Mapping Worm-Bioavailable Chemical Space for Improved Probe Discovery and Characterization with Caenorhabditis elegans

by

Andrew Robert Burns

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Molecular Genetics
University of Toronto

© Copyright by Andrew Robert Burns 2013
Mapping Worm-Bioavailable Chemical Space for Improved Probe Discovery and Characterization with *Caenorhabditis elegans*

Andrew Robert Burns
Doctor of Philosophy
Graduate Department of Molecular Genetics
University of Toronto
2013

Abstract

Small molecule modulators of protein activity are powerful reagents to probe the biology of living systems. Small molecules have been used repeatedly to uncover protein functions that were otherwise hidden from more conventional approaches. One way to find new small molecule probes is to screen libraries of chemicals for those that induce phenotypes in model organisms. Chemical screens in yeast, worms, fish, and plants have all resulted in the discovery of novel probes of protein function.

The nematode worm *Caenorhabditis elegans* has great potential for the discovery and characterization of small molecule probes. Its small size, ease of culture, and genetic tractability make this animal amenable to high-throughput chemical screens, as well as mode-of-action studies of newly identified bioactives. One drawback to using *C. elegans* as a screening tool for novel small molecule probes is its resistance to pharmacologically active drug-like compounds. Given its extensive xenobiotic defenses, poor bioavailability may explain this resistance. Indeed, there are a number of bioactive compounds that are ineffective when applied to whole worms that perturb their target proteins in dissected animals. Thus, target
access is likely a major limiting factor for pharmacological efficacy in *C. elegans*, although this idea has never been tested.

In this thesis, I describe a novel HPLC-based method for measuring the accumulation of exogenous compounds in worms, and I use this method to test the accumulation of hundreds of drug-like molecules in *C. elegans*. I show for the first time that worms resist drug-like compound accumulation, and that accumulation is generally required for bioactivity. I describe the construction of a computer-based tool that can identify worm-bioavailable molecules, and I highlight structural features that influence accumulation. I show that screening worm-bioavailable molecules increases the bioactive discovery rate of *C. elegans* chemical screens.

Next, I describe novel bioactive heterobiaryl amide compounds discovered through a *C. elegans* chemical screen. I apply insights gleaned from the mapping of worm-bioavailable chemical space to facilitate their characterization, and preliminary data suggest that these compounds are agonists of the worm metabotropic glutamate receptor MGL-1. Thus, the worm-bioavailable chemical space I describe facilitates small-molecule probe discovery and characterization.
Acknowledgements

First, I would like to thank my supervisor Peter Roy. Roughly seven years ago Peter’s excitement and commitment to science overshadowed his unfortunate affinity for the rock bands Rush and Nickelback, and I joined his lab. Peter’s meticulous nature and skill as a scientist are an example for any young scientist, and I am grateful to have had him as a mentor. I am a more confident scientist today than I was seven years ago, and I owe this in large part to Peter. I would also like to thank my extended family of mentors Sean Cutler, Henry Krause, and Andrew Spence. I have learned a great deal from them as well.

Second, I thank the members of the Roy Lab, both past and present. In alphabetical order by last name they are: Mariam Alexander, Simon Alfred, Rachel Bagg, Alex Byrne, Kevin “Boo” Chan, Serena D’Souza, Scott Dixon, Trevor “T-Dawg” Kwok, Genna “G’Luz” Luciani, Nicole Ricker, Ashwin “Shwin” Seetharaman, Guillermo “G’Mo” Selman, and Victoria “V Dubs” Wong. They have all contributed to my personal growth, both in and out of science. In my mind, they are not only esteemed colleagues, but good friends as well. Thanks for the good times.

Finally, I would like to thank my family and friends outside of science. In particular I would like to call attention to my existence, and the efforts my parents made towards achieving it. My folks, Brian and Joan Burns, are undoubtedly two of the finest people living. From day one they have provided me with limitless love and support, and they have been a constant source of stability and strength for me throughout my graduate career. I bow humbly.

Given the outstanding contributions my parents have made to my existence and success, I dedicate this thesis to them. Moms and Pops, thanks again.
Note to Reader

This thesis is an account of two projects that I carried out over the course of my doctoral studies. The first project involves the mapping of a worm-bioavailable chemical space to facilitate the discovery and characterization of small-molecule probes using Caenorhabditis elegans. The second project involves the discovery of a novel family of bioactive heterobiaryl amide compounds from a C. elegans phenotype-based chemical screen, and their subsequent chemical-genetic characterization. I have presented the two projects separately because they are not directly related, although the characterization of the heterobiaryl amide compounds was facilitated by the insights gleaned from the worm-bioavailable chemical space I define in project 1. The first project is presented in Chapters 1 to 3 of this thesis, and the second project is presented as a self-contained unit in Chapter 4.
# Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgements ......................................................................................................................... iv

Note to Reader .................................................................................................................................. v

Table of Contents ............................................................................................................................ vi

List of Tables ....................................................................................................................................... ix

List of Figures ...................................................................................................................................... x

List of Appendices ........................................................................................................................... xi

Chapter 1. General Introduction ...................................................................................................... 1
   1.1 Chemical Genetics: Probing Biology with Small Molecules ...................................................... 2
      1.1.1 Forward Chemical Genetics and Phenotype-Based Screens .............................................. 4
      1.1.2 Reverse Chemical Genetics and Target-Based Screens .................................................... 7
      1.1.3 Assessing Essential Gene Function with Small Molecules ............................................... 8
      1.1.4 Circumventing Functional Redundancy with Small Molecules ......................................... 9
      1.1.5 Probing Temporal Requirements of Protein Function with Small Molecules ................. 11
   1.2 Probe Discovery in an Optimized Chemical Space .................................................................. 13
      1.2.1 What is Chemical Space? .................................................................................................. 13
      1.2.2 Biogenic Bias of Screening Libraries ............................................................................... 13
      1.2.3 Structural Diversity of Screening Libraries ...................................................................... 15
      1.2.4 Quality Control for Screening Libraries ........................................................................... 17
      1.2.5 Bioavailability and the Lipinski Rule-of-Five ................................................................... 19
   1.3 Probe Discovery and Characterization using *C. elegans* ....................................................... 21
      1.3.1 *Caenorhabditis elegans*: A Premier Animal Model ......................................................... 21
      1.3.2 Probe Discovery with *C. elegans* ................................................................................... 22
      1.3.3 Small Molecule Mode-of-Action Studies using Worms .................................................... 25
   1.4 Research Problem: Overcoming the Resistance of *C. elegans* to Pharmacological Perturbation ................................................................................................................. 27

Chapter 2. Developing a Structure-Based Model to Predict Small Molecule Accumulation in *C. elegans* ............................................................................................................................. 31
   Abstract ........................................................................................................................................... 31
   2.1 Introduction .............................................................................................................................. 32
   2.2 Results ...................................................................................................................................... 35
      2.2.1 An HPLC-Based Small Molecule Accumulation Assay for *C. elegans* ......................... 35
      2.2.2 A Survey of Drug-Like Small Molecule Accumulation in Worms .................................. 39
      2.2.3 Exogenous Compounds Can Accumulate as Metabolites in *C. elegans* ....................... 43
4.2.1 A Phenotype-Based Chemical Screen in Worms Uncovers a Heterobiaryl Amide Compound (HBAC) that Induces Motility Defects .......................................................... 111
4.2.2 The HBACs are Structurally Similar to Known mGluR Modulators.......................... 113
4.2.3 HBAC-A Causes Premature Starvation-Induced Death (SID) in Worms, Consistent with it Modulating MGL-1 or MGL-2 Activity ........................................... 115
4.2.4 A SAM-Selected Structural Analog of HBAC-A, called HBAC-B, is a More Potent Inducer of Premature SID ................................................................. 117
4.2.5 HBAC Treatment Induces Premature SID by Engaging an Excessive Starvation Response, Consistent with MGL-1 or MGL-2 Agonism........................................... 118
4.2.6 The HBAC Sensitivity Profile of mgl Mutants Suggests MGL-1 Agonism as a Possible Mode-of-Action ................................................................................... 119
4.2.7 MGL-2 Activity is Not Modulated by the HBACs .................................................. 121
4.3 Discussion ................................................................................................................... 124
4.4 Materials and Methods ............................................................................................. 128
  4.4.1 Worm Culture and Strains .................................................................................. 128
  4.4.2 C. elegans Liquid-Based Chemical Screen .......................................................... 129
  4.4.3 Purchased Chemicals .......................................................................................... 130
  4.4.4 Swimming Videos ............................................................................................. 130
  4.4.5 Starvation Survival Assay ................................................................................ 130
  4.4.6 Molecular Cloning and Expression Construct Preparation .............................. 132
  4.4.7 mGluR Cell-Based Functional Assay ............................................................... 133

Appendix. Characterization of C. elegans Xenobiotic Metabolites by Mass Spectrometry ...135

Additional Files ............................................................................................................... 141

References ....................................................................................................................... 142
List of Tables

Table 2-1. Characterization of *C. elegans* xenobiotic metabolites ............................................. 44
Table 2-2. Accurate mass determinations for 4 representative xenobiotic metabolites ........... 45
Table 4-1. mGluRs with reported subtype-selective modulators ......................................................... 110
List of Figures

Figure 1-1. Examples of biologically active small molecules .................................................. 3
Figure 1-2. Elucidating the biological roles of proteins using forward and reverse chemical genetics .......................................................................................................................... 6
Figure 1-3. Finding an optical chemical space for probe discovery ....................................... 14
Figure 1-4. High-throughput screening of chemicals for bioactivity and target identification in C. elegans .......................................................................................................................... 23
Figure 1-5. C. elegans is resistant to pharmacological perturbation ........................................ 29

Figure 2-1. A survey of exogenous drug-like molecule accumulation in C. elegans .............. 36
Figure 2-2. The 1,4-dihydropyridine structures analyzed for accumulation in worms ............ 38
Figure 2-3. The complete small molecule accumulation dataset (1,132 compounds) ........... 41
Figure 2-4. The distribution of small molecule concentrations in worms for the accumulating compounds in the complete Spectrum dataset ................................................................. 42
Figure 2-5. β-glucosidase digestions of select xenobiotic metabolites ................................. 47
Figure 2-6. A machine-learned model predicts the accumulation of exogenous small molecules in C. elegans ............................................................................................................... 48
Figure 2-7. Comparison of Naïve Bayesian classifier machine-learned models trained using distinct molecular descriptors .................................................................................................. 50
Figure 2-8. Prominent substructures that influence small-molecule accumulation in C. elegans ................................................................................................................................. 52
Figure 2-9. Lipinski property distributions for the 483-compound training set for machine-learning .............................................................................................................................. 54
Figure 2-10. Phenotypic and accumulation analysis of Fenbufen and three analogs ............... 56
Figure 2-11. The C. elegans structure-based accumulation model (SAM) enriches for structurally diverse compounds with distinct bioactivities in the worm ........................................ 58

Figure 3-1. Theoretical Venn diagrams of human- and worm-bioavailable chemical spaces ...... 84
Figure 3-2. Methods to maximize bioavailability in the context of C. elegans high-throughput chemical screening ........................................................................................................... 93
Figure 3-3. Reaction scheme for the generation of SMoC-tagged chemical libraries ............. 101

