strongly alkaline pH avoidance mediated by ASH sensory neurons in *C. elegans*. *Neuroscience Letters,* 2013*; 555*, 248.

Schöneberg T, Schulz A, Biebermann H, Hermsdorf T, Römpler H *et al*. Mutant G-protein-coupled Receptors as a Casue of Human Diseases. *Pharmacology & Therapeutics*, 2004; 104: 173-206.

Stiernagle, T. Maintenance of *C. elegans.WormBook,* 2006.

Swierczek, N. A., Giles, A. C., Rankin, C. H., & Kerr, R. A. High-throughput behavioral analysis in *C. elegans*. *Nature Methods,* 2011; *8*(7), 592-598.

Synapse transmission (KIN450-Neurophysiology). (2013). http://kin450-neurophysiology.wikispaces.com/synaptic+transmission

Teng M, Dekkers M, Ng B, Rademakers S, Jansen G *et al*. Expression of Mammalian GPCRs in *C. elegans* Generates Novel Behavioral Response to Human Ligands. *BMC Biology*, 2006; 4: 22.

Thurmond R, Gelfand E, and Dunford P. The Role of Histamine H1 and H4 Receptors in Allergic Inflammation: the Search for New Antihistamines. *Nature Reviews Drug Discovery*, 2008; 7: 41-53.

What are the parts of the nervous system? (2014). Obtained from http://www.nichd.nih.gov/health/topics/neuro/conditioninfo/Pages/parts.aspx

Appendices

Appendix I. Procedures for Preliminary Experiments

Pipetting vs. Stamping

1. After placing the desired *C. elegans* onto the OmniTray plate [Nunc #242811] (see procedure of Worm Spreading/Distribution in Appendices), perform the following protocol:
    a. Pipetting: place the spot sample reference sheet in the bottom of the plate, and pipette 1μL of desired chemical to the corresponding spot on the plate.
    b. Stamping: place 90 μL of each of the desired chemical in a well of a 96-well-plate [Chemglass-CLS-3506-S5]; use the PDMS stamp to transfer all chemicals to the OmniTray plate

    Record worm behavior with gVision MATLAB program for 10 minutes.

    Analyze data with Navin code and MATLAB program.

Rectangular OmniTray Plate vs. 60-mm Circular Petri Dish Plate

1. Obtain, wash and deliver enough *C. elegans* to rectangular OmniTray plate and/or 60-mm circular petri dish plate as protocol of *Worm Spreading/Distribution*
2. Pipetting same amount (1 μL of chemical for each spot; spot same number of spots) to each type of plates
3. Record and analyze the worm behavior with MATLAB programs

Generation of Selectively Responsive Male Strain of *C. elegans*

eat-4;him-5:

1. Prepare at least three plates of cross between three L4 *eat-4* hermaphrodite and eight young males of *him-5* (Parental plate 1 or P1 generation)
2. Remove the *him-5 C. elegans* from the P1 plate the next day
3. In 4-5 days, prepare at least three crosses between three L4 *eat-4* hermaphrodite and eight young male progeny from P1 (Parental plate 2 or P2 generation)
4. Remove the eight young male progeny from P1 on P2 plate the next day
5. In 4-5 days, clone at least 25-30 plates from each of the P2 progeny plates with the presence of male *C. elegans* (select L4 or younger; F1 generation)
6. In 4-5 days, clone at least 25-30 plates from each of the F1 progeny plates with the presence of male *C. elegans* (select L4 or younger; F2 generation)

7. Repeat Step 6 for F2 and F3 progeny plates (F3 and F4 generations, respectively) to aim for stable *eat-4;him-5* line
8. PCR eat-4 from male progeny from F4 to confirm the genotypic characteristic of the cross line

Generation of *DRA211* males or *DRA212* males (Selectively responsive strain)

1. Prepare 10+ L4 or young adults of 211 and/or 212 hermaphrodites on corresponding plates
2. Wrap the plates tightly with parafilm
3. Place the plates into 33C water bath for 3 hours of heat shock treatment (plates are facing up--the lid is closer the air)
4. Take out the plates and store properly
5. After 1-2 days, cross the males with the fluorescent markers (8 males) from the heat shock plates to L4 hermaphrodites of their corresponding strains (3 L4 with fluorescent markers)
6. Maintain the strains through crossing method as mentioned in Step 5

Worm Spreading/Distribution

1. Pick the desired *C. elegans* to 1-1.5 mL of M9 solution in 2 mL eppendorf tube (or wash off a full plate of young adult/male *C. elegans* with 1-1.5 mL of M9 solution to 2 mL eppendorf tube)
2. Let the *C. elegans* settle in the bottom of the eppendorf tube.
3. Take off 0.75-1 mL of washing M9 solution from the tube, and replace with fresh M9 solution.
4. Gently invert the tube for washing.
5. Repeat Steps 2-4 at least twice.
6. After wash, let the *C. elegans* settle in the bottom of the eppendorf tube.
7. Pipette 25 µL of *C. elegans* to each of the center and spots close to the four corners of the rectangular OmniTray plate
8. Let the M9 dry and the *C. elegans* evenly distribute on the plate (>10 min)

Appendix II. PDMS Stamp Procedure

1. Obtain a weight boat and place it on top of a scale and zero it.
2. Measure 180 grams of PDMS on the weigh boat.
3. Carefully mix in 20 grams of PDMS plastomer curing agent into the PDMS.
4. Stir the PDMS with a spatula vigorously to mix the curing agent. This should be done for about 10 minutes to ensure that the PDMS is mized uniformly for even cross-linking.
5. Place the PDMS mixture into a bell-jar dessicator that is connected to a vacuum pump. Make sure the jar is sealed by the vacuum and leave it in for 1 hour to get rid of the bubbles.
6. While the PDMS is in the vacuum, get the 96 well plate [Chemglass-CLS-3506-01] and cut up cardboard and tape them around the plate to create walls. This is done to add thickness to the top of the stamp. Figure 41 shows an example.
7. After an hour, take the PDMS out and pour it into the well plate.
8. Place it inside the vacuum again for 5-10 minutes to get rid of bubbles during pouring.
9. Afterwards, put the mold into an incubator at 56^{o}C for 3 hours.
10. After 3 hours, take the PDMS out of the incubator and let it cool down for a few minutes.
11. Once the PDMS is cooled, take it out of the mold by pulling it out. It takes some force to pull out but be careful not to break it while doing it.

Figure 41: 96 well plate with cardboard taped around it to create walls to add thickness to the stamp

Appendix III: eat-4 Genotyping PCR for eat-4; him-5

1. Worm Lysis
2. Label the PCR tubes [Olympus Plastics #24-705] (one sample per tube)
3. Add 20 µL of Lysis buffer with pro-K [Invitrogen] to each PCR tube
4. Transfer worm sample to the lysis buffer
    a. Single Worm Lysis
        i. Pick 1 male worm to the PCR tube from Step 2(1 worm for each sample each tube) and make sure the worm is in the liquid
    b. Full Plate Worm lysis
        i. Rinse the sample plate with M9 solution and transfer to a new clean 2.5 mL Eppendorf tube
        ii. Once the worms are settled to the bottom of the Eppendorf tube, pipette 20 µL of worm pellets to the PCR tube with lysis buffer from Step 2
5. Place the PCR tube to thermocycler with the following the temperature setting of lysis cycle:
    a. 65°C for 1 hour
    b. 95°C for 15 min
    c. 4°C forever

II. Genotyping PCR

1. Set up the master mix for PCR (per reaction)

Materials without Apex Master Mix PCR Product	**Amount (µL)**
DNTPs [Roche]	1.25
PCR buffer [Roche]	5
Taq Polymerase [Roche]	1
Template DNA (worm lysis from I)	5
10µM *eat-4* (ky5) forward primer [IDT # 115797287]	1.5
10µM *eat-4* (ky5) reverse primer [IDT #1157972288]	1.5
PCR H2O [UltraPureGentrox #18-195]	34.75
TOTAL	50

Materials with Apex Master Mix PCR Product	**Amount (µL)**
Apex Hot Start Master Mix Blue with Apex Buffer I (including DNTPs, Buffer, Taq polymerase)	25
Template DNA (worm lysis from I)	2.5
10µM *eat-4* (ky5) forward primer	2.5
10µM *eat-4* (ky5) reverse primer	2.5
PCR H2O	17.5
TOTAL	50

2. Place the PCR tubes to the thermocycler
    a. PCR Cycle for materials without Apex Master Mix PCR product
        i. 95°C for 2 min
        ii. 95°C for 30s
        iii. 53°C for 30s
        iv. 72°C for 1min15s
        v. repeat Steps ii to iv for 25-30 cycles
        vi. 72°C for 7 min
        vii. 4°C forever
    b. PCR Cycle for materials with Apex Master Mix PCR product
        i. 95°C for 15 min
        ii. 95°C for 30s
        iii. 53°C for 30s
        iv. 72°C for 1min15s
        v. repeat Steps ii to iv for 34 cycles
        vi. 72°C for 5 min
        vii. 4°C forever

III. DNA Gel Electrophoresis

1. Dissolve 1.5g agarose to 150 mL 1x TAE by heating up the solution with microwave
2. Let the solution cool down to approximate 50°C
3. At the meantime, assemble the gel rack and cones apparatus

4. When the agarose solution is cool enough, add a drop of EtBr dye, swirl till well mixed, and pour the agarose solution to the gel rack apparatus
5. Pop off the bubbles on the gel with sterile pipette tips
6. Let the gel solidify and carefully take off the cone.
7. Place the gel to electrophoresis apparatus and cover the gel with 1x TAE
8. Obtain 25 µL of PCR products for each sample from II respectively, and add 5 µL of loading dye and mix completely.
9. Load 5 µL of DNA standard latter to the gel
10. Load 15 µL of each sample to the gel with gel map recorded in the lab notebook
11. Run gel at 100 V and 400mA for approximate 1.5 hour or until the bands reach about 3 cm to the bottom of the gel
12. Image the gel
13. Expected fragment size:
    a. WT: ~1235 bp
    b. eat-4 mutant: ~621 bp

Appendix IV.Navin Code Procedure

1. To setup the Navin tracking software for the first time, place the folder containing the files in the desired placed.
2. Next, open up the file "navin_code_steup.m" in MATLAB. Change the path in the file to match where the folder was placed. Figure A42 shows what needs to be changed. Save the file and run it. This file must be executed each time the program is used when MATLAB is opened.