Figure 4-1. Schematic diagram of a metabotropic glutamate receptor dimer .......................... 108
Figure 4-2. A newly identified heterobiaryl amide compound induces motility defects in C. elegans ............................................................................................................................ 112
Figure 4-3. The heterobiaryl amide compounds (HBACs) are structurally similar to established mGluR modulators ........................................................................................................... 114
Figure 4-4. The HBACs cause premature starvation-induced death by engaging an excessive starvation response ................................................................................................................. 116
Figure 4-5. The mgl mutant HBAC sensitivtiy profile ............................................................ 120
Figure 4-6. MGL-2 and mGluR5 are not HBAC target proteins ............................................. 122
Figure 4-7. Sequence alignment of the amino-terminal domains of rat mGluRs 1-8 and worm MGL-1 ........................................................................................................................... 123

Figure A-1. Mass spectrometry analysis of xenobiotic metabolites ........................................ 136
List of Appendices

**Appendix.** Characterization of *C. elegans* Xenobiotic Metabolites by Mass Spectrometry ......135
Chapter 1

General Introduction

Chapter 1. General Introduction

1.1 Chemical Genetics: Probing Biology with Small Molecules

Biological organisms, from the simplest prokaryotes to the most complex vertebrates, rely on a constellation of endogenous and exogenous small molecules to regulate their development, physiology, and behaviour (Alberts et al., 2007). Many of these small molecules bind and modulate the activity of specific cellular protein targets to achieve their respective biological activities. Small molecule ligands of proteins can be used to impart information within a cell (Alberts et al., 2007), between different cells in the same organism (Cutler et al., 2010; Motola et al., 2006), between individuals of the same species (Diggle et al., 2006; Kim et al., 2009; Srinivasan et al., 2008), and they can also serve as a means of communication or self-defense between distinct species (Miller, 2002) (Fig. 1-1). Small-molecule ligands have a varied range of structural design and complexity, underpinning the diverse biological processes that they can affect (Fig. 1-1). Given that small molecule signaling is an integral part of organic life, small molecule modulators of protein activity have great potential as tools to probe the biology of living systems and to modify disease states. Indeed, the majority of FDA-approved drugs are small molecule modulators of protein function (Paolini et al., 2006), and small molecules have been used countless times as reagents to investigate diverse biological processes (O'Connor et al., 2011; Spring, 2005; Stockwell, 2000; Yeh and Crews, 2003).

The approach of using small molecules to probe gene function is referred to as “chemical genetics” (Stockwell, 2000). In contrast with classical genetics, which relies primarily on mutant
Figure 1-1. Examples of biologically active small molecules. (A) Adenosine triphosphate (ATP) – a small molecule present in all cells that carries the energy currency of life (Alberts et al. (2007)). It is a ligand for many proteins, including kinases. The hydrolysis of ATP to form ADP releases ~7.3 kcal/mol of energy. (B) Δ4-dafachronic acid – An endocrine hormone produced by the nematode *Caenorhabditis elegans* that regulates reproductive development and metabolism through its target protein DAF-12, a nuclear hormone receptor (Motola et al. (2006)). (C) S-(+)-abscisic acid – A plant hormone that regulates development and adaptive stress responses by agonizing PYR/PYL receptors, a family of START proteins (Cutler et al. (2010)). (D) Daumone – A pheromone produced by *C. elegans* hermaphrodites that causes male attraction and arrests development in conditions of high population density. The effects of daumone are mediated, in part, by its interaction with its target proteins SRBC-64 and SRBC-66, which are G-protein coupled receptors (Kim et al. (2009), Srinivasan et al. (2008)). (E) Pseudomonas quinolone signal (PQS) – An interbacterial signaling molecule used by *Pseudomonas aeruginosa* for quorum sensing. PQS modulates the activity of a number of cellular protein targets (Diggle et al. (2006)). (F) Penicillin G – An antibiotic produced by the *Penicillium* mould, whose mode-of-action involves inhibiting bacterial cell wall synthesis (Miller (2002)).
alleles to alter the activity of specific genes, chemical genetics employs small molecule ligands to modulate the activity of specific gene products, such as proteins. The small molecules used in chemical genetic experiments can be natural products or synthetic chemicals. The chemical genetic method affords four main advantages over the classical genetic approach. First, target activity can be titrated from no activity to full activity in a continuous fashion – a feat that cannot be achieved by even the most comprehensive allelic series. Second, the kinetics of target activity can be assayed, which is difficult to accomplish with a genetic modification present throughout an organism’s life. Third, functional redundancy of genes, which is not revealed by single gene mutations, can be circumvented with selective small molecule agonists or promiscuous antagonists of functionally redundant genes. Fourth, small molecules are easily transferred between organisms, whereas genetic alleles are not. Although conditional alleles of genes, as well as hypomorphic and hypermorphic alleles, can be used to control the timing and magnitude of target activity, it is not always possible to generate such alleles for every gene of interest. Thus, the power of chemical genetics lies in its ability to interrogate systems that are otherwise out-of-reach for classical genetic approaches.

In this section, I will briefly describe basic chemical-genetic strategies, and I will provide examples of how small molecule probes have been used to reveal protein functions inaccessible to classical genetic perturbation.

1.1.1 Forward Chemical Genetics and Phenotype-Based Screens

Classical forward genetic screens in model organisms have been used to identify the roles of many genes involved in a variety of biological processes (Jorgensen and Mango, 2002; Page and Grossniklaus, 2002; Patton and Zon, 2001; St Johnston, 2002). A forward genetic screen
typically involves the random mutagenesis of a large number of organisms, screening the mutants for a phenotype of interest, and subsequently identifying the mutated genes via genetic mapping. Phenotype-based genetic screens are powerful because they afford an unbiased exploration of the genome for genes involved in a biological process of interest.

The forward chemical genetic approach also begins with an unbiased phenotypic screen; however, mutagenized genomes are replaced with libraries of hundreds or thousands of small molecules (Fig. 1-2A). Purchasable chemical libraries are available from numerous vendors (Irwin and Shoichet, 2005). The library chemicals are arrayed in multi-well plates, typically with one distinct compound per well, and are screened for small molecules that induce a desired phenotype in the organism or cell type of interest. The number of bioactive compounds (or “hits”) obtained from a forward-genetic screen will depend on the assay and the library being screened. Identifying the target protein of a bioactive compound is often difficult and is typically the slow step for any chemistry-to-gene screen. However, there are a number of molecular, genetic, biochemical, and cell-biological approaches for target identification (see (Andrusiak et al., 2012; Rix and Superti-Furga, 2009; Sato et al., 2010; Tashiro and Imoto, 2012; Titov and Liu, 2012) for a thorough discussion of small-molecule target identification methods). In addition, the chemical structure of a novel bioactive can sometimes suggest potential target proteins, based on its structural similarity to known protein ligands (Keiser et al., 2007). A variety of searchable databases and computational tools are available for such in silico structure-guided approaches to target identification (www.Click2Drug.org). Forward chemical genetic screens have identified novel small molecule modulators of many proteins, and have implicated these proteins in the regulation of a broad range of biological processes (O’Connor et al., 2011; Spring, 2005; Stockwell, 2000; Yeh and Crews, 2003).
A. Forward Chemical Genetics

(1) Phenotype-Based Screen

Apply small-molecules to the organism or cell-type of choice

Identify compounds that induce an interesting phenotype (e.g. paralysis)

(2) Target Identification

Identify protein target using genetic and/or biochemical methods

B. Reverse Chemical Genetics

(1) Target-Based Screen

Apply small-molecules to purified target protein

Identify compounds that modulate protein activity (e.g. enzyme inhibitor)

(2) Phenotype Assessment

Control (no small molecule)  Treatment (plus small molecule)

Figure 1-2. Elucidating the biological roles of proteins using forward and reverse chemical genetics. (A) The forward chemical genetic method begins with an organism- or cell-based screen for chemicals that induce an interesting phenotype. The protein target of a small molecule recovered from such a phenotype-based screen can then be identified using genetic and/or biochemical approaches. (B) The reverse chemical genetic method begins with a target-based screen to identify compounds that modulate the activity of a specific target protein. Small molecules recovered from a target-based screen can then be applied to an organism or cell-type of interest to determine the phenotypic outcome of protein modulation. Both forward and reverse chemical genetics can identify small molecule probes of protein function.
1.1.2 Reverse Chemical Genetics and Target-Based Screens

In contrast to forward chemical genetics, reverse chemical genetics endeavours to explore the biological function of a pre-determined protein target (Fig. 1-2B). Instead of probing the system with thousands of chemicals to find the few that induce a given phenotype, one small molecule with established activity against a specific protein is used to elucidate that protein’s biological role. In this regard, reverse chemical genetics is a relatively biased method of interrogating gene function, akin to the use of genetic knockouts or RNAi knockdown to selectively perturb gene activity in classical reverse genetic experiments. Reverse chemical genetics has been used numerous times to elucidate gene function in a number of systems (O’Connor et al., 2011; Spring, 2005; Stockwell, 2000; Yeh and Crews, 2003).

If a small molecule modulator of a given protein is not already available, it can be found using target-based chemical screening. Target-based screens are done by purifying the target protein of interest, developing *in vitro* protein binding or enzymatic assays that are amenable to miniaturization in multi-well plates, and screening hundreds of thousands of chemicals (or more) for compounds that bind and modulate the activity of the purified protein (Fig. 1-2B). Target-based screens are typically much faster than phenotypic chemical screens, and they permit the screening of many more chemicals in a relatively short amount of time. Also, in contrast to phenotypic screens, the target is known *a priori*. However, unlike bioactive compounds identified from phenotype-based screens, small molecules discovered through target-based screening may not be active *in vivo*, and they may not be specific for the target against which they were screened. For these reasons, additional control experiments must be performed to ensure that the modulation of target activity is not an artefact of the screening
assay (Jadhav et al., 2010; Thorne et al., 2010), and structural modifications must often be made to the initial “hit” compounds to accommodate improved bioavailability and specificity of the probe molecules. Target-based screens have uncovered small molecule modulators of many proteins, and these compounds have been used as tools in a variety of reverse chemical genetic experiments ((Stockwell, 2000) and references therein).

1.1.3 Assessing Essential Gene Function with Small Molecules

Aspects of biology that are inaccessible to classical genetic approaches can be interrogated using chemical genetics. For instance, genes that are essential to an organism’s survival can be difficult to investigate by classical genetic methods, particularly if conditional or weak loss-of-function alleles are not available, since mutations in these genes are lethal. However, a small molecule modulator of an essential protein can be applied at a dose that dampens but does not eliminate activity, or at a time point in development when the protein is no longer required for viability. For example, Sebolt-Leopold et al. found that treating 4 to 6-week old colon tumour-bearing mice with the MEK1 kinase inhibitor PD 184352 decreased tumour size, consistent with in vitro data showing that PD 184352 affects the cell cycle and growth of cultured colon tumour cells (Sebolt-Leopold et al., 1999). Performing the analogous reverse genetic experiment with tumour processes from adult Mek1 mutant mice was impossible since Mek1-deficient mice die early in embryogenesis (Giroux et al., 1999). Thus, a small molecule probe elucidated a role for MEK-1 in colon tumour processes in vivo, that could not be revealed by classical genetic methods.
1.1.4 Circumventing Functional Redundancy with Small Molecules

Classical genetics aims to assign function to genes. This is typically accomplished by observing the phenotypic consequence of mutating or deleting a gene of interest, and relating phenotype to genotype. However, some genes produce no observable phenotype when deleted. For example, only ~34% of genes in the yeast *Saccharomyces cerevisiae* genome are essential for normal growth under standard laboratory conditions ((Hillenmeyer et al., 2008) and references therein). This genetic redundancy can be explained in part by the functional redundancy of duplicate genes, and also by genetic buffering between disparate genes in functionally related pathways ((Gu et al., 2003) and references therein). One approach towards elucidating the biological roles of functionally redundant genes is the systematic exploration of double mutant (or double gene knockdown) combinations, and their resultant phenotypes, to define genetic interactions of functionally related genes. This approach has been used successfully in a number of systems to uncover gene functions that were otherwise concealed (Byrne et al., 2007; Costanzo et al., 2010; Horn et al., 2011). Although systematically investigating the synthetic phenotypes of redundant gene pairs is a powerful means of uncovering their function, this approach is not technically feasible for all organisms and it does not circumvent functional redundancy of three or more genes. Clearly, alternative approaches for the study of redundant genes are required.