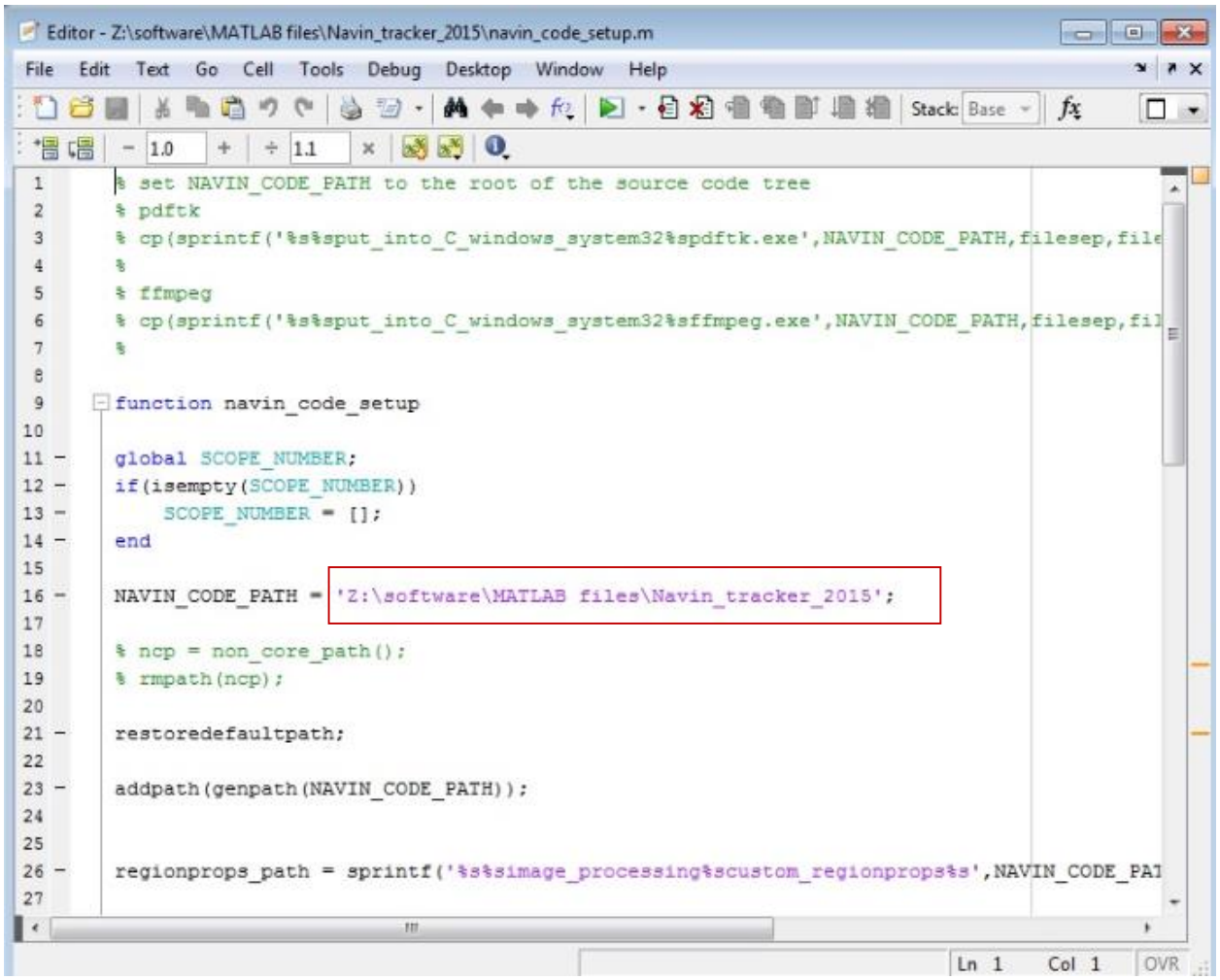

Figure 42: The red box indicates where the path of the folder is inputted

.

3. Place the video of the animal behavior in the same folder as the video used for scaling.
4. Find that folder in the MATLAB directory.
5. To run the tracking software, type in the following line of code :

TrackerAutomatedScript('animalvideo.avi', 'scale', 'holepunch' , 'holepunch.avi')

Replace the avi file names for your behavioral video and scale video (holepunch). FigureA43 shows an example.

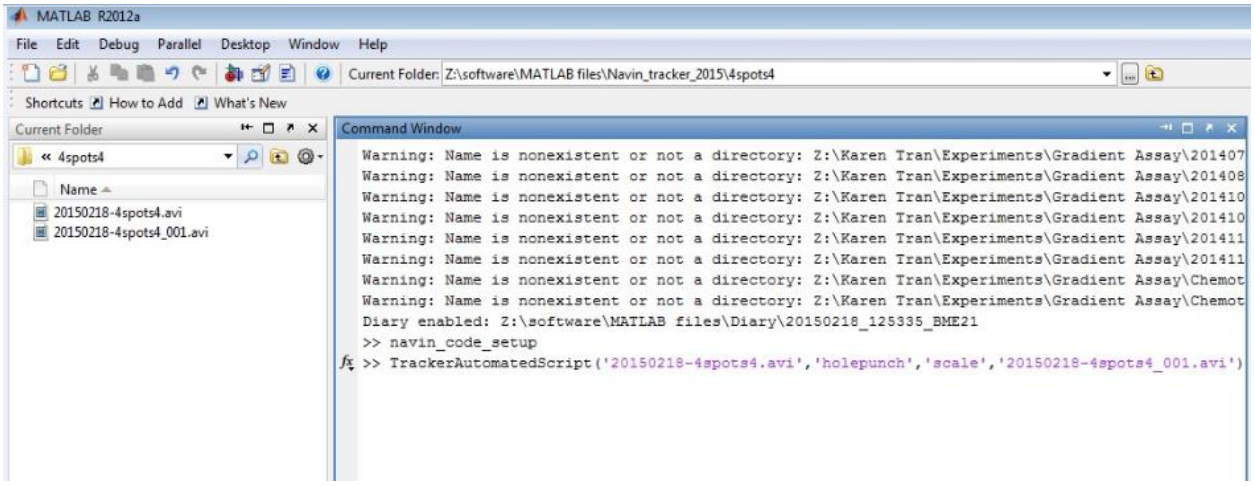

Figure 43: In the Figure, the current folder displayed is where the videos are located. In the command window, the TrackerAutomatedScript is inputted

6. Run the code and afterwards MATLAB will prompt you to indicate where the hole is that is used for scaling. Simply click around the circumference of the circle in the video and double click the left mouse button when finished. Figure A44, A45, and A46 shows examples of the scaling process with a hole punch.

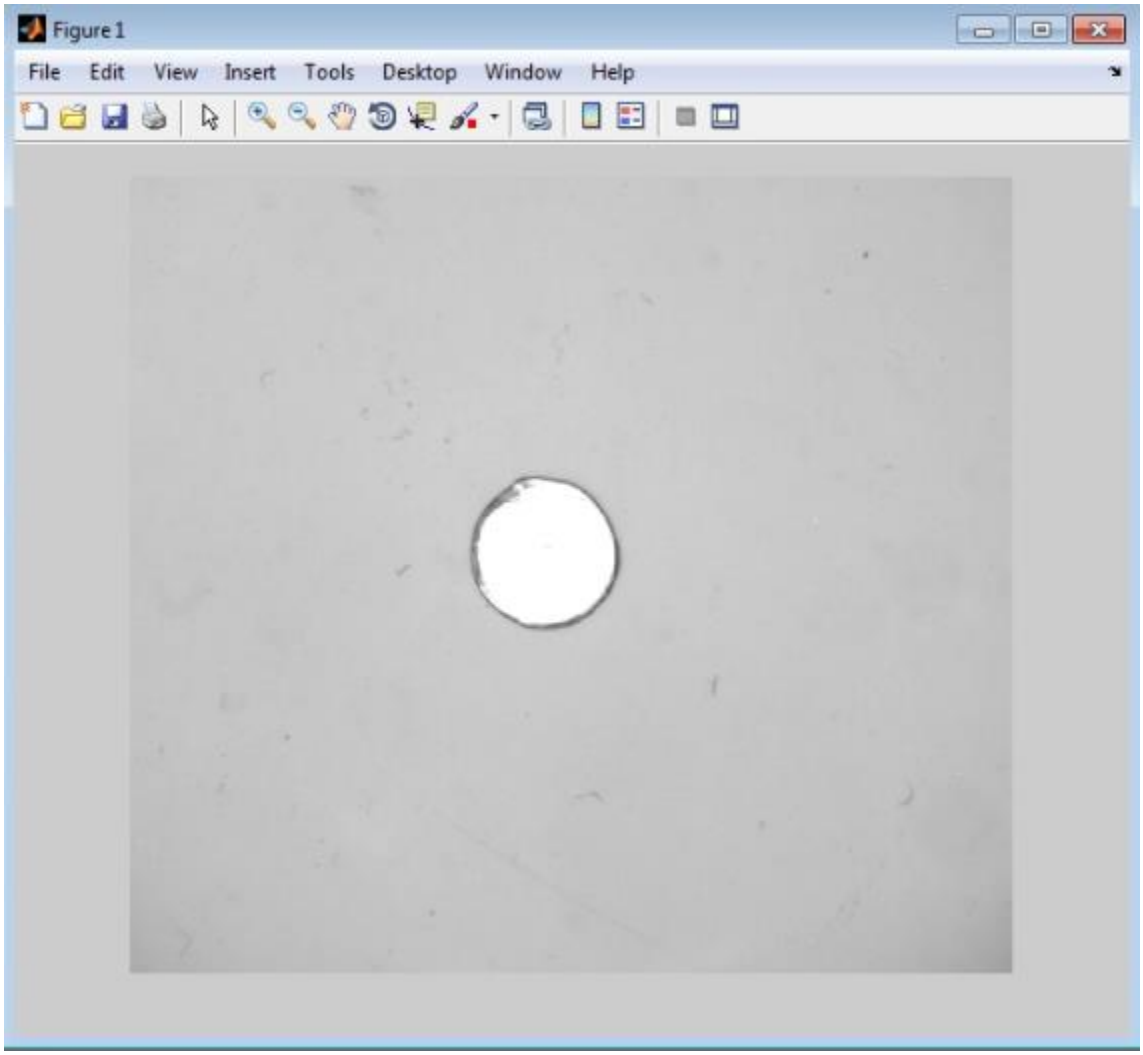

Figure 44: This Figure shows the hole punch that was recorded that will pop up

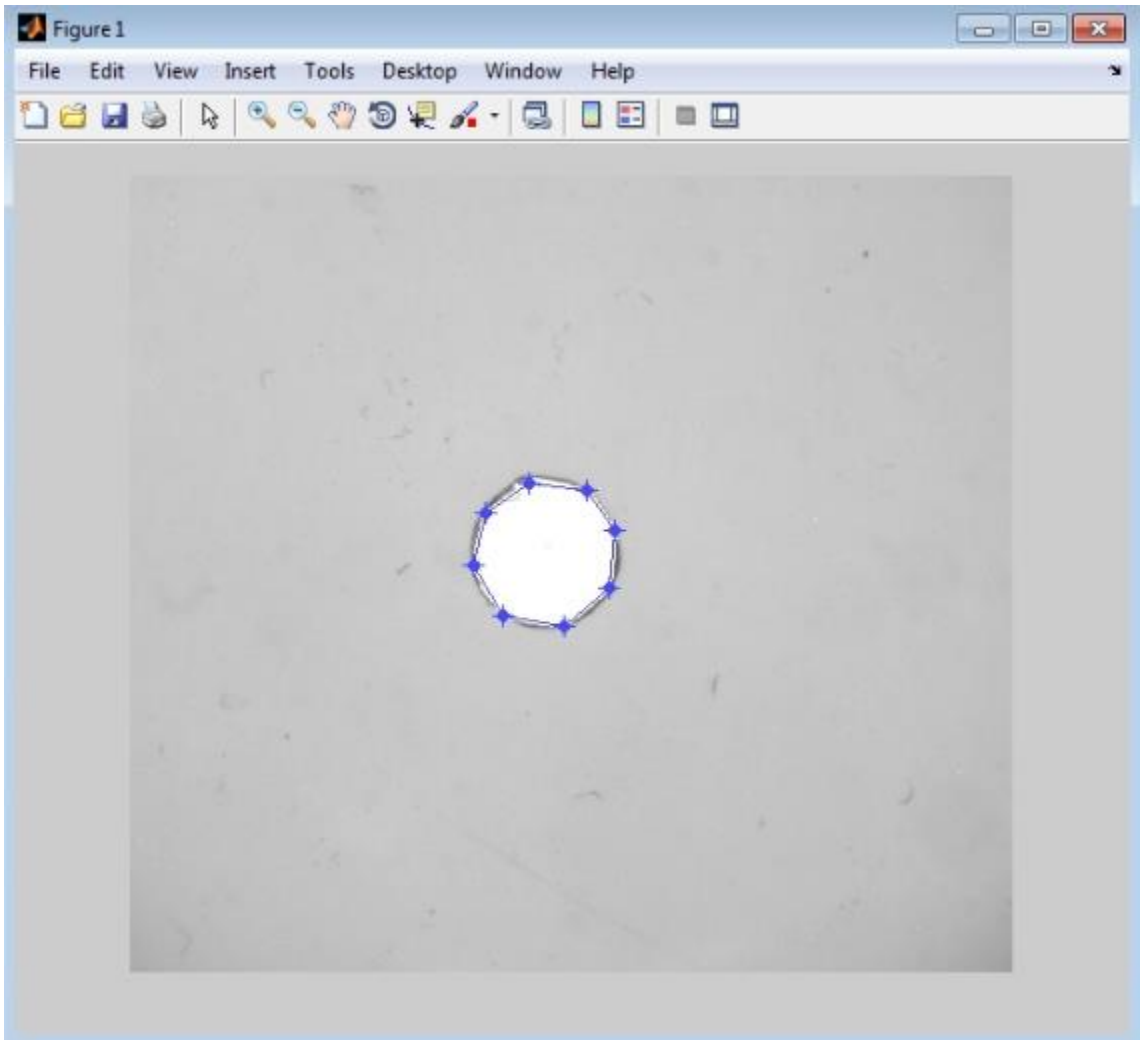

Figure 45: This Figure shows the points that had been labeled around the hole punch to define the circle

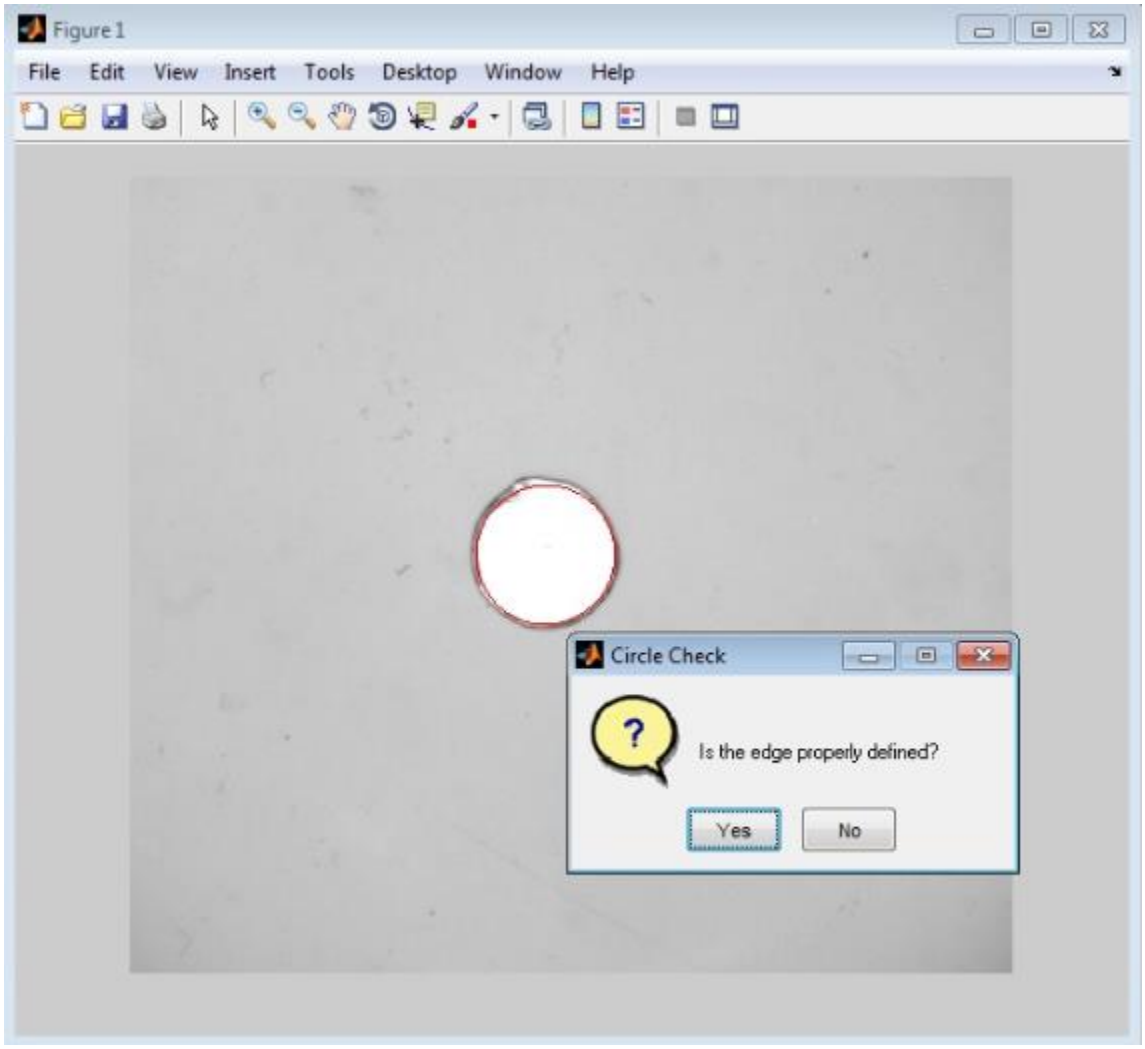

Figure 46; This Figure shows that the circle has been defined and will ask the user to confirm it

7. Afterwards, another prompt will pop up to locate the worms. The program can recognize some of the worms but usually manual input is required to get all the worms. FigureA47 and A48 show examples of the prompts that the user will encounter. To add a worm use the left mouse button. To remove a worm, use the right mouse button. Click the center mouse wheel button when finished.

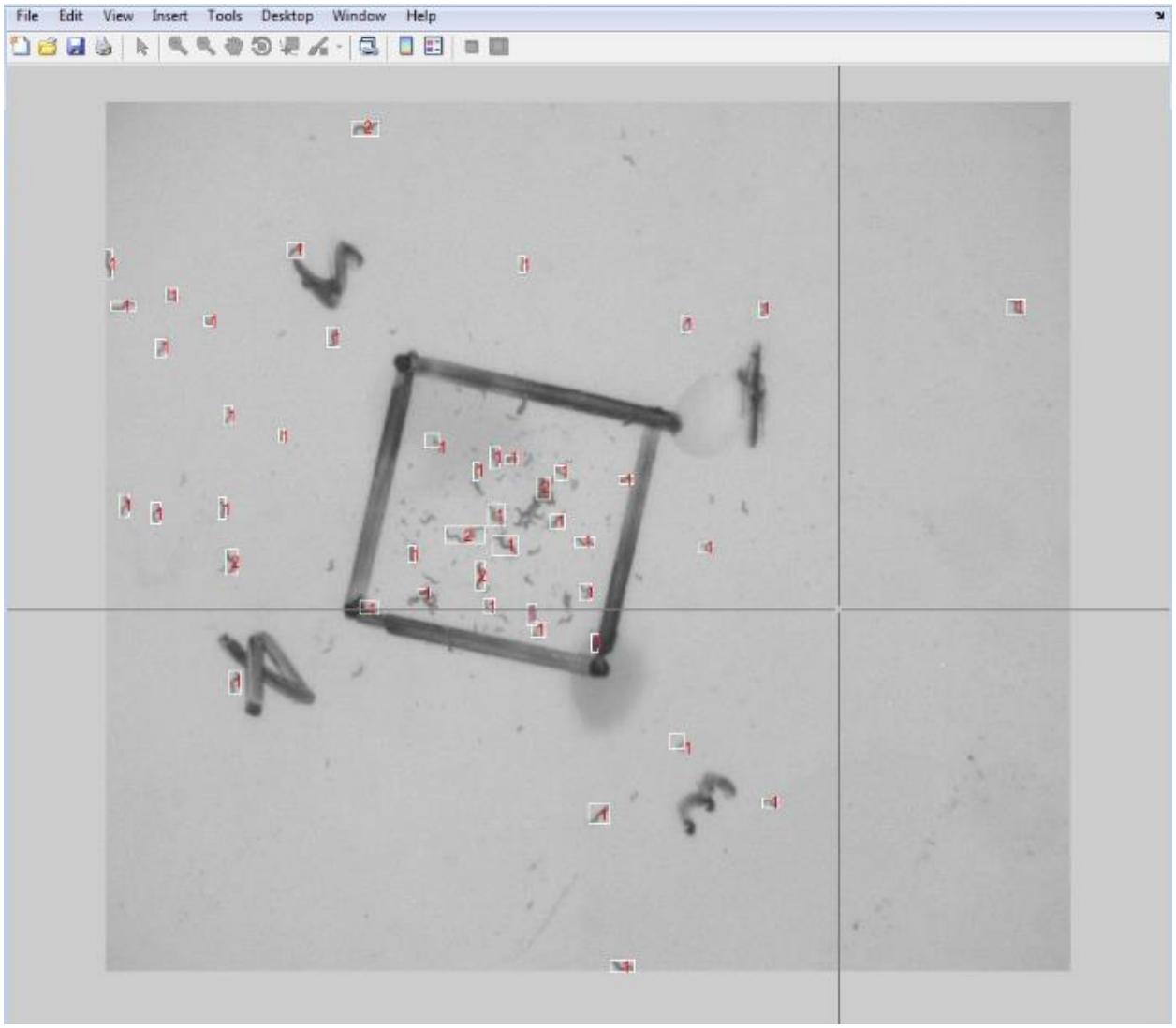

Figure 47: This Figure shows an example of manually indication of the worms. Each mouse click will generate a box around the worm