Chemical genetics provides a means for sidestepping genetic redundancy. For example, a promiscuous small molecule inhibitor of two or more functionally redundant proteins can reveal their function. From a phenotype-based chemical screen, Chen et al. (2006) identified “pluripotin”, a small molecule that allows murine embryonic stem cells to be propagated in an undifferentiated, pluripotent state in the absence of feeder cells, serum, and leukemia
inhibitory factor (Chen et al., 2006). Pluripotin acts by inhibiting both ERK1 kinase and RasGAP, and this dual inhibition is critical for its function, since the genetic knockdown or chemical inhibition of either protein alone is insufficient to maintain the cells in an undifferentiated, pluripotent state. Thus, pluripotin unveils a function for ERK1 and RasGAP in embryonic stem cell differentiation.

In addition to promiscuous inhibitors, a selective small molecule agonist (or activator) can reveal the function of one member of a normally redundant protein family. For example, the functionally redundant *PYR/PYL* receptors for abscisic acid (ABA), an endogenous plant hormone that regulates development and adaptive stress responses, were found using a seed-selective small molecule agonist of *PYR1* called “pyrabactin” (Park et al., 2009). Pyrabactin was identified from a forward chemical genetic screen in *Arabidopsis thaliana* for small molecules that inhibit germination and cell expansion, and thus phenocopy ABA-treated seeds. Additionally, ABA- and pyrabactin-treated seeds have a highly correlated pattern of gene expression, suggesting they activate a common receptor. The *PYR1* target of pyrabactin was uncovered through the isolation of pyrabactin-resistant mutant lines and subsequent map-based cloning of *Pyr1* (*pyrabactin*-resistance 1) mutant alleles. Many of the *Pyr1* alleles are putative nulls, suggesting that pyrabactin is a *PYR1*-selective agonist. A search of the genome revealed 13 additional genes with sequence similarity to *Pyr1*, called *Pyl1*-*Pyl13* (for *Pyr1*-like), representing a family of START domain-containing proteins. Data from genetic and biochemical experiments with ABA strongly suggest that the *PYR/PYL* proteins are receptors for ABA signaling (Nishimura et al., 2009; Park et al., 2009). Of note, more than 20 years of classical genetic interrogation of the ABA signaling pathway failed to identify the ABA receptor protein(s), likely because ABA is a non-selective agonist of the functionally redundant *PYR/PYL*
receptors, and forward genetic screens are unlikely to uncover mutations in two or more functionally redundant genes.

Small molecules can also reveal redundant gene function on a genome-wide scale. For example, Hillenmeyer et al. (2008) treated the yeast “non-essential” gene deletion collection with over 400 distinct small molecules (and other environmental stressors), and found that 97% of genes that are non-essential in normal media are required for growth in at least one experimental condition (Hillenmeyer et al., 2008). In this case, chemical-genomic interrogation of an organism revealed a function for almost every gene in its genome. Thus, small molecules are powerful reagents to unmask the biological roles of functionally redundant genes.

1.1.5 Probing Temporal Requirements of Protein Function with Small Molecules

An important utility of small molecule probes is the ease with which they can be applied and removed from the system being studied, enabling temporal and kinetic analyses of protein activity. Conversely, genetic knockouts and null alleles are present throughout the life of an organism, and may reveal only the earliest roles of the gene in development. Conditional regulation of gene activity can be achieved by using temperature-sensitive (ts) mutant alleles (O'Rourke et al., 2011) or inducible transgenes (Lewandoski, 2001), however ts alleles are not available for every gene, and inducible transgenes may not afford the tight temporal “gene off, gene on” resolution required for the process under investigation (Kim et al., 2008).

Furthermore, small molecules can often be applied under standard assay conditions to cells or organisms that are genotypically wild-type. In contrast, modifications to the environment (i.e. temperature shifts), and changes to the genetic background are usually needed for conditional
gene regulation by ts alleles and inducible transgenes, which can confound the interpretation of results.

There are a number of examples in the literature of how small molecules have been used to define the temporal requirements of proteins for a variety of biological processes (Kim et al., 2008; Mayer et al., 1999; O'Connor et al., 2011; Spring, 2005; Wang et al., 2003; Wolff et al., 2003; Yeh and Crews, 2003). One particularly notable example is the chemical-genetic dissection of how the level and timing of hedgehog (Hh) signaling specifies three distinct muscle cell types (the muscle pioneers (MPs), the superficial slow fibers (SSFs), and the medial fast fibers (MFFs)) in the developing zebrafish (Wolff et al., 2003). Using a reverse chemical genetic approach, Wolff et al. (2003) treated zebrafish embryos with cyclopamine, a known small molecule antagonist of the Hh transducer protein smoothened, at different concentrations and times, and looked at the identities of cells in the developing myotome. They found that the highest levels of Hh signaling are required for MP specification, followed by slightly lower levels for MFF induction, and that SSF specification occurs in response to the lowest Hh levels of the three. Furthermore, MP specification precedes that of the MFFs, requiring early and continuous Hh signaling, while Hh signaling at a relatively later stage is sufficient for MFF induction. Notably, the authors’ fine tuning of the timing and magnitude of Hh activity with cyclopamine would not have been possible with traditional genetic methods in the fish.
1.2 Probe Discovery in an Optimized Chemical Space

1.2.1 What is Chemical Space?

The ensemble of all theoretically possible molecules has been termed “chemical space” (Dobson, 2004), an analogy based on the “astronomical space” in which our planet resides. Our physical universe is expansive, and it is made up primarily of empty space. Similarly, the chemical universe is vast, comprised of upwards of $10^{60}$ synthesizable small organic molecules (Bohacek et al., 1996), and most of these compounds likely have no biological use, assuming even the most modest selectivity of proteins for their ligands (Hert et al., 2009). In fact, small molecule chemical space is so big that we could never hope to synthesize every molecule (i.e. the combined mass of 10μg of each compound would outweigh the observable astronomical universe (Hert et al., 2009)). Therefore, for the purposes of probe discovery, it is critically important to define the region of chemical space that is of biological relevance and utility (Fig. 1-3), which will help guide the synthesis of novel chemical matter and inform the construction of small molecule screening libraries.

1.2.2 Biogenic Bias of Screening Libraries

Given the enormous size of small molecule chemical space and the likelihood that most of this space has no biological value, it is perhaps surprising that chemical screens succeed at all in identifying biologically active molecules. One reason chemical-genetic and drug discovery screens bear fruit is that the molecules used for screening are structurally biased towards the chemistry of the natural world (Hert et al., 2009). The biogenic bias of purchasable chemical libraries was quantified by Hert et al. (2009) who analyzed the structures of a generated database (GDB) comprised of all 26 million synthetically possible molecules containing 11 or
Figure 1-3. Finding an optimal chemical space for probe discovery. In the oversimplified diagram above, the rectangle outlines the space occupied by all theoretically possible molecules (the arrows imply the vastness of this space). There are regions within chemical space comprised of small molecules that are bioavailable to the intended organism or cell type (white circle). There is a distinct, but overlapping, area of chemical space populated by molecules that can modulate protein activity (black circle). Of the molecules that affect protein activity, some may act through covalent modifications (green circle), some may act non-selectively (yellow circle), and some may be toxic to biological organisms (red circle). It is the grey area of chemical space that is of greatest value for probe discovery, because it is occupied by non-toxic compounds that can access target proteins in vivo and selectively modulate their activity via non-covalent mechanisms. The diagram is not to scale.
fewer first-row atoms, as well as the purchasable subset of these molecules, and compared
them to the structures of metabolites and natural products retrieved from the KEGG and DNP
databases, respectively (Hert et al., 2009). They found that the purchasable subset of the GDB
is almost 1,000-fold more structurally similar to the natural products and metabolites, relative
to the entire GDB. Thus, the success of chemical-genetic and drug discovery screens likely
results from the screening of molecules that resemble naturally occurring protein ligands,
which, by design, occupy a biologically relevant chemical space.

Although the biogenic bias of screening libraries is largely serendipitous, the idea of
synthesizing chemical libraries with molecular scaffolds that resemble natural products is not
new (Breinbauer et al., 2002; Ertl et al., 2008; Hert et al., 2009; Newman and Cragg, 2007;
Wetzel et al., 2011). An interesting insight from the analysis of the biogenic bias found in
purchasable screening libraries is that, despite the bias, only 17% of natural product molecular
scaffolds are found in the purchasable chemical space (Hert et al., 2009). Therefore, 83% of
naturally occurring scaffolds are not found in purchasable libraries, providing a starting point
for the intentional biasing of future compound collections towards bioactivity.

1.2.3 Structural Diversity of Screening Libraries

An important consideration for the construction of a chemical library is the structural diversity
of its constituent molecules (O’Connor et al., 2012). The diversity of the chemical space
screened is related to the diversity of the target space probed. Obviously, screening the same
molecule (or very similar molecules) over and over again will identify ligands for very few
proteins. Thus, it is important to maximize the molecular diversity of the chemicals screened
for probe discovery.
Four main components of molecular diversity have been suggested (Galloway et al., 2010):

1. Scaffold diversity, which is the variety of molecular “skeletons” or “backbones” present in the screening library; 2. Appendage diversity, which refers to the assortment of branches, side chains, or chemical moieties that decorate common scaffolds; 3. Functional group diversity, which defines the range of functional groups present in the compounds; 4. Stereochemical diversity, which is the variation in the spatial orientation of the appendages and functional groups stemming from the core scaffolds. On the whole, it is the molecular shape diversity of screening libraries that is most important, since small molecules fit into the binding pockets of their target proteins in three-dimensional space. All four diversity parameters described contribute to a molecule’s three-dimensional shape; however, it has been shown that the central scaffold diversity is the most critical determinant of three-dimensional shape diversity, with the appendages and functional groups being of relatively minor importance (Sauer and Schwarz, 2003). Therefore, the scaffold diversity of screening libraries has the most impact on functional diversity.

Many commercially available chemical libraries are constructed via combinatorial synthesis methods (Krier et al., 2006; O’Connor et al., 2012). Combinatorial chemistry allows for the fast and simultaneous synthesis of large chemical libraries by bringing together sets of reactive modules in such a way that compounds with every modular combination can be formed (Sauer and Schwarz, 2003; Thompson and Ellman, 1996). In most cases, one of the modular building blocks is constant, and by default is the “scaffold” for that set of combinatorially synthesized structures. Not surprisingly, large commercially available combinatorial libraries have historically suffered from a lack of scaffold diversity (Krier et al.,
2006), despite considerate appendage diversity, and as a result these libraries likely suffer from a lack of functional diversity as well.