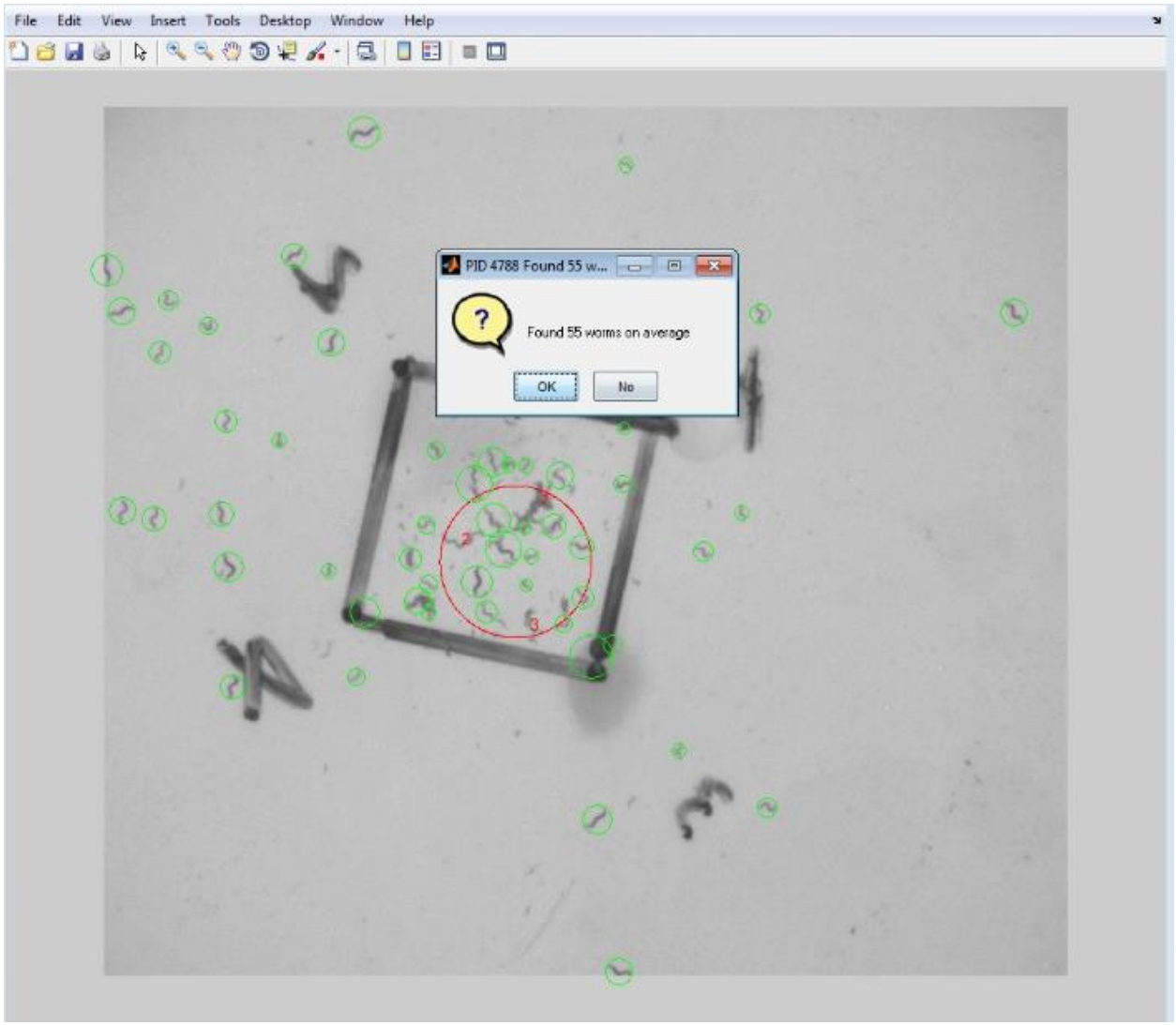

Figure 48: This Figure shows the window pop up that occurs after the user has finished spotting the worms. The program will ask the user to confirm the worms spotted. Clicking Ok will continue the script. Clicking No will prompt the window to manually label the worms again

8. After all the worms have been indicated and confirmed, the program will run on its own to track the different animals. The time it takes to finish tracking the video is variable and depends on a number of factors such as number of frames recorded, numbers of animals on the plate, and the resolution of the video.

Appendix V. gVisionProcedure

1. Open up MATLAB and type in "gVision" in the command window to open up the recording software. FigureA49 shows the window that will pop up.

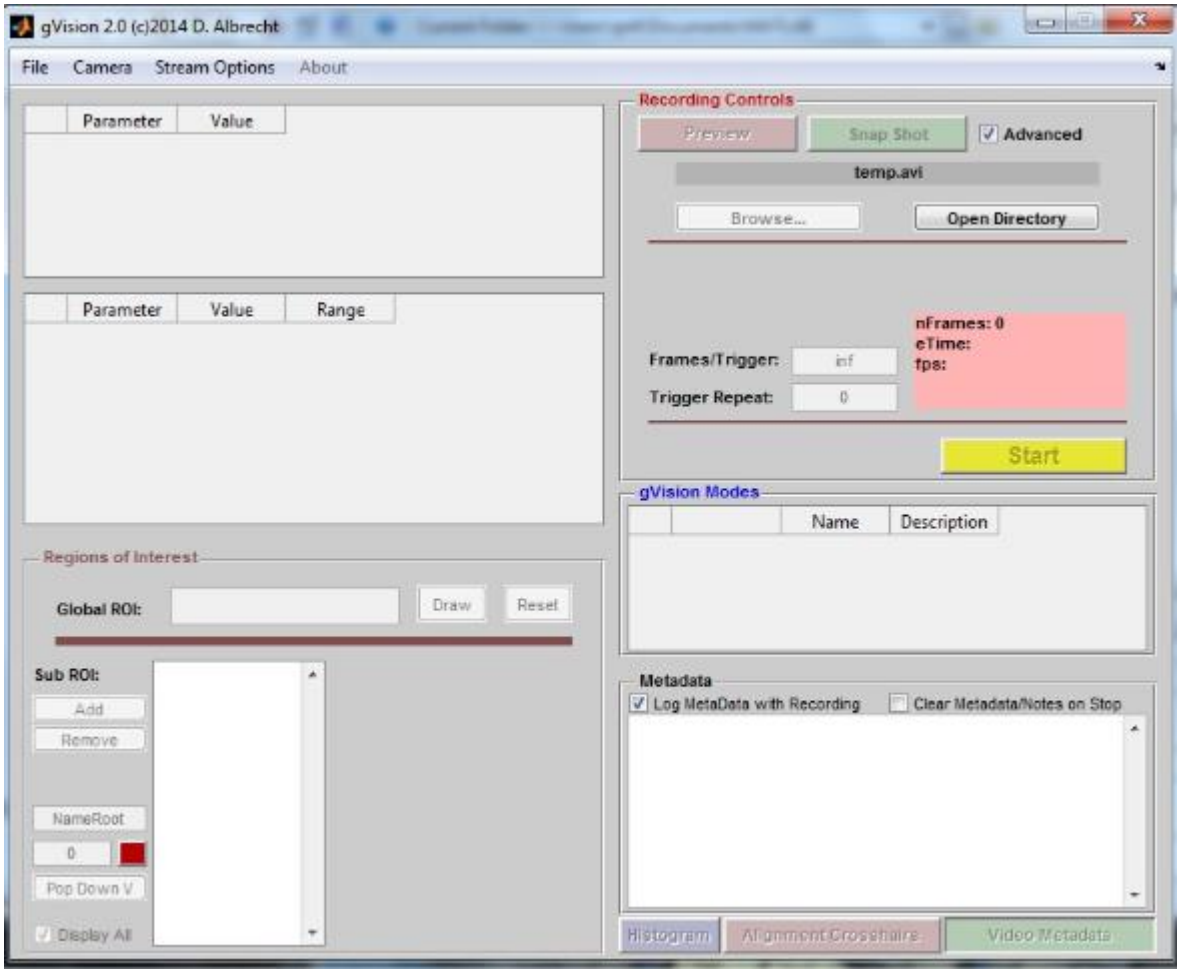

Figure 49: This Figure shows the gVision program window that pops up

2. Click file and click load state. This is to select the desired resolution for the recording. FigureA50 shows an example.

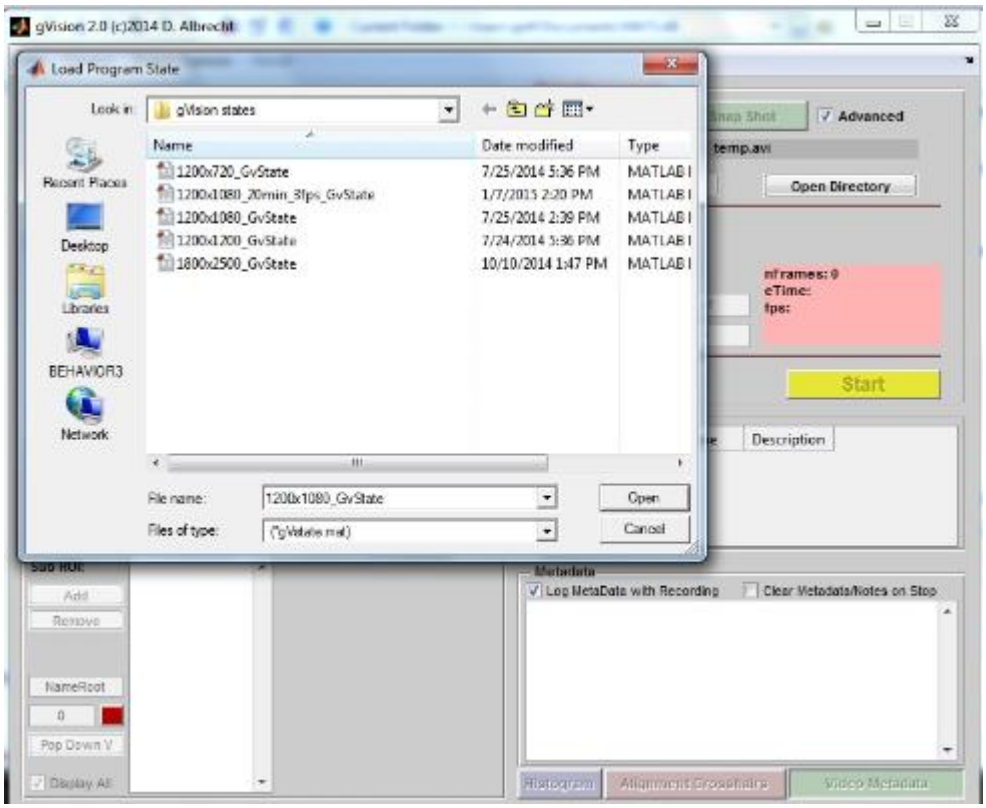

Figure 50: This Figure shows a number of files that correspond to different resolutions

3. Pick a state and the settings should load into gVision. FigureA51 and A52 shows examples of the windows that appear.

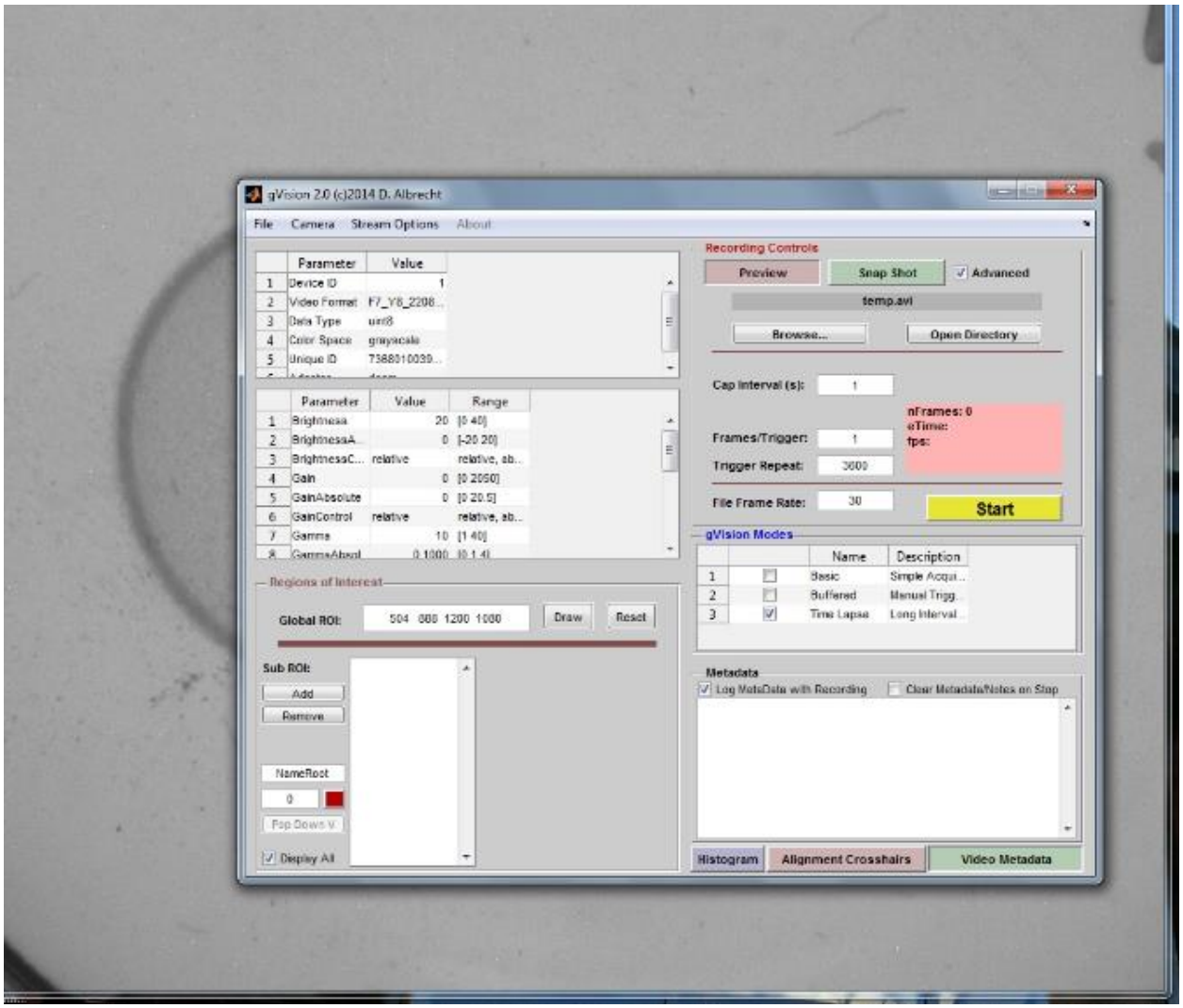

Figure 51: This Figure shows the settings has been loaded into gVision

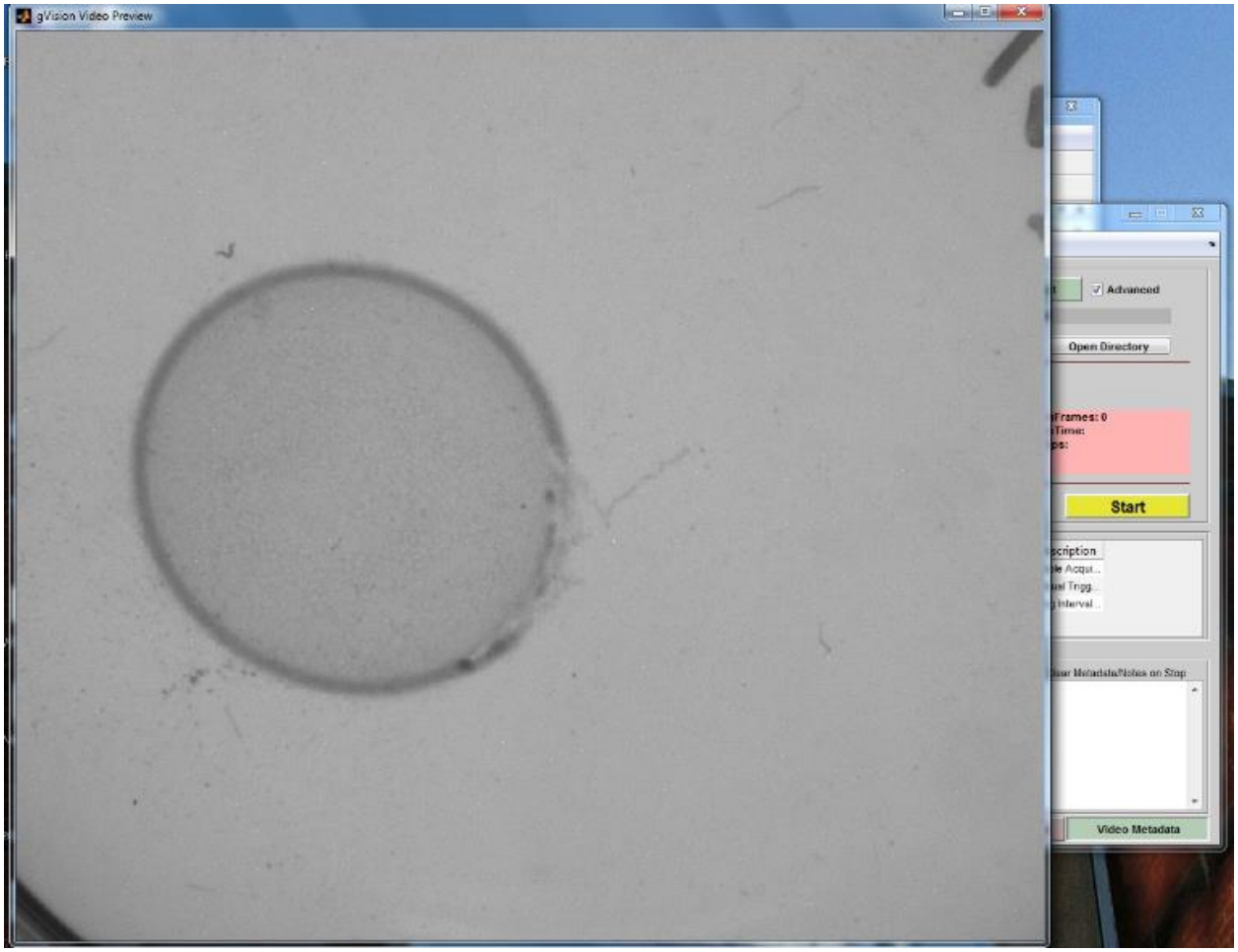

Figure 52

4. Place the agar plate with animals in them onto the platform under the camera. Adjust the camera and lighting to make sure the animals are zoomed in properly and that they are in focus.
5. Adjust settings on the gVision panel before recording. Figure A53 shows and describes what settings can be inputted.

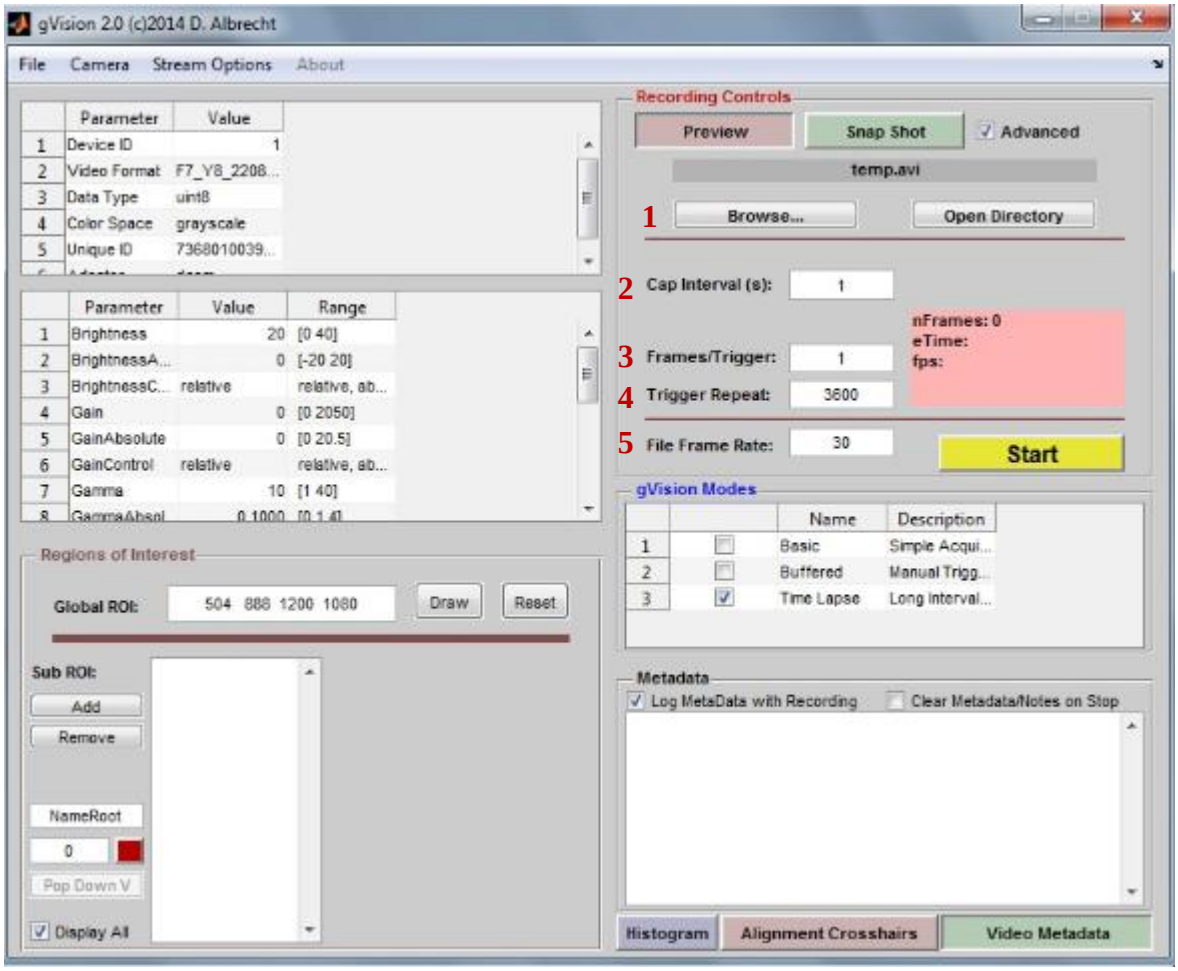

Figure 53: (1) Browse to the folder where you want the video file to be saved. (2) Number of frames recorded per second. (3) Numbers of frames you want per trigger. (4) How many times you want to trigger the recording? At 1 frame/trigger, this basically means how many frames you want to record. (5) The frame rate of the video output desired

6. Before starting the recording, click stream options and make sure the video compression is set to "motion jpg avi" to make sure the video is as small as possible. Also, uncheck the timestamp overlay. FigureA54 shows where these options are.