One approach for the synthesis of chemical libraries with diverse scaffolds has been termed diversity-oriented synthesis (DOS) ((Galloway et al., 2010; O'Connor et al., 2012) and references therein). Similar to combinatorial synthesis, DOS of chemical libraries can be achieved in a modular and simultaneous manner. However, modules with distinct chemistries from those used in combinatorial synthesis are employed to generate libraries with a large fraction of unique core scaffolds. For example, a starting module might contain many different functional (or reactive) groups that exhibit unique condition-specific reaction chemistries (see (Galloway et al., 2010; O'Connor et al., 2012) and references therein). Using distinct reaction conditions, these functional groups can be paired intramolecularly to create diverse scaffolds. Chemical screens with libraries constructed using DOS have uncovered modulators of proteins that had no known prior ligands (O'Connor et al., 2012). For instance, a target-based screen of ~10,000 DOS compounds for molecules that bind Sonic hedgehog (Shh) protein in vitro identified the molecule robotnikinin (Stanton et al., 2009). Downstream analyses with robotnikinin suggest that it inhibits Shh in vivo. Previous phenotype-based screens with chemical libraries derived from combinatorial synthesis failed to identify an antagonist of the Shh protein.

1.2.4 Quality Control for Screening Libraries

There is a subset of molecules in chemical space that should be omitted from chemical screening libraries, in order to avoid false positives in target-based screens, as well as in vivo toxicities (Kazius et al., 2005; Rishton, 2003). Compounds that lead to false positives in target-
focused screens include protein-reactive electrophiles (Rishton, 2003), “frequent hitters” (Roche et al., 2002), and “promiscuous inhibitors” (McGovern et al., 2002). Protein-reactive electrophiles act through the covalent modification of target proteins, and are typically reactive towards biological nucleophiles including the target protein, serum protein and glutathione. Also, electrophilic false positives are generally unstable, and are prone to decomposition by solvolysis or hydrolysis, providing further impetus for their removal from chemical libraries. Electrophilic groups that should be avoided in chemical screens have been identified (Rishton, 2003), and they should be omitted from small molecule screening collections. “Promiscuous inhibitors” act by self-associating into colloidal aggregates in the aqueous screening media, and these aggregates then adsorb target proteins non-selectively and inhibit activity through protein denaturation (Coan et al., 2009; McGovern et al., 2002). Promiscuous inhibitors are a major source of false positives in target-based screening. The formation of colloidal aggregates is disrupted by treatment with detergent, and a simple detergent-based method has been developed that can be used to assay for promiscuous inhibition and rule out this type of false positive (Feng and Shoichet, 2006). It has recently been shown that colloidal aggregation can reduce the efficacy of bioactive compounds in cell culture (Owen et al., 2012), suggesting that this process is relevant for phenotype-based screens as well. “Frequent hitters” are molecules that show up as screening hits in many different assays against a variety of protein targets (Rishton, 2003; Roche et al., 2002). Frequent hitters can be promiscuous inhibitors (as discussed above), or molecules that disrupt the screening assay; such as fluorescent molecules in a fluorescence-based screen. Computational models have been developed that identify frequent hitters based on their structures (Rishton, 2003; Roche
et al., 2002), and these computer-based methods can be used to filter out unwanted false positives from screening libraries.

Molecules that are generally toxic to biological organisms should also be avoided in screening libraries. Toxic compounds can perturb living systems in non-specific ways (e.g. through the covalent modification of cellular macromolecules (Kazius et al., 2005; Williams and Naisbitt, 2002)), and are generally not ideal candidates for biological probes or drugs. In some cases, there are molecular substructures that are common to toxic compounds, referred to as toxicophores (Kazius et al., 2005). For example, aromatic nitro and sulfonate bonded carbon groups are often found in mutagenic molecules (Kazius et al., 2005). Indeed, ethyl methane sulfonate is a mutagen commonly used in forward genetic screens. There is a plethora of software available for the in silico structure-based prediction of toxic molecules (Click2Drug, ; Liao et al., 2011). These computational tools can be used to filter out toxicophores from screening libraries.

1.2.5 Bioavailability and the Lipinski Rule-of-Five

A critical parameter for the in vivo efficacy of a bioactive compound is target accessibility. In 1991, the main cause for clinical attrition of pharmaceutical drug candidates was adverse pharmacokinetic and bioavailability properties, accounting for 40% of all attrition (Kola and Landis, 2004). By 2000, the percentage of clinical failures attributed to adverse pharmacokinetics decreased 4-fold and poor bioavailability was no longer the root cause for attrition. Efforts to define property-based rules of small-molecule drug absorption and permeation likely contributed to the observed decrease (van de Waterbeemd and Gifford, 2003; Zhu et al., 2011). A particularly influential example is Lipinski’s “Rule-of-Five”, which was
built by analyzing the physicochemical properties of a subset of the world drug index that achieved clinical exposure (Lipinski et al., 1997). The hypothesis was that “poorer physicochemical properties would predominate in the many compounds that enter into and fail to survive pre-clinical stages and Phase I safety evaluation”. Lipinski et al. found that small molecules will generally have unfavorable absorption and permeation properties if they have a calculated octanol/water partition coefficient (clogP) greater than 5, a molecular weight above 500, and more than five hydrogen bond donors or more than 10 hydrogen bond acceptors (Lipinski et al., 1997). The Rule-of-Five serves as a useful property-based computational tool to exclude molecules from high-throughput screens that are likely to have poor oral bioavailability in humans, and loosely defines a human orally-bioavailable chemical space. Not surprisingly, the majority of purchasable chemical libraries are purposefully designed to conform to the Rule-of-Five (Irwin and Shoichet, 2005).

Bioavailability considerations are of particular importance for whole-organism phenotype-based chemical screens, in which target access is limited by the pharmacokinetics of the compounds being screened. Lipinski’s Rule-of-Five describes a chemical space with good oral absorption properties for humans, but it likely does not define a chemical space with optimal bioavailability for other species. For example, a comparison of the physicochemical properties between human orally bioavailable drugs and antibiotics against four distinct pathogenic species of bacteria reveals that these molecules occupy organism-dependent areas of chemical space (Hopkins and Bickerton, 2010). Thus, phenotype-based chemical screens in model organisms would likely benefit from the construction of their own species-specific bioavailability filters. These filters could then be applied to enrich screening libraries with compounds that are more likely to access potential target proteins in vivo, in the screening
system of choice. To my knowledge, the modeling of bioavailable chemical spaces for non-mammalian model organisms has never been done.

1.3 Probe Discovery and Characterization using C. elegans

1.3.1 Caenorhabditis elegans: A Premier Animal Model

The nematode worm Caenorhabditis elegans is a well established model organism used to investigate the biology of animals. Over the past 35+ years, experiments with C. elegans have uncovered novel principles fundamental to animal biology (Ambros, 1989; Austin and Kimble, 1989; Brenner, 1974; Ellis and Horvitz, 1986; Fire et al., 1998; Hedgecock et al., 1990; Lee et al., 1993; Yochem and Greenwald, 1989). The utility of the worm for gene discovery and characterization is a consequence of its many virtues. C. elegans is genetically tractable and easy to culture due to its small size (~1mm fully grown), short lifecycle (3 days), simple dietary requirements (E. coli and cholesterol), hermaphroditism, diploidy, and isogenicity. The anatomy, cell lineage, and neural connectivity of the worm have been described in detail (Sulston and Horvitz, 1977; Sulston, 1983; White et al., 1986). The worm is optically transparent, which facilitates the observation of cell components with or without the use of fluorescent labels and dyes. The C. elegans genome sequence is complete, revealing a high degree of genetic conservation between worms and mammals (Consortium, 1998; Ruvkun and Hobert, 1998); thus, information garnered through C. elegans genetic experiments is translatable to more complex organisms. Furthermore, the growing list of C. elegans disease models ((Kaletta and Hengartner, 2006) and references therein) positions the worm as a tool for the discovery of novel drugs to treat human ailments.
1.3.2 Probe Discovery with *C. elegans*

Historically, *C. elegans* has been employed as a gene discovery tool using classical genetic approaches. However, the worm has great potential as a chemical-genetic model system, particularly for the discovery of novel small molecule probes via phenotype-based screens (Fig. 1-4A). This potential has not gone unnoticed, and a number of compounds with novel bioactivities have been identified from *C. elegans*-based chemical screens (Breger et al., 2007; Gosai et al., 2010; Hihi et al., 2008; Kwok et al., 2006; Lemieux et al., 2011; Min et al., 2007; Petrascheck et al., 2007; Samara et al., 2010). Since the worm is only 1mm in length fully grown and has a short 3 day lifecycle, *C. elegans* small molecule screens can be performed relatively quickly in multi-well plates, at any life stage, over multiple generations. In this regard, *C. elegans* distinguishes itself from other chemical-genetic model systems, since it is the only animal model against which libraries of small molecules can be screened for bioactivity in the context of the whole animal at any life stage. Furthermore, *C. elegans* has a short ~3-week lifespan, which permits screens for life-extending chemicals. For instance, Petrascheck *et al.* (2007) performed a large-scale screen of 88,000 compounds to identify molecules that extend the lifespan of adult *C. elegans* (Petrascheck *et al.*, 2007). One molecule identified from their screen extended the life of worms by 20%, and is structurally similar to known serotonin receptor antagonists. The authors found that worms treated with the FDA-approved serotonin receptor antagonist mianserin also live longer than untreated worms, and that its activity is likely mediated through the dual antagonism of worm serotonin and octopamine receptors (the putative *C. elegans* starvation-sensing receptors (Horvitz *et al.*, 1982; Niacaris and Avery, 2003; Suo *et al.*, 2006)). Experiments combining food restriction with mianserin treatment suggest that mianserin extends lifespan through ageing mechanisms.
Figure 1-4. High-throughput screening of chemicals for bioactivity and target identification in *C. elegans*. (A) The workflow for a high-throughput chemical screen in *C. elegans*. Liquid handling, worm sorting, and imaging can be fully automated with the use of robots and automated imaging platforms. (B) Upon identifying a compound that is bioactive in *C. elegans*, forward genetic screens can be performed to identify candidate pathways, targets, and docking sites of the molecule of interest (see Burns et al. (2006) for more details). In the diagram, pink represents a solution of the compound of interest.
associated with dietary restriction, although mianserin treatment alone does not affect feeding. Thus, a small molecule screen in *C. elegans*, and the downstream characterization of a “hit” compound, implicates the perception of starvation, despite adequate food intake, in the regulation of ageing. Notably, such a large-scale screen for compounds that extend the life of adult animals could not have been performed, in a reasonable timeframe, with any model system other than *C. elegans*.

Several recent technological advances will continue to propel *C. elegans* into the forays of high-throughput chemical screening (Burns et al., 2006; Giacomotto and Segalat, 2010; Gosai et al., 2010; O’Rourke et al., 2009). There are a number of commercially available automated liquid-handling systems that can be used to standardize the addition of media and chemicals to screening plates, and our group has made use of such technologies for genetic screens in *C. elegans* (Byrne et al., 2007). Worm deposition in *C. elegans*-based screens can be automated using the COPAS Biosort (Union Biometrica) (Pulak, 2006). This machine sorts worms based on user-defined selection criteria, such as size and fluorescence, and then precisely dispenses defined numbers of worms into the wells of multi-well assay plates in a high-throughput fashion. The COPAS Biosort has been used by a number of groups to automate worm handling for both chemical and genetic screens in *C. elegans* (Burns et al., 2006; Byrne et al., 2007; Gosai et al., 2010; Okoli et al., 2009). Methods for automated image acquisition and analyses of *C. elegans* samples cultured in multi-well plates have been developed (Buckingham and Sattelle, 2008, 2009; Burns et al., 2006; Gosai et al., 2010). Our lab uses a high-throughput digital imager (HiDi 2100; Elegenics) to automate plate handling and image acquisition of worms on top of solid media in multi-well plates (Burns et al., 2006). Combined, these technologies facilitate a throughput that might be considered high even by industrial standards (i.e., tens of
thousands of chemicals per day (Gosai et al., 2010)), transforming *C. elegans* into a platform for high-throughput, whole-organism, phenotype-based small molecule probe discovery.