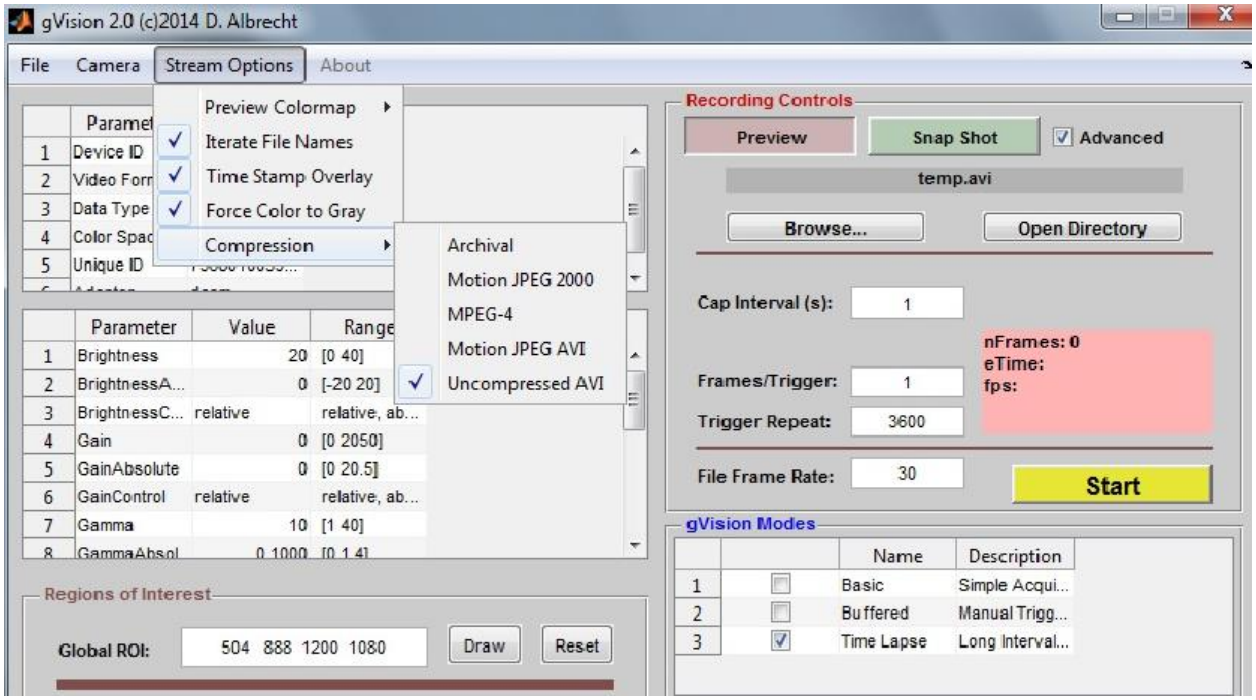

Figure 54: This Figure shows the stream options that needs to be changed before the recording

7. When ready, hit start to record a video of animal behavior.
8. When the video is finished, remove the plate of animals from the platform and place a agar plate with a hole punch.
9. Hit start again on gVision to record a short video of the hole punch to use for scale. Use the same settings and simply just stop the recording after about 10 seconds.
10. When finished, close the program and clean-up the area.

Appendix VI. Code

After processing the video through the Navin code, the team created MATLAB functions to allow us to examine the worm behavior relative to each of the chemical spots. This functions also placed the data into a useable form for the machine learning algorithms built into MATLAB. We decided to utilize the Classification Tree machine learning program for creating our models and to examine our data.

Part 1. Processing the data

The output from the Navin code gives us a structure with each column containing information on each event that occurs in the video. The Navin code collects data on the positions of the event, the frame number the event occurs in, the animal behavior state as a representative number and the pixel conversion. The Navin code was most interested in the frequency of events that occur at different time points throughout the experiment. We made functions called Processing, Training_Processing and Data Manipulation that allowed us to view the events relative to the center of each chemical spot. The struct2mat and safecat functions provided by Professor Dirk Albrecht were utilized to run the functions.

Processing reads in the MATLAB files from the Navin code, pulls out all of the positions, states, pixel conversation value and frame number. It also associates each event with a specific chemical spot if it is within a certain distance from the center of the spot. Finally, we associated the event with a behavioral class depending on the assigned state. We created four general classes of interest for these experiments: forward, pause, aversive and other (turn behaviors). The Processing code only removes the data that is unassociated with a particular chemical spot. Training_Processing varies from Processing by having an additional output which is a cell array containing the chemical state of each of the chemical spots. The commented code is below:

Processing code:

```
function [chemspots_mat,fulldata_mat] = Processing(linkedTracks_mat,mov)
%Program Writer: Rebecca Lee
%Processing intakes cell arrays containing the names of any number of
%linkedTracks files or behavior movie files. It then outputs two matrices.
%chemspots_matcotains the locations of all the chemical spots, the. fulldata_mat is matrix containing
%location, state, closest chemical spot number, distance to said chemical
%spot, pixel conversion value and a behavioral class number (1-4).

chemspots_mat = [];
fulldata_mat = [];

tic
num_tracks = length(linkedTracks_mat);

%This loop will go through each track and movie in the file then start building the
%final matrices and arrays.
```

```
for a = 1:num_tracks
   %Pull out the movie and the linkedTracks file of interest
track = linkedTracks_mat{a};
movie = mov{a};
load (track)

   %This component utilizes the movie that was inputted to create an
   %interacTable image for inputting the location of each of the spots.
   %The image is displayed, the chemical spots can be clicked in the
   %image, hit ENTER and then the code will be able to read in the spot
   %locations and then it will label the spots.
MovieObj = VideoReader(movie);
Mov = read(MovieObj, 1);
Mov = Mov(:,:,3); %Pulls the Blue from the Image to make image B&W
imshow(imadjust(Mov,[],[],2));
txt = 'Click on stimulus spots. When done, hit ENTER.'; title(txt);
   [X,Y] = ginput;
chem_spots = [X Y];
   [cs_row,cs_col] = size(chem_spots);
for b = 1:cs_row
      %Add label to image
text(chem_spots(b,1),chem_spots(b,2),['Spot ' num2str(b)], 'Color', 'r', 'FontWeight', 'bold');
end
chemspots_mat=vertcat(chemspots_mat,chem_spots);

   %Load in the different data columns of interest.
   x = struct2mat(2,linkedTracks,[],{'SmoothX'});
   y = struct2mat(2,linkedTracks,[],{'SmoothY'});
   s = struct2mat(2,linkedTracks,[],{'State'});
   f = struct2mat(2,linkedTracks,[],{'Frames'});
   x=double(x);
   y=double(y);
   s=double(s);
   f=double(f);
   %Convert each column out of single precision values into doubles, convert
   %from rows into columns and convert the state values to make them more
   %manageable.
   x=x';
   y=y';
   s=s';
   f=f';
   s = s*10;
   s = round(s);
   %All of the separate matrices are combined into the full_data matrix.
   %X AND Y ARE IN PIXELS
   %Insert the ratio of pixel to mm into the fifth column
   mm2p_ratio = linkedTracks.PixelSize;
pix = 1/ mm2p_ratio;
event_num=length(x);
   p = zeros(event_num,1);
p(:,:)=pix;
currentdata_mat = [x y s f p];

   %This are the state codes that are associated with different types of
   %behavior
```

```
all_beh = [11,10,47,57,71,51,31,41,20,30,40,50,70,80,74,75,43,34,53,35];
for d = 1:event_num
      % Add in Behavior class
beh_type = currentdata_mat(d,3);
location=find(all_beh==beh_type);
if location == 1
currentdata_mat(d,6)=1;
elseif location == 2
currentdata_mat(d,6)= 2;
elseif location >= 3 && location <=8
currentdata_mat(d,6)= 3;
else
currentdata_mat(d,6)= 4;
end
end

   %Find and associate with nearest chemical spot
   %Something SCREWY IS HERE
   [chem_r,chem_c]=size(chem_spots);
   [chemspot_r,chemspot_c]=size(chemspots_mat);
for e = 1:event_num
for f = 1:chem_r
         distance =sqrt((currentdata_mat(e,1)-chem_spots(f,1))^2+(currentdata_mat(e,2)-chem_spots(f,2))^2);
distance = distance/(currentdata_mat(e,5));
if distance <=4.5
currentdata_mat(e,7)=distance;
spot_num = (chemspot_r)-(chem_r-f);
currentdata_mat(e,8)=spot_num;
end
end
end

   [currentdata_r,currentdata_c] = size(currentdata_mat);
   currentdata2_mat=[];
   %Remove the events that are unrelated to an chemical spot
for z = 1:currentdata_r
ifcurrentdata_mat(z,8)~=0
         currentdata2_mat = vertcat(currentdata2_mat,currentdata_mat(z,:));
end
end

   %Take each "current" data set and concatenate into the full data
   %matrices and arrays
fulldata_mat = vertcat(fulldata_mat,currentdata2_mat);

end



toc
end
```