### 1.3.3 Small Molecule Mode-of-Action Studies using Worms

As I described, the power of *C. elegans* as an animal model system lies in the relative simplicity of its biology and genetics. This simplicity can be taken advantage of to define the mode of action of bioactive molecules. Tens of thousands of mutagenized *C. elegans* worms can be screened for mutants that are resistant to the effects of a given bioactive molecule in a relatively short timeframe (Fig. 1-4B). Depending on the design, these forward genetic screens may yield mutations in the target’s biochemical pathway, the target itself, or even residues within the target that comprise the small molecule’s binding site (Burns et al., 2006). Forward genetic screens in *C. elegans* have been used successfully to characterize the modes of action for a number of bioactive molecules (Allard and Colaiacovo, 2010; Fitzgerald et al., 2006; Jones et al., 2005; Kaminsky et al., 2008; Kokel et al., 2006; Kwok et al., 2008; Kwok et al., 2006; Lackner et al., 2005; Petrascheck et al., 2007). For example, the protein targets of the novel class of anthelmintics, called amino-acetonitrile derivatives (AADs), were discovered through a screen for resistant *C. elegans* mutants (Kaminsky et al., 2008; Rufener et al., 2010). Kaminsky *et al.* obtained 44 AAD-resistant mutants from a screen of *C. elegans* mutant F2s, of which 36 fell into a single complementation group. Genetic mapping and subsequent DNA sequencing revealed 27 independent mutations in the *C. elegans* **acr-23** gene. **acr-23** encodes a nicotinic acetylcholine receptor (nAChR) of the DEG-3 group. The authors found that isolates of the parasitic nematode *Haemonchus contortus* that are resistant to AADs had lost at least part of the **acr-23** gene homolog and at least part of the **des-2** gene homolog (another member of the
DEG-3 group of nAChRs). These results suggest that the AADs function by activating nAChRs of the DEG-3 subfamily. Indeed, subsequent electrophysiological analysis has shown that the AADs act as positive allosteric modulators of *H. contortus* DEG-3/DES-2 channels (Rufener et al., 2010). One of the AADs, monepantel, is currently marketed as an anthelmintic therapy for sheep (Kaminsky et al., 2009). The use of *C. elegans* in defining the mode of action of the AADs is a clear example of the utility of this simple worm for identifying the targets of bioactive compounds.

Forward genetic screening for mutants resistant to bioactive molecules is a powerful approach to target identification, and can be used to determine the mode-of-action of compounds that induce relatively non-specific phenotypes, such as lethality, that do not immediately suggest candidate target proteins. However, the careful characterization of chemical-induced phenotypes can often implicate candidate protein targets, by exploiting the existing knowledge of specific gene perturbations and their resultant phenotypes in the organism being studied (Titov and Liu, 2012). The genetic perturbation of *C. elegans* has been ongoing for the last 35+ years, and has revealed phenotypes that are characteristic of specific gene mutations. This information is archived and readily available for public use (WormBase), allowing for a cross comparison of chemical-induced phenotypes with the existing knowledge base. For example, a forward chemical-genetic screen in our lab recently uncovered a small molecule named *dafadine* that causes *C. elegans* worms to constitutively enter a long-lived and stress-resistant alternative developmental stage called dauer (Luciani et al., 2011). In addition to the constitutive dauer (Daf-c) phenotype, dafadine also causes worms to have protruding vulvae (Pvl) and distal-tip cell migration (Mig) defects. The *C. elegans* genes *daf-9* and *daf-12* are the only two worm genes known to cause a penetrant combination of Daf-c, Pvl, and Mig
phenotypes when mutated (Antebi et al., 1998; Gerisch et al., 2001; Jia et al., 2002), suggesting that one or both of their protein products could be the target of dafadine. The DAF-9 protein is a cytochrome p450 enzyme that synthesizes cholesterol-like hormones called dafachronic acids (see Fig. 1-1 for the structure) that inhibit dauer formation through their interaction with the nuclear hormone receptor DAF-12 (Motola et al., 2006). The ability of dafadine to inhibit the activity of the DAF-9 or DAF-12 proteins was tested using functional cell-based assays, revealing that dafadine is a DAF-9 inhibitor. Thus, in addition to forward genetic approaches, the wealth of publicly available C. elegans genetic, phenotypic, and biochemical data can be mined to inform the mode-of-action of novel bioactive compounds.

1.4 Research Problem: Overcoming the Resistance of C. elegans to Pharmacological Perturbation

C. elegans is free-living for its entire life cycle and has adapted to survive the dynamic stresses of a soil environment. As a consequence, C. elegans has evolved extensive physical and enzymatic defenses to resist the uptake and accumulation of exogenous compounds in its tissues. The physical barriers comprise an intestine through which solutes are rapidly pumped (Avery and Shtonda, 2003) and a four-layered collagenous cuticle that lines its exterior, as well as its oral and rectal cavities (Cox et al., 1981). In addition, the C. elegans genome encodes a number of putative xenobiotic detoxification genes including 86 cytochrome P450 (CYP) and 72 UDP-glucuronosyltransferase (UGT) enzymes (Lindblom and Dodd, 2006), as well as 15 P-glycoprotein (PGP) and nine multidrug resistance protein (MRP) pumps, many of which are implicated in xenobiotic efflux (Sheps et al., 2004). The expression of detoxifying genes can be
induced by xenobiotics in *C. elegans* (Laing et al., 2010; Menzel et al., 2005) and this induction is likely mediated by one or more of the 284 *C. elegans* nuclear hormone receptor (*nhr*) genes (Lindblom and Dodd, 2006).

Not surprisingly, the worm is generally resistant to perturbation by drug-like pharmacologically active molecules, despite the conservation of many known and well-characterized pharmacological targets (Kwok et al., 2006) (Fig. 1-5A). Furthermore, phenotype-based chemical screens in zebrafish embryos, commonly considered to be “drug-permeable”, have higher hit rates than *C. elegans* chemical screens when comparable experimental conditions are considered (Burns et al., 2006; Kwok et al., 2006; Peterson et al., 2000) (Fig. 1-5B). Given its extensive xenobiotic defenses, a reasonable explanation for this resistance is that exogenous molecules have limited bioavailability in *C. elegans* tissues. Indeed, molecules having no effect on whole worms can perturb their target proteins if they are provided direct access (Franks et al., 2002; Jospin et al., 2002; Ruiz-Lancheros et al., 2011). Thus, target accessibility is likely a major limiting factor for drug efficacy in *C. elegans*, although this idea has never been tested experimentally. The obvious concern is that *C. elegans* chemical screens will miss potential small-molecule modulators of protein function, simply because the compounds screened do not achieve effective concentrations at their target.

Here, I set out to answer two specific questions: 1) Does poor pharmacological efficacy in *C. elegans* result from a general resistance to the accumulation of exogenous drug-like compounds in its tissues? 2) If so, can a “worm-bioavailable” subset of drug-like chemical space be defined, and will selecting molecules from this space improve the rate at which bioactive molecules are identified from *C. elegans* chemical screens?
Figure 1-5. *C. elegans* is resistant to pharmacological perturbation. (A) All of the 1,280 molecules included in Sigma’s Library of Pharmacologically Active Compounds (LOPAC) have established activity at one or more protein receptor. When tested against *C. elegans* at a relatively high concentration of 25μM, only 3.8% of the LOPAC molecules produce an observable effect on growth, behaviour, or morphology despite many conserved pharmacological targets in the worm (Kwok et al. (2006), Burns et al. (2006)). (B) A small molecule developmental screen in zebrafish resulted in a 70% higher hit rate relative to a similar screen in *C. elegans*, despite a 25-fold lower screening concentration (Kwok et al. (2006), Burns et al. (2006), Peterson et al. (2000)). In both screens, molecules from a Chembridge DIVERSet collection were tested for the induction of obvious developmental and morphological defects.
To answer these questions, I developed a high-throughput high performance liquid chromatography (HT-HPLC) assay that measures the accumulation of exogenous molecules in worm tissue, and tested the accumulation of hundreds of drug-like molecules in *C. elegans*. In Chapter 2 I describe the results of this experiment. In brief, I find that worms are resistant to drug-like compound accumulation, and that accumulation is generally required for bioactivity. Using my small molecule accumulation data, a machine-learned model was built that selects molecules which have a high likelihood of accumulating in worms based on their structures. I show that pre-selecting “worm-bioavailable” molecules using my structure-based accumulation model improves the rate at which bioactive compounds are identified from chemical screens in *C. elegans*. Thus, I establish poor bioavailability as a mechanism by which worms resist pharmacological perturbation, and describe a worm-bioavailable chemical space from which small-molecules can be selected to achieve higher hit rates from *C. elegans* phenotype-based chemical screens.
Chapter 2

Developing a Structure-Based Model to Predict Small Molecule Accumulation in C. elegans

The work in this chapter was performed by me with a few exceptions. Victoria Wong provided technical assistance with bulk worm culture under my supervision in the Roy lab. Computational analyses were performed by Iain Wallace and Jan Wildenhain at the University of Toronto, with guidance from me, my supervisor (Peter Roy), Mike Tyers, Gary Bader, Guri Giaever, and Corey Nislow. Mass spectrometry (MS) and tandem MS-MS measurements of small molecule metabolites were performed by Songqin Pan at the University of California, Riverside (subsequent data analysis was performed by me). Accurate mass MS was performed by Alex Young of the Advanced Instrumentation for Molecular Structure laboratory at the University of Toronto. My supervisor (Peter Roy) and committee member Sean Cutler provided guidance for all aspects of the work presented here, particularly at the outset. The work in this chapter has been published as: Burns, A.R., Wallace, I.M., Wildenhain, J., Tyers, M., Giaever, G., Bader, G.D., Nislow, C., Cutler, S.R., Roy, P.J. (2010) A predictive model for drug bioaccumulation and bioactivity in Caenorhabditis elegans. Nat Chem Biol 6(7): 482-3.
Chapter 2. Developing a Structure-Based Model to Predict Small Molecule Accumulation in C. elegans

Abstract
The resistance of Caenorhabditis elegans to pharmacological perturbation limits its use as a screening tool for novel small bioactive molecules. One strategy to improve the hit rate of small-molecule screens is to preselect molecules that have an increased likelihood of reaching their target in the worm. To learn which structures evade the worm’s defenses, I performed the first survey of the accumulation and metabolism of over 1,000 commercially available drug-like small molecules in the worm. I discovered that fewer than 10% of these molecules accumulate to concentrations greater than 50% of that present in the worm’s environment. Using my dataset, a structure-based accumulation model was developed that identifies compounds with an increased likelihood of bioavailability and bioactivity, and I describe structural features that facilitate small-molecule accumulation in the worm. Preselecting molecules that are more likely to reach a target by first applying the model to the tens of millions of commercially available compounds will undoubtedly increase the success of future small-molecule screens with C. elegans.

2.1 Introduction
The many virtues of the tiny nematode C. elegans make it a promising model system for the discovery and characterization of novel bioactive compounds. Because of the worm’s rapid life cycle, small size and hermaphroditism, libraries of small molecules can be screened for
bioactivity in the context of the whole animal and over its entire life cycle in a high-throughput fashion (Burns et al., 2006). Furthermore, the power of C. elegans genetic analysis has repeatedly uncovered the mechanism of action of both small-molecule tools (Kokel et al., 2006; Kwok et al., 2008; Kwok et al., 2006; Petrascheck et al., 2007) and novel anthelmintics (Jones et al., 2005; Kaminsky et al., 2008). The worm also shares extensive genetic conservation with more complex animals, and the growing list of human disease models in the worm further distinguish C. elegans as a unique tool for the discovery of novel therapeutics (Kaletta and Hengartner, 2006).