Training_Processing code:

```
function [chemspots_mat,chem_class,fulldata_mat] = Training_Processing(linkedTracks_mat,mov)
%Processing intakes cell arrays containing the names of any number of
```

```
%linkedTracks files or behavior movie files. It then outputs two matrices.
%chemspots_matcotains the locations of all the chemical spots, the. fulldata_mat is matrix containing
%location, state, closest chemical spot number, distance to said chemical
%spot, pixel conversion value and a behavioral class number (1-4).

chemspots_mat = [];
chem_class = {};
fulldata_mat = [];

tic
num_tracks = length(linkedTracks_mat);

%This loop will go through each track and movie in the file then start building the
%final matrices and arrays.
for a = 1:num_tracks
   %Pull out the movie and the linkedTracks file of interest
track = char(linkedTracks_mat(a));
movie = char(mov(a));
load (track)

   %This component utilizes the movie that was inputted to create an
   %interacTable image for inputting the location of each of the spots.
   %The image is displayed, the chemical spots can be clicked in the
   %image, hit ENTER and then the code will be able to read in the spot
   %locations and then it will label the spots.
MovieObj = VideoReader(movie);
Mov = read(MovieObj, 1);
Mov = Mov(:,:,3); %Pulls the Blue from the Image to make image B&W
imshow(imadjust(Mov,[],[],2));
txt = 'Click on stimulus spots. When done, hit ENTER.'; title(txt);
   [X,Y] = ginput;
chem_spots = [X Y];
   [cs_row,cs_col] = size(chem_spots);
for b = 1:cs_row
      %Add label to image
text(chem_spots(b,1),chem_spots(b,2),['Spot ' num2str(b)], 'Color', 'r', 'FontWeight', 'bold');
end
chemspots_mat=vertcat(chemspots_mat,chem_spots);

   %This allows the user to create a cell array that contains the chemical
   %spot type at each spot in the image
   [chem_r,chem_c] = size(chem_spots);
currentchem_class = [];
for c = 1:chem_r
num = num2str(c);
fprintf(['This is spot ' num ' in ' track])
class = cellstr(input('\nWhat is the class of this spot? ','s'));
currentchem_class = vertcat(currentchem_class,class);
end

   %Load in the different data columns of interest.
   x = struct2mat(2,linkedTracks,[],{'SmoothX'});
   y = struct2mat(2,linkedTracks,[],{'SmoothY'});
   s = struct2mat(2,linkedTracks,[],{'State'});
   f = struct2mat(2,linkedTracks,[],{'Frames'});
```

```
    x=double(x);
    y=double(y);
    s=double(s);
    f=double(f);
    %Convert each column out of single precision values into doubles, convert
    %from rows into columns and convert the state values to make them more
    %manageable.
    x=x';
    y=y';
    s=s';
    f=f';
    s = s*10;
    s = round(s);
    %All of the separate matrices are combined into the full_data matrix.
    %X AND Y ARE IN PIXELS
    %Insert the ratio of pixel to mm into the fifth column
    mm2p_ratio = linkedTracks.PixelSize;
pix = 1/ mm2p_ratio;
event_num=length(x);
    p = zeros(event_num,1);
p(:,:)=pix;
currentdata_mat = [x y s f p];

    %This are the state codes that are associated with different types of
    %behavior
all_beh = [11,10,47,57,71,51,31,41,20,30,40,50,70,80,74,75,43,34,53,35];
for d = 1:event_num
        % Add in Behavior class
beh_type = currentdata_mat(d,3);
location=find(all_beh==beh_type);
if location == 1
currentdata_mat(d,6)=1;
elseif location == 2
currentdata_mat(d,6)= 2;
elseif location >= 3 && location <=8
currentdata_mat(d,6)= 3;
else
currentdata_mat(d,6)= 4;
end
end

    %Find and associate with nearest chemical spot
    [chem_r,chem_c]=size(chem_spots);
    [chemspot_r,chemspot_c]=size(chemspots_mat);
for e = 1:event_num
for f = 1:chem_r
          distance =sqrt((currentdata_mat(e,1)-chem_spots(f,1))^2+(currentdata_mat(e,2)-chem_spots(f,2))^2);
distance = distance/(currentdata_mat(e,5));
if distance <=4.5
currentdata_mat(e,7)=distance;
spot_num = (chemspot_r)-(chem_r-f);
currentdata_mat(e,8)=spot_num;
end
end
end
```

```
   [currentdata_r,currentdata_c] = size(currentdata_mat);
   currentdata2_mat=[];
   %Remove the events that are unrelated to an chemical spot
for z = 1:currentdata_r
ifcurrentdata_mat(z,8)~=0
        currentdata2_mat = vertcat(currentdata2_mat,currentdata_mat(z,:));
end
end

   %Take each "current" data set and concatenate into the full data
   %matrices and arrays
chem_class = vertcat(chem_class,currentchem_class);
fulldata_mat = vertcat(fulldata_mat,currentdata2_mat);

end
toc
end
```

Data_Manipulation reads in the outputs from the Processing code and then intakes user inputs to select parameters for organizing the data. The user has control over the frame numbers of interest and the distances considered. Because we were interested in behavior relative to the center of each chemical spot we decided to further refine our analysis by also examining the events within different radii from the center spot. For example, the total distance we were interested in was 4.5mm from the center of the chemical spot. If we set 5 distance bins, each bin will be 0.9 mm. In this situation, all the events in distance bin 1 will be within 0-0.9mm from the center of the chemical spot. All events in distance bin 2 will be within 0.9-1.8mm from the center of the chemical spot. The output of Data Manipulation are arrays and matrices of the count of events for each distance and behavior or interest. Additionally, there are matrices and arrays containing the relative percentages of behavior occurring within each distance. The commented code is below:

```
function [count_mat,percentage_mat,count_array, percentage_array,chemspot_array,spotevents_array,variables] =
Data_Manipulation(fulldata_mat,chemspots_mat)
%Program Writer: Rebecca Lee
%Data_Manipulation intakes the data set from Processing and allows the user
%to input details on the type of final data set they are interested in
%creating. Will output matrices and arrays that can then be used for data analysis and
%training data - all behavior types and chemical spot types

%INPUTS AND VARIABLES (Allows the user to control variables of the data
%set.

tic
%Setting distance bins
max_bin = 4.5;
dis_bin = str2double(input('How many distance bins do you want? ','s'));
disbin_r = max_bin/dis_bin;
```

```
%Setting frames of interest
max_frame = max(fulldata_mat(:,4));
fprintf(['\nThe max frame number is ' num2str(max_frame)])
minframe = str2double(input('\nWhat is the minimum frame number of interest? ','s'));
maxframe = str2double(input('What is the maximum frame number of interest? ','s'));

%Setting the distance bins of interest
fprintf(['\nThere are ' num2str(dis_bin) ' distance bins for this data set.'])
min_bin = str2double(input('\nWhat is the minimum distance bin of interest? ','s'));
max_bin = str2double(input('What is the maximum distance bin of interest? ','s'));

variables = [dis_bin;minframe;maxframe;min_bin;max_bin];

[fulldata_r,fulldata_c]=size(fulldata_mat);
[chemspot_r,chemspot_c]=size(chemspots_mat);

framedata_mat = [];
chemspot_array = zeros(5,dis_bin,chemspot_r);
spotevents_array = cell(1,1,chemspot_r);
percentage_mat = [];
count_mat = [];

%DATA CLEANUP
%Remove frames we aren't interested in
for a = 1:fulldata_r
iffulldata_mat(a,4)>=minframe&&fulldata_mat(a,4)<=maxframe
framedata_mat = vertcat(framedata_mat,fulldata_mat(a,:));
end
end
[framedata_r,framedata_c]=size(framedata_mat);

%Create count array and cell array containing the events associated with each event
for b = 1:chemspot_r
for c = 1:framedata_r
iframedata_mat(c,8) == b
beh_type = fulldata_mat(c,6);
event_dis = fulldata_mat(c,7);
disbin_type = ceil(event_dis/disbin_r);
ifdisbin_type == 0
disbin_type = 1;
end
        chemspot_array(beh_type,disbin_type,b)=chemspot_array(beh_type,disbin_type,b)+1;
        spotevents_array{1,1,b}=vertcat(spotevents_array{:,:,b},fulldata_mat(c,:));
end
end
end
chemspot_array(5,:,:)=sum(chemspot_array(1:4,:,:));

%Calculate percentages
percentage_array = zeros(5,dis_bin,chemspot_r);
for d = 1:chemspot_r
for e = 1:4
if e ==1
```

```
percentage_array (1,:,d) =(chemspot_array(1,:,d))./(chemspot_array(5,:,d));
elseif e == 2
percentage_array (2,:,d) =(chemspot_array(2,:,d))./(chemspot_array(5,:,d));
elseif e == 3
percentage_array (3,:,d) =(chemspot_array(3,:,d))./(chemspot_array(5,:,d));
elseif e == 4
percentage_array (4,:,d) =(chemspot_array(4,:,d))./(chemspot_array(5,:,d));
end
percentage_array (5,:,d) =(sum(percentage_array (1:4,:,d)));
end
end

%Reorganize percent data into a matrix
[percentage_r,percentage_c,percentage_d]=size(percentage_array);
for f = 1:percentage_d
holdthis = [];
for g=1:4
holdthis=horzcat(holdthis,percentage_array(g,min_bin:max_bin,f));
end
percentage_mat = vertcat(percentage_mat,holdthis);
end

%Remove the NaN values
[percmat_r,percmat_c]=size(percentage_mat);
for h = 1:percmat_r
fori = 1:percmat_c
ifisnan(percentage_mat(h,i))
percentage_mat(h,i)= 0;
end
end
end

%Reorganize counts into a matrix
[count_r,count_c,count_d]=size(chemspot_array);
for j = 1:count_d
holdthis = [];
for k=1:4
holdthis=horzcat(holdthis,chemspot_array(k,min_bin:max_bin,j));
end
count_mat = vertcat(count_mat,holdthis);
end
count_array = chemspot_array;

toc
end
```

Part 2. Machine Learning

For the machine learning components of the code we utilized the built-in MATLAB functions. Classification tree analysis was conducted because we were using behavioral data and we could not assume that the correlations and the averages were similar between tests. Classification Tree analysis is capable of handling any data type and has many options of later optimization after the model is generated.

There are two main components of the machine learning components of the code. There was the code that created the training data and the model for the tests,theModel_Gen script. Then there was the script that tested a specific experiment using a model, the Tree_Test_Script.