Unfortunately, C. elegans is relatively resistant to perturbation by pharmacologically active molecules. For example, pharmacological agents must often be applied to the worm at concentrations that are orders of magnitude higher than those used in mammalian cell culture (Broeks et al., 1995; Choy and Thomas, 1999; Kwok et al., 2006; Rand and Johnson, 1995). Moreover, our lab has found that only 2% of pharmacologically active compounds can induce a robust phenotype in the worm when screened at a concentration of 25 μM (Burns et al., 2006; Kwok et al., 2006). Screening compounds at a higher concentration may overwhelm the worm’s xenobiotic defenses, but doing so can be prohibitively costly and would likely result in molecules precipitating out of solution in the screening medium. Circumventing the resistance of C. elegans to bioactive compounds would increase its utility as a small-molecule screening tool.

C. elegans has extensive physical and enzymatic xenobiotic defenses that may render many pharmacological tools ineffective. The physical barriers include a four-layered cuticle that lines its exterior and oral and rectal cavities (Cox et al., 1981), as well as an intestine through which solutes are rapidly pumped (Avery and Shtonda, 2003). The worm’s genome is
replete with predicted xenobiotic detoxification enzymes, including 86 cytochrome P450s, and 60 ATP-binding cassette transporters, many of which likely function as xenobiotic efflux pumps (Lindblom and Dodd, 2006). Compounds that are ineffective when applied to whole animals can readily antagonize their targets if they are provided with direct access (Franks et al., 2002; Jospin et al., 2002; Kwok, 2006). Hence, it is likely that C. elegans is generally resistant to exogenously applied pharmacologicals because they fail to accumulate to effective concentrations within its tissues. Modeling the properties of small molecules that promote accumulation in the worm would greatly facilitate the discovery of new biological probes and drug leads.

Typically, only a small fraction of the more than 13 million commercially available small molecules (Irwin and Shoichet, 2005) is screened by any one academic laboratory. Many bioactive compounds therefore remain undiscovered within the unscreened fraction of purchasable chemical space. For instance, the largest C. elegans chemical screen to date used 88,000 compounds (Petrascheck et al., 2007), which covers less than 0.7% of available chemical space. Even this relatively large-scale screen leaves more than 99% of all possible bioactives undiscovered within the unscreened fraction of available chemical space. Hence, it is essential to maximize the number of bioactive compounds in the fraction of molecules that is screened. Given that bioavailability is a prerequisite to bioactivity, one way to increase the hit rate of a chemical screen is to develop a property-based computer model that predicts small-molecule accumulation within the screening system of choice. The model can then be used to select molecules with an increased likelihood of bioavailability from the purchasable chemical space and improve the chances of finding a hit.
In this chapter, I describe the development of a high-throughput HPLC-based technique to measure the bioaccumulation of exogenous compounds and their metabolites in *C. elegans*. This method was used to assay the accumulation of more than 1,000 commercially available drug-like molecules in whole animals. Using the accumulation data, a machine-learned model was constructed that distinguishes accumulating from non-accumulating compounds on the basis of their structural properties. When applied to two naive chemical libraries, the structure-based accumulation model enriches for structurally distinct compounds with diverse bioactivities in the worm.

### 2.2 Results

#### 2.2.1 An HPLC-Based Small Molecule Accumulation Assay for *C. elegans*

As a first step toward developing a predictive model of bioaccumulation in *C. elegans*, I surveyed over 1,000 commercially available drug-like molecules for their ability to accumulate in worm tissue. To do this, I modified a previously described HPLC-based assay (Kwok et al., 2006) to a 96-well-plate format (Materials and Methods 2.4.2 and 2.4.3). Briefly, I incubated ~5,000 fourth-larval-stage worms in 40 μM of compound for 6 h in filter plates. I then drained the incubation buffer from the wells, washed the worms, re-suspended them in fresh buffer, transferred them to new plates and then lysed the worms chemically. I used a reverse-phase HPLC system coupled with a diode array detector (HPLC-DAD) to separate and visualize the components of the worm lysate (Fig. 2-1A and Materials and Methods 2.4.2 and 2.4.3). This method measures small-molecule bioaccumulation in *C. elegans*, as opposed to bioavailability, because it detects compounds that are taken up by the worms during incubation and then
Figure 2-1. A survey of exogenous drug-like molecule accumulation in *C. elegans*. (A) Heat-mapped HPLC-DAD chromatograms. Retention time is shown on the x-axis, and absorbance wavelength is shown on the y-axis. The scale of absorbance intensity, in milli-absorbance units (mAU), is shown on the right. The DMSO peak (red arrow), peak of worm contents (green arrows), and small-molecule peak (yellow arrows) are indicated. Molecules that have a retention time of less than 1.25 minutes or absorb light exclusively between 190 and 220 nm cannot be detected with our protocol because they would be masked by either the peak of worm contents, or by the solvent, respectively. (B) Pie charts showing the fraction of accumulating structures for 387 compounds from the Spectrum library (Microsource), and for 23 nematicides derived from a 10K DIVERSet library (Chembridge) and the 1K HitsKit library (Maybridge).
remain in the worms after washing. Similar HPLC approaches have been developed to identify drugs and drug metabolites for clinical and forensic purposes (Herre and Pragst, 1997; Herzler et al., 2003). Though the use of a diode array detector limits the detection space to compounds with conjugated systems of pi bonds (for example, aromatics), I chose this approach over a mass spectrometry–based approach because of the cost effectiveness and relatively high-throughput nature of the HPLC-DAD technique.

Here, I define an accumulating molecule as one that (i) is detectable by my HPLC protocol in at least two of three replicate lysates, (ii) is undetectable in the no-worm sham trials (to control for compounds that precipitate in the filter-plate wells (Materials and Methods 2.4.4) and (iii) remains detectable after the worms are washed with the highest concentration of sodium dodecyl sulfate (SDS) that does not result in any obvious physiological changes (to control for small-molecule association with the cuticle (Materials and Methods 2.4.5)). To control for accumulated contaminants or spontaneous oxidation products of the parent molecule, compounds that accumulate as putative metabolites (see below) are reprocessed using dead worms to ensure that the putative metabolites are dependent on living worms (Materials and Methods 2.4.6).

As a proof of principle, I used my high-throughput (HT) HPLC method to examine the accumulation of 21 members of the 1,4-dihydropyridine (DHP) family of L-type calcium channel antagonists after either a 0.5- or 6-h co-incubation with worms (Fig. 2-2). Using a low-throughput approach, our lab previously examined the accumulation of 12 of these DHPs after a 2 h incubation period and found that four of the 12 DHPs accumulate, including the three bioactive molecules nemadipine-A, nemadipine-B and felodipine, and one inactive DHP called analog 7 (Kwok et al., 2006). Using my HT-HPLC approach, all three bioactive DHPs were found
Figure 2-2. The 1,4-dihydropyridine structures analyzed for accumulation in worms. Compounds in red accumulate in the worm after a 6 h incubation in 40μM of the molecule. Compounds with an asterisk induce egg laying defective, abnormal morphology, and slow growth phenotypes in wild-type worms at a 25μM concentration. The compound with a cross induces a slow growth phenotype at a 25μM concentration in wild-type worms. Compounds 1-12 have been assayed at least in triplicate for phenotype in wild-type worms at a 25μM concentration. Compounds 13-21 have not been assayed for bioactivity in the worm.
to accumulate at both time points, as well as DHP analog 7 at the 0.5-h time point and DHP analog 1 at the 6-h time point. Hence, my HT-HPLC protocol can robustly detect the accumulation of exogenous small molecules in the worm, and the results obtained using the HT method correlate well with the low-throughput approach.

2.2.2 A Survey of Drug-Like Small Molecule Accumulation in Worms

There are currently more than 13 million commercially available small molecules (Irwin and Shoichet, 2005), a quantity that greatly exceeds what could likely be sampled in any single C. elegans screen. Therefore, it is important to carefully select the compounds that will be screened in order to maximize the number of bioactives obtained. Property-based modeling of small-molecule accumulation in worms could be used to prioritize purchasable chemical space so as to enrich for compounds that have a greater likelihood of accessing a biologically relevant target. Even though the overwhelming majority of commercially available compounds obey Lipinski’s rule-of-five and are considered “drug-like” (Irwin and Shoichet, 2005; Lipinski et al., 1997), the rate at which molecules are identified as bioactive against C. elegans remains relatively low (Burns et al., 2006; Kwok et al., 2006). To better understand the molecular properties of commercially available drug-like molecules that facilitate their accumulation in C. elegans, I used my HT-HPLC assay to measure the accumulation of the Spectrum collection of 2,000 pharmacologically active compounds (MicroSource Inc.) in whole worms. I chose the Spectrum library because it is enriched for popular small bioactive compounds that include human drugs and pharmacological tools (Kocisko et al., 2003), and the majority of these compounds obey Lipinski’s rule-of-five. In addition, our lab has previously investigated this collection for bioactivity in wild-type worms (Burns et al., 2006; Kwok et al., 2006), which
allowed me to compare bioaccumulation and bioactivity. Of these 2,000 Spectrum molecules, 1,096 were detectible when I processed 7.5 nmol of each molecule using our HPLC-DAD system.

Next, I determined which of the 1,096 detectable Spectrum molecules accumulate in wild-type worms after a 6-h incubation. I found that 96 compounds (9.3%) accumulate out of a set of 1,027 (69 molecules were eliminated from consideration because the sample was either lost during processing or considered a false positive after control experiments). I call this dataset the “complete Spectrum dataset” (Fig. 2-3A). The majority of accumulating compounds (64%) are found in worms at concentrations that are less than the 40-μM assay concentration, with a median concentration of 26 μM (Fig. 2-4). I discovered a bias in my accumulation dataset such that molecules with relatively low limits of detection are enriched for accumulating compounds, and molecules with relatively high detection limits are enriched for non-accumulating compounds (see Materials and Methods 2.4.7 to 2.4.10 for the approach taken to estimate the limits of detection for the compounds in the dataset and for my analysis of bias). I therefore applied a 19-μM detection-limit cutoff to the complete Spectrum dataset to make an unbiased assessment of the fraction of small drug-like molecules that accumulate in worms (Materials and Methods 2.4.10 and 2.4.11). I found that only 26 (6.7%) of the 387 molecules that passed the cutoff accumulate using this unbiased criterion (Fig. 2-1B). These data support the conclusion that worms are generally resistant to the accumulation of exogenous drug-like molecules and provide a foundation to investigate the properties that govern small-molecule accumulation in worms.
Figure 2-3. The complete small molecule accumulation dataset (1,132 compounds). (A) The complete Spectrum dataset (1027 molecules). (B) The nematicides (27 molecules). (C) The 1,4-dihydropyridine (DHP) family (21 molecules). (D) The Fenbufen analogs (4 molecules). (E) The LOPAC (Sigma) molecules (53 molecules). P - Parent Structure, M - Metabolite, P/M – Parent and/or Metabolite.
Figure 2-4. The distribution of small molecule concentrations in worms for the accumulating compounds in the complete Spectrum dataset. The concentrations of the most abundant accumulating species (i.e. parent or metabolite) for each molecule were used to generate the histogram. 7 of the 96 accumulating molecules accumulate only as unquantifiable metabolites, due to their proximity to the peak of worm contents (green arrows in Fig. 2-1A), and were not included in this distribution.
2.2.3 Exogenous Compounds Can Accumulate as Metabolites in C. elegans

Of the molecules that I have found to accumulate in worms, 36% accumulate as metabolites. The majority of these metabolites have a similar spectral absorption profile as their respective parental compounds but typically elute from the reverse-phase HPLC column faster than the parental molecule. To verify that the metabolites are bona fide derivatives of the parental compounds, 17 metabolites from 12 parental compounds were examined by mass spectrometry and tandem MS-MS (Table 2-1, Fig. A-1, and Materials and Methods 2.4.12). For 11 of the 12 compounds, the analyses show that the metabolites are parent derived, indicating that the majority of the novel peaks observed on the HPLC chromatograms represent bona fide metabolites of the parental compound. For one of the 12 compounds (compound 12 in Table 2-1), I could not identify abundant masses specific to the purified metabolite fractions relative to the DMSO control fractions.

I deduced the modification(s) made to several of the parental compounds on the basis of the mass of the metabolites and their fragmented derivatives in the MS-MS analysis (Table 2-1 and Fig. A-1). The modifications made to parental compounds 4, 5 and 6 are consistent with drug metabolism in mammals whereby phase I enzymes functionalize the drug through demethylation or reduction for subsequent metabolism by phase II conjugating enzymes (Eddershaw and Dickins, 2004; Manchee et al., 2004). The diol metabolite of compound 8 likely results from the activity of the phase I enzymes cytochrome P450 and epoxide hydrolase. For compounds 1, 2, 3 and 7, only conjugated derivatives were found, including glucosidated and sulfated derivatives, suggesting that these compounds are modified directly by phase II–like enzymes. To verify the correctness of my proposed biotransformations, accurate mass determinations were performed for four representative metabolites (Table 2-2 and Materials
Table 2-1. Characterization of *C. elegans* xenobiotic metabolites.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Parent Structure</th>
<th>Parent (P) Molecular Weight</th>
<th>Metabolite 1 (M1) Mass</th>
<th>Metabolite 1 Enzymatic modification</th>
<th>Metabolite 2 (M2) Mass</th>
<th>Metabolite 2 Enzymatic modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>240</td>
<td>402</td>
<td>O-hexosidation&lt;sup&gt;a&lt;/sup&gt; [P + 162]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>284</td>
<td>446</td>
<td>O-glucosidation&lt;sup&gt;b&lt;/sup&gt; [P + 162]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>244</td>
<td>406</td>
<td>O-hexosidation [P + 162]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.png" alt="Structure 4" /></td>
<td>238</td>
<td>386</td>
<td>Demethylation O-glucosidation&lt;sup&gt;a,b&lt;/sup&gt; [P – 14 + 162]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td><img src="image5.png" alt="Structure 5" /></td>
<td>225</td>
<td>211</td>
<td>Demethylation [P – 14]</td>
<td>373</td>
<td>O-glucosidation&lt;sup&gt;b&lt;/sup&gt; [M1 + 162]</td>
</tr>
<tr>
<td>6</td>
<td><img src="image6.png" alt="Structure 6" /></td>
<td>282</td>
<td>284</td>
<td>Reduction&lt;sup&gt;c&lt;/sup&gt; [P + 2]</td>
<td>446</td>
<td>O-glucosidation&lt;sup&gt;b,d&lt;/sup&gt; [M1&lt;sub&gt;284&lt;/sub&gt; + 162]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>268</td>
<td>Demethylation&lt;sup&gt;d&lt;/sup&gt; [P – 14]</td>
<td>430</td>
<td>O-glucosidation&lt;sup&gt;b,d&lt;/sup&gt; [M1&lt;sub&gt;268&lt;/sub&gt; + 162]</td>
</tr>
<tr>
<td>7</td>
<td><img src="image7.png" alt="Structure 7" /></td>
<td>213</td>
<td>455</td>
<td>N-glucosidation N-sulfation [P + 162 + 80]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td><img src="image8.png" alt="Structure 8" /></td>
<td>240</td>
<td>274</td>
<td>Aryl epoxidation Epoxide hydrolysis&lt;sup&gt;e&lt;/sup&gt; (dil formation) [P + 16 + 18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td><img src="image9.png" alt="Structure 9" /></td>
<td>210</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td><img src="image10.png" alt="Structure 10" /></td>
<td>196</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td><img src="image11.png" alt="Structure 11" /></td>
<td>208</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td><img src="image12.png" alt="Structure 12" /></td>
<td>284</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<sup>a</sup> – Accurate mass was measured for these four metabolites (Table 2-2)
<sup>b</sup> – β-glucosidase digestions confirm these metabolites to be O-linked β-glucosides (Fig. 2-5 and Materials and Methods 2.4.14)
<sup>c</sup> – These metabolites were co-purified by HPLC
<sup>d</sup> – These metabolites were co-purified by HPLC
<sup>n/a</sup> – Abundant masses specific to these metabolite fractions, relative to controls, could not be found

A dash indicates that the metabolite is a bona fide derivative of the parent molecule; however the respective biotransformation could not be unambiguously determined
Table 2-2. Accurate mass determinations for 4 representative xenobiotic metabolites.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Adduct</th>
<th>Number of Charges</th>
<th>Elemental Composition</th>
<th>Calculated m/z (amu)</th>
<th>Measured m/z (amu)</th>
<th>PPM Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_M1</td>
<td>+Na</td>
<td>1</td>
<td>C_{21}H_{22}O_{8}Na</td>
<td>425.1206</td>
<td>425.1208</td>
<td>0.2</td>
</tr>
<tr>
<td>4_M1</td>
<td>+Na</td>
<td>1</td>
<td>C_{21}H_{22}O_{7}Na</td>
<td>409.1257</td>
<td>409.1265</td>
<td>1.7</td>
</tr>
<tr>
<td>6_M2_{268}</td>
<td>+H</td>
<td>1</td>
<td>C_{22}H_{23}O_{9}</td>
<td>431.1336</td>
<td>431.1357</td>
<td>4.7</td>
</tr>
<tr>
<td>8_M1</td>
<td>+H</td>
<td>1</td>
<td>C_{14}H_{15}N_{2}O_{4}</td>
<td>275.1026</td>
<td>275.1030</td>
<td>1.3</td>
</tr>
</tbody>
</table>

a – The structures and the mass spectrometry analysis of the metabolites can be found in Table 2-1 and Figure A-1
For all four metabolites the measured masses are within 5 p.p.m. of the calculated monoisotopic masses, indicating that the elemental compositions are correct. Furthermore, metabolites of compounds 2, 4, 5, and 6 were confirmed to be O-linked β-glucosides by β-glucosidase digestion (Fig. 2-5 and Materials and Methods 2.4.14). To my knowledge, this work provides the first in vivo survey of xenobiotic metabolism in *C. elegans* and reveals many similarities of drug metabolism between worms and mammals.

### 2.2.4 A Predictive Structure-Based Accumulation Model (SAM)

The majority of the purchasable drug-like compounds I have so far tested fail to accumulate in worms. Given that most commercially available chemical libraries are designed to have similar drug-like properties as the molecules I have tested (Irwin and Shoichet, 2005), I sought to better understand the attributes of these compounds that influence their accumulation in worms and to develop a generally applicable model that can predict bioavailability on the basis of these attributes.

To build a predictive small-molecule structure-based accumulation model (SAM) for worms, I consolidated all of the 1,132 molecules I have assayed for accumulation into one dataset (Fig. 2-3). After applying the 19-μM detection limit cutoff and removing duplicate structures, 483 molecules remained in the SAM training set, of which 74 accumulate in worms (Fig. 2-6A). These 483 compounds were used to train an ECFP_4-Naive Bayesian classifier to distinguish accumulating from non-accumulating molecules on the basis of their structural features (Fig. 2-6B and Materials and Methods 2.4.15). ECFP_4 is a two-dimensional circular fingerprint descriptor that represents each compound as a series of small fragments that are built by starting from each non-hydrogen atom in the compound and extending out up to four
Figure 2-5. β-glucosidase digestions of select xenobiotic metabolites. Metabolites 2_M1 (A), 4_M1 (B), 5_M2 (C), 6_M2 (D), 1_M1 (E), 3_M1 (F), 5_M1 (G), and 7_M1 (H) were treated with β-glucosidase (see Table 2-1 and Fig. A-1 for the proposed structures). HPLC chromatograms are shown for the parent structure alone, the products of mock digestions without β-glucosidase, and the products of β-glucosidase digestions. Metabolites 5_M1 and 7_M1 served as negative controls for the digestions. Retention time and absorbance scales are the same as in Figure 2-1A.
Figure 2-6. A machine-learned model predicts the accumulation of exogenous small molecules in C. elegans. (A) A 483-molecule training set was used for machine-learning. (B) Building the ECFP_4-Bayesian structure-based accumulation model (SAM). $F_{\text{accum}}$ is the number of accumulating compounds in which the fragment is present, $F_{\text{total}}$ is the total number of compounds in which the fragment is present, $P(\text{acc})$ is the probability that a compound accumulates in the training set, and the Laplacian correction ($lc$) = $1/P(\text{acc})$. An accumulation score is generated for any molecule by summing the scores for each fragment in the molecule. (C) The distribution of accumulation scores for the accumulating and non-accumulating subsets of molecules in the training set. (D) 5-fold cross-validation of the model. Test sets were ranked by the model and the enrichment of accumulating compounds, relative to random, was plotted for the top-scoring 5%, 10%, 20%, 50%, and 100% of compounds. Enrichment for a perfect model was also plotted relative to random. Error bars represent ± 1 standard deviation. (E) Biological validation of the SAM. A set of compounds from a 50K DIVERSet library (Chembridge) were ranked using the SAM, and three subsets were chosen at random from the top-scoring 5%, the bottom-scoring 5%, and the entire set (i.e. random-scoring molecules), respectively. These molecules were tested for accumulation in worms using our HT-HPLC method. Pie charts show the fraction of accumulating molecules for each subset.
bond lengths. In total, 4,698 fragments were derived from the 483 compounds in the training set using the ECFP_4 fingerprint descriptor. The Naive Bayesian classifier identifies the fragments that are over-represented and under-represented in the accumulating subset of molecules (Rogers et al., 2005; Xia et al., 2004). Over-represented fragments receive positive scores, and under-represented fragments receive negative scores. The accumulation score of a molecule is then calculated by summing up the scores of its respective fragments. There was relatively little overlap in the scores of accumulating and non-accumulating subsets of the 483-molecule training set (Fig. 2-6C), indicating that there are fragments that distinguish accumulating from non-accumulating structures in worms.

The predictive power of the SAM was estimated by five independent five-fold cross-validation experiments (Fig. 2-6D). On average, the SAM performs about three times better than random; the top scoring 5% and 10% of molecules are 3.5- and 2.9-fold enriched for accumulation compared to randomly selected compounds, respectively, indicating that the SAM can be successfully applied to independent datasets. When compared to three additional Naive Bayesian models, trained using distinct molecular descriptors (MDL PublicKeys (Durant et al., 2002), physicochemical properties and Lipinski properties (Lipinski et al., 1997)), the SAM outperforms all three as determined by five-fold cross-validation (Fig. 2-7).