Model_Gen asks the user for the files for the training data and then runs Training_Processing and Data_Manipulation on the data sets given. These training data sets are then used to create a model using the fitctree function in MATLAB. Additionally, a 4 fold cross validation model is made to allow us to later test the performance of the models generated. At the end, the script save the model, the cross validation and all the training data. The commented code is below:

```
%Program Writer: Rebecca Lee
%Reads in the training data sets of interest, prepares and then makes
%models from the datasets.

tic
%Input the new data
track = input('What is the linkedTrack of interest? Add files in a cell array. ');
movie = input('What is the movie of interest? Add files in a cell array. ');

%Processing the new data
[trainingchemspots_mat,trainingchemclass,trainingfulldata_mat] = Training_Processing(track,movie);
[trainingcount_mat,
trainingperc_mat,trainingcount_array,trainingperc_array,trainingchemspot_array,trainingspotevents_array,variables]
= Data_Manipulation(trainingfulldata_mat,trainingchemspots_mat);

%Create a model
[tree_model]=fitctree(trainingperc_mat,char(trainingchemclass));
[cvtree_model]=fitctree(trainingperc_mat,char(trainingchemclass),'KFold',4);

%Save the files
uisave({'trainingcount_mat','trainingperc_mat','trainingcount_array','trainingperc_array','trainingchemspot_array','tra
iningspotevents_array'})
uisave({'tree_model'})
uisave({'cvtree_model'})
toc
```

Tree_Test_Script asks the user for the filename of the model and the file for the experiment of interest. It can analyze the same experiment multiple times. For each test of the experiment, the script will save the data from the experiment, the final chemical decision of the program and the confidence scores of the program. The commented code is below:

```
%Program Writer: Rebecca Lee
%Classification Tree This script will utilize previously made
%models to determine the classification of chemicals in a new data set.

tic
%Load the models
model = input('What is the filename of the model? ','s');
```

```
load(model)

%Input the new data (will only take one experiment at a time)
track = {input('What is the linkedTrack of interest? ','s')};
movie = {input('What is the movie of interest? ','s')};

%Repeat the test
repeat = input('How many times do you want to repeat the test? ');

%Information for plotting
plate_row = input('How many rows are there in the plate? ');
plate_col = input('How many columns are there in the plate? ');
total = plate_row*plate_col;
breaks = total/plate_col;

%Processing the new data
for a = 1:repeat
   [testchemspots_mat,testfulldata_mat] = Processing(track,movie);
   [testcount_mat, testperc_mat,testcount_array,testperc_array,testchemspot_array,testspotevents_array] =
Data_Manipulation(testfulldata_mat,testchemspots_mat);

   %Test the processed data
fprintf(['This is test ' num2str(a) '\n'])
fprintf ('These are the results for the classification tree test:\n')
   [tree_results,scores] = predict(tree_model,testperc_mat)

   %Make the images of scores
   %I am not sure how fitctree decides which column is aversive and which
   %one is neutral. I suspect it has to do with which was the first state
   %defined in the original training set. BE WARNED THAT THIS MAY NOT HOLD
   %TRUE FOR ALL DATA SETS.
neu_scores = scores(:,2)';
aver_scores = scores(:,1)';
neu_matrix = neu_scores(1:plate_col);
aver_matrix = aver_scores(1:plate_col);

for b = 1:(breaks-1)
min = (b*plate_col)+1;
max = (min)+(plate_col-1);
neu_matrix = vertcat(neu_matrix,neu_scores(min:max));
aver_matrix = vertcat(aver_matrix,aver_scores(min:max));
end

imagesc(neu_matrix)
caxis([0 1])
colormap('hot')
colorbar
title ('Neutral Scores')
print ('-dpsc','-append',['Scores for ' track{1} ' with ' model '.ps'])

imagesc(aver_matrix)
caxis([0 1])
colormap('hot')
colorbar
```

```
title ('Aversive Scores')
print ('-dpsc','-append',['Scores for ' track{1} ' with ' model '.ps'])


uisave({'tree_results','scores','testfulldata_mat','testcount_mat','testchemspots_mat','testperc_mat','testcount_array','te
stperc_array','testchemspot_array','testspotevents_array'})
end

toc
```

Part 3. Plots

We wanted to be able to graphically see the locations of the event occurring in the experiments. This was done by a script called All_Plots. This script utilized the hist2 function provided by Professor Albrecht. This script runs Processing to obtain all the events in the plate and then runs Data Processing to get the events sorted by spot. The script plots a heat map of all events, heat maps for each of the behavior classes. Additionally, the program can create a plot of all events, forward events, pause events, and other events for any specific spots of interest in the experiment. All the plots are appended to a pdf that is outputted into the path folder. The comment code is below:

```
%Program Writer: Rebecca Lee
%All Plots is a script that intakes the data of a plate then makes plots of
%the data. The user has the option to control the frames, bin size and the
%behavior classes examined in the plots. This code assume that the exact
%classes of the spots are unknown so the chemcial classes of the spots
%cannot be controlled.

%Read in the file of interest
track = {input('What is the linkedTrack of interest? ','s')};
movie = {input('What is the movie of interest? ','s')};
interest_spots = input('Are there any particular spots of interest? \nInput a matrix of the spots of interest. If there is
not specific spot of interest input a matrix containing 0.\n');

%Process the data
tic
[plotchemspots_mat,plotfulldata_mat] = Processing(track,movie);
[~,~,~,~,~,plotspotevents_array] = Data_Manipulation(plotfulldata_mat,plotchemspots_mat);
toc

%Plot the full data
tic
x1 = plotfulldata_mat(:,1);
y1 = plotfulldata_mat(:,2);
pix = plotfulldata_mat(1,5);
Figure();
hist2(x1,y1,pix,pix)
axisij
colorbar
title('Location of all events')
```

```
print('-dpsc','-append',['All Plots_' track{1} '.ps'])

%Plot each of the behavior classes of the data
num_events = length(x1);
pause_events = [];
forward_events = [];
aver_events = [];
other_events = [];
for a = 1:num_events
ifplotfulldata_mat(a,6)==1
pause_events = vertcat(pause_events,plotfulldata_mat(a,1:2));
elseifplotfulldata_mat(a,6)==2
forward_events = vertcat(forward_events,plotfulldata_mat(a,1:2));
elseifplotfulldata_mat(a,6)==3
aver_events = vertcat(aver_events,plotfulldata_mat(a,1:2));
else
other_events = vertcat(other_events,plotfulldata_mat(a,1:2));
end
end

holdingpattern = {pause_eventsforward_eventsaver_eventsother_events};

for b = 1:4
data = holdingpattern{b};
Figure();
   hist2(data(:,1),data(:,2),pix,pix)
axisij
colorbar
if b == 1
title('All pause events')
elseif b == 2
title('All foward events')
elseif b == 3
title('All aversive events')
else
title('All other events')
end
print('-dpsc','-append',['All Plots_' track{1} '.ps'])
end

%Examine specific spots in the plate
num_interest = length(interest_spots);
for c = 1:num_interest
ifinterest_spots(c)==0
break
end
spot = plotspotevents_array{:,:,c};
spotevents = length(spot);
Figure();
   hist2(spot(:,1),spot(:,2),pix,pix)
axisij
colorbar
title(['Location of all events in Spot ' num2str(c)])
print('-dpsc','-append',['All Plots_' track{1} '.ps'])
```

```
   %Plot each of the behavior classes of the data for the particular spot
spotpause_events = [];
spotforward_events = [];
spotaver_events = [];
spotother_events = [];
for d = 1:spotevents
if spot(d,6)==1
spotpause_events = vertcat(spotpause_events,spot(d,1:2));
elseif spot(d,6)==2
spotforward_events = vertcat(spotforward_events,spot(d,1:2));
elseif spot(d,6)==3
spotaver_events = vertcat(spotaver_events,spot(d,1:2));
else
spotother_events = vertcat(spotother_events,spot(d,1:2));
end
end

spotholdingpattern = {spotpause_eventsspotforward_eventsspotaver_eventsspotother_events};

for e = 1:4
spotdata = spotholdingpattern{e};
Figure();
      hist2(spotdata(:,1),spotdata(:,2),pix,pix)
axisij
colorbar
if e == 1
title(['All pause events at Spot ' num2str(c)])
elseif e == 2
title(['All foward events at Spot ' num2str(c)])
elseif e == 3
title(['All aversive events at Spot ' num2str(c)])
else
title(['All other events at Spot ' num2str(c)])
end
print('-dpsc','-append',['All Plots_' track{1} '.ps'])
end
end

clearvars

toc
```

Appendix VII. Code Procedure

All of the components were scripts. There are different scripts to complete different actions. If you want to generate a model run the Model_Gen script. If you want to use a previously generated model to predict the chemical spot type of a unknown array, utilize the Tree_Test Script.

The model generation scripts and the model testing scripts can only be run in MATLAB 2014 and newer. This is because the functions for Classification tree model generation and model testing changed name between MATLAB 2012/ MATLAB 2013 and MATLAB 2014. There are older functions that can complete these actions.

To run any of the scripts:

1. Place all of the MATLAB scripts, and the files and videos to be analyzed into the same folder.
2. Open MATLAB
3. Open the Path in MATLAB to the folder containing the scripts, files and videos.
4. Run the script of interest by typing the name in the command line of MATLAB
5. The automated program will prompt the user for any inputs it needs.

3. An ability to design a system, component, or process to meet desired needs *within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability* (ABET 3c) *while incorporating appropriate engineering standards* (ABET Criterion 5) (*need to assess each of these separately, but since 'or' and "such as"*not all need to be met separately).

i) multiple realistic constraints (*economic, environmental, social, political, ethical, health and safety, manufacturability*) – **pages 21-26**

ii) appropriate engineering standards - **pages 27-48**

4. An ability to function on multidisciplinary teams (3d). **pages _____**

6. An understanding of professional and ethical responsibilities (3f)

i) Professional – **page 67**

ii) Ethical – **page 68**

7. An ability to communicate effectively (3g). **pages _____**

8. The broad education necessary to understand the impact of engineering solutions in a global, *economic*, *environmental*, and societal context (3h). (both economic AND environmental need to be addressed)

i) Economic – **page64-65**

ii) Environmental – **page 66**

10. A knowledge of contemporary issues (3j). **page 69**

Appendix VIII. Worm NGM Media

Notes: Types of experimental plan represents the recipe of the media and the thickness of the media on the plate

	Standard Recipe	Experimental Plan					
		regular; thin	1.5x agar; regular+thin	2x agar; regular+thin	2.5x agar; regular+thin	3x agar; regular+thin	3.5x agar; regular+thin
	1 L	10 mL	25 mL	25 mL	25 mL	25 mL	25 mL
NaCl	3 g	0.03 g	0.075 g	0.075 g	0.075 g	0.075 g	0.075 g
Peptone	2.5 g	0.025 g	0.125 g	0.125 g	0.125 g	0.125 g	0.125 g
Agar	17 g	0.17 g	0.6375 g	0.85 g	1.0625 g	1.275 g	1.4875 g
dH2O	975 mL	9.75 mL	24.375 mL	24.375 mL	24.375 mL	24.375 mL	24.375 mL
Heat up to melt, cool down till 50˚C then add							
	Standard Recipe	Experimental Plan					
		regular; thin	1.5x agar; regular+thin	2x agar; regular+thin	2.5x agar; regular+thin	3x agar; regular+thin	3.5x agar; regular+thin
	1 L	10 mL	25 mL	25 mL	25 mL	25 mL	25 mL
5mg/mL cholesterol	1 mL	10 µL	25 µL	25 µL	25 µL	25 µL	25 µL
1M $CaCl_2$	1 mL	10 µL	25 µL	25 µL	25 µL	25 µL	25 µL
1M $MgSO_4$	1 mL	10 µL	25 µL	25 µL	25 µL	25 µL	25 µL
1M KPO_4	25 mL	250 µL	625 µL	625 µL	625 µL	625 µL	625 µL

For example:
“regular; thin” = regular media recipe pours thin layer of plate
“1.5x agar; regular+thin” = 1.5x agar of media pours regular thickness (10 mL) and thin layer of plates

1. Prepare NGM plates by following the recipe above: mix the corresponding chemicals, autoclave the media, cool down the media to around 50˚C, and pour plate.
2. Add a drop of tested chemical into the plate as well as some adult N2 worms.
3. Observe the chemical diffusion and the interaction with the worms.