To biologically validate the SAM and to assess its general applicability, it was used to rank the compounds of a 50K DIVERSet library (Chembridge Corp.) and I experimentally tested its predictions. As a first step, an in-house predictive model of UV-visible absorbance was applied to the 50K DIVERSet library to increase the probability that any chosen molecule could be detected by the diode array detector (Materials and Methods 2.4.16). Second, library molecules that were >85% similar to any molecule in the SAM training set were excluded from
Figure 2-7. Comparison of Naïve Bayesian classifier machine-learned models trained using distinct molecular descriptors. Using the 483-compound training set for machine-learning (Figure 2-6A), four distinct Naïve Bayesian classifiers were generated using four different molecular descriptors: 1. ECFP_4 fingerprints (see Materials and Methods 2.4.15), 2. MDL PublicKeys (Durant et al. (2002)), 3. Physicochemical properties (hydrophobicity, molecular weight, molecular polar surface area, molecular solvent accessible area, # of hydrogen-bond acceptors, # of hydrogen-bond donors, # of rotatable bonds, # of rings, # of aromatic rings), 4. Lipinski properties (Lipinski et al. (1997)). Five independent 5-fold cross-validations were performed for each model, and the enrichment of accumulating compounds in the test sets was plotted for the top-scoring 5%, 10%, 20%, and 50% of molecules. Error bars represent the standard error of the mean.
consideration. After applying these two filters, the SAM was applied to the remaining 4,993 DIVERSet molecules. I randomly selected 23 molecules from the top-scoring 5%, 29 molecules from the bottom-scoring 5% and 28 molecules from the entire set (which I refer to as random-scoring molecules). I assayed the accumulation of these 80 molecules in worms using my HT-HPLC method. After controls, and after applying the 19-µM detection limit cutoff, 15 of the top-scoring, 25 of the bottom-scoring and 18 of the random-scoring compounds remained in the test set. I found that only 1 out of 25 (4%) of the bottom-scoring molecules accumulate, representing a 2.8-fold under-enrichment relative to the random-scoring compounds (Fig. 2-6E). By contrast, I found that 9 out of 15 (60%) of the top-scoring molecules accumulate, representing a 5.5-fold enrichment compared to the random-scoring compounds and a 15-fold enrichment relative to the bottom-scoring molecules (Fig. 2-6E). These results indicate that my structure-accumulation model can be applied to diverse libraries and successfully predict the accumulation of exogenous drug-like compounds in worms on the basis of their structural features.

2.2.5 Structural Features that Influence Accumulation

To identify which of the 4,698 ECFP_4 fragments most strongly influence the accumulation of exogenous compounds in worms I compiled the top-scoring and bottom-scoring five structurally non-redundant and non-overlapping fragments from the top and bottom 30 features learned by the SAM (Fig. 2-8 and Materials and Methods 2.4.17). Strikingly, the top five and bottom five fragments are present in ~50% of the accumulating and non-accumulating compounds, respectively (Fig. 2-8A,B).
Figure 2-8. Prominent substructures that influence small-molecule accumulation in *C. elegans*. *(A)* The top 5 non-overlapping fragments learned by the structure-based accumulation model (T1 to T5). An atom denoted with an “A” could be one or more of the atoms C, N, O, or S (where appropriate). Dashed lines indicate double bond character. Pie charts show the fraction of accumulating and non-accumulating compounds in the training set with each of the 5 top-scoring non-overlapping fragments. *(B)* The bottom 5 non-overlapping fragments learned by the structure-based accumulation model (B1 to B5). Pie charts show the fraction of accumulating and non-accumulating compounds in the training set with each of the 5 bottom-scoring non-overlapping fragments. *(C-F)* Clustergrams of compounds containing the “unfused” biaryl scaffold (C), the methyl piperazine scaffold (D), the 2- or 3-phenyl-chromen-4-one scaffolds (E), and the “fused” biaryl scaffold (F). Each column, represented by a number or a dash on top of the box, represents a distinct molecule of the class depicted by the structural scaffold above. The enrichment of accumulating compounds and the hypergeometric p-value associated with that enrichment are shown at the bottom of each clustergram.
To interpret how the top five features facilitate accumulation, I considered the structural scaffolds from which they were derived within the SAM training set. Four scaffold types enriched for accumulating molecules were identified: (i) the ‘unfused’ biaryl scaffold, (ii) the methyl piperazine scaffold, (iii) the 2- or 3-phenyl-chromen-4-one scaffolds and (iv) the ‘fused’ biaryl scaffold (Fig. 2-8C-F). Seventy-two distinct compounds in the training set are composed of these four scaffolds, and they account for 41% of the accumulating compounds in the training set. Notably, 10 out of 12 accumulating ‘unfused’ biaryl structures accumulate as metabolites in the worm, and all but one induce penetrant lethality at 25 μM, suggesting that this scaffold may be a useful building block for the development of novel nematicides.

Three of the bottom five fragments (B1, B2 and B4 in Fig. 2-8B) are derived from compounds that contain a carboxylic acid group, an aliphatic hydroxyl group or a sulfonyl group. It is not surprising that these fragments are over-represented in the non-accumulating subset of molecules, for a number of reasons. First, these features contain hydrogen and oxygen atoms that can make hydrogen bond contacts with water molecules, promoting solubility in aqueous buffer and preventing cell membrane permeability. Second, these functional groups are known to be sites of phase II conjugation reactions in mammals, such as glucuronidation, glycosylation and sulfation (Manchee et al., 2004). Thus, compounds with these features are more likely to be modified and excreted by the organism. In agreement with these data, an analysis of the Lipinski property distributions for the accumulating and non-accumulating subsets of the training set shows that the accumulating compounds generally have fewer hydrogen bond donors and acceptors than the non-accumulating molecules and that accumulating structures typically have a greater LogP than non-accumulating structures (Fig. 2-9). Despite these differences, the Lipinski properties perform the least well of all the
Figure 2-9. Lipinski property distributions for the 483-compound training set for machine-learning. The distributions of the four Lipinski properties (LogP, molecular weight, # of hydrogen bond acceptors, # of hydrogen bond donors) are plotted for the accumulating and non-accumulating subsets of the training set for machine-learning. aLogP is a computationally derived estimate of the LogP of an organic compound.
approaches tested to predict accumulation (Fig. 2-7). It is presently unclear how the remaining two fragments (B3 and B5 in Fig. 2-8B) might antagonize small-molecule accumulation in the worm.

I observed that compounds with scaffolds that facilitate accumulation will generally accumulate in worms unless they have one or more of the bottom-scoring fragments B1–B5 (Fig. 2-8C-F). Hence, the features that antagonize small-molecule accumulation generally act as the ‘master’ determinants of accumulation. An example of how the negative features influence bioaccumulation is provided by my study of the biphenyl fenbufen, a nonsteroidal anti-inflammatory drug (Kerwar, 1983), and three fenbufen analogs (Fig. 2-10 and Materials and Methods 2.4.18). Fenbufen and fenbufen analog 3 both contain carboxylic acid groups and fail to accumulate in worms. By contrast, fenbufen analogs 1 and 2 both lack carboxylic acid groups and accumulate as metabolites (compounds 9 and 10 in Table 2-1). The fenbufen analogs that accumulate as metabolites also induce penetrant lethality at 25 uM, but the non-accumulating analogs do not.

2.2.6 The SAM Enriches for Compounds with Distinct Bioactivities

I previously found that the only DHPs that show bioactivity are those that can accumulate in worm tissue (Fig. 2-2), suggesting that bioaccumulation is generally required for bioactivity. Additional observations made here further support this idea. First, only those fenbufen analogs that accumulate are bioactive (Fig. 2-10). Second, only two molecules from my unbiased Spectrum dataset are bioactive in the worm, and both accumulate (Fig. 2-3A). Finally, 17 out of 23 previously described nematicides (Burns et al., 2006; Kwok et al., 2006) accumulate to concentrations greater than 19 μM in the worm (Fig. 2-1B).
Figure 2-10. Phenotypic and accumulation analysis of Fenbufen and three analogs. (A) The structures of Fenbufen and the Fenbufen analogs. (B) Dose-response curves for the four Fenbufen analogs. The number of adult worms in each well four days after being seeded with two L4 stage animals is shown. The error bars represent the standard error of the mean for three trials. (C) HPLC chromatograms of the lysates of worms incubated for 6 h in 40 µM of the four Fenbufen analogs. 7.5 nmol of each Fenbufen analog was processed by HPLC and the peak corresponding to the parent structure of each analog is indicated with a green arrow. The major accumulating metabolites of Fenbufen Analogs 1 and 2 are indicated with white arrows. Axes and legend are the same as that in Figure 2-1A.
Given that bioaccumulation is correlated with bioactivity, I anticipated that the SAM would also enrich for bioactive molecules in *C. elegans*. To test whether the SAM can enrich for molecules that are bioactive in the worm, it was used to score the 10K DIVERSet library (Chembridge Inc.) that our lab previously screened for the induction of gross phenotypes in wild-type worms (Burns et al., 2006; Kwok et al., 2006). The molecules that are structurally distinct from those in the SAM’s training set (\( n = 9,740 \)) were ranked according to their accumulation score and then I determined whether the top-scoring molecules are enriched for phenotype. Thirty percent (14 out of 47) of the bioactive molecules in the library were present in the top-scoring 5% of the compounds, representing a six-fold enrichment of bioactives (\( P < 3.5 \times 10^{-8} \), Fig. 2-11A). To further test the SAM’s ability to enrich for compounds with distinct bioactivities, it was used to rank 1,040 compounds that were screened by the National Institute of Neurological Disorders and Stroke (NINDS) for the correction of neuronal defects in a worm model of Huntington’s disease (PubChem BioAssay, AID: 1599). The top 5% of molecules that are structurally distinct from the compounds in our SAM’s training set (\( n = 826 \)) are 3.8-fold enriched for bioactive molecules relative to random (\( P < 0.02 \), Fig. 2-11A). Hence, applying the SAM to naive libraries can greatly increase the efficiency by which novel bioactive molecules are identified.

### 2.2.7 SAM-Selected Molecules are Structurally Diverse

Applying structure-based filters to a chemical library will inevitably narrow its structural diversity. I therefore investigated the impact of the SAM on structural diversity using three different approaches using the Tanimoto molecular similarity scoring system with ECFP_4 fingerprints (Materials and Methods 2.4.15). Pairwise Tanimoto coefficients range from 0,
Figure 2-11. The *C. elegans* structure-based accumulation model (SAM) enriches for structurally diverse compounds with distinct bioactivities in the worm. (A) Phenotype enrichment curves for the molecules in the top-scoring 5% of the 10K DIVERSet (Chembridge) and NINDS (PubChem BioAssay, AID: 1599) libraries ranked by the SAM. The percentage of actives obtained as the compounds are traversed from the highest to the lowest ranking compound is shown. The enrichment value and the hypergeometric p-value associated with the enrichment are indicated for the top-scoring 5%. The compounds in the 10K DIVERSet library were scored previously for any obvious gross phenotypes in wild-type *C. elegans* (Burns et al. (2006), Kwok et al. (2006)). The compounds in the NINDS library were scored for the correction of neuronal defects in a *C. elegans* model of Huntington's disease (PubChem BioAssay, AID: 1599). (B) Structural diversity of the compounds in (A). The distributions of all pair-wise Tanimoto similarity scores are shown for the top-scoring 5% of compounds, as well as a random-scoring 5%, from the 10K DIVERSet and NINDS libraries. Three different random-scoring distributions were generated for each library; due to considerable overlap only one distribution is shown for clarity. (C) Scaffold composition of the compounds in (A). Murcko scaffold similarity networks are shown for the compounds in the top-scoring 5% of the 10K DIVERSet and NINDS libraries. Scaffolds (nodes) are connected if they have a pair-wise ECFP_4/Tanimoto similarity score ≥ 0.7. Red nodes indicate active compound scaffolds, whereas grey nodes indicate inactive compound scaffolds. (D) Similarity networks for the active compounds in the top-scoring 5% of the 10K DIVERSet and NINDS libraries. Compounds (red nodes) are connected if they have a pair-wise ECFP_4/Tanimoto similarity score ≥ 0.3. Network visualization was performed using Cytoscape (Cline et al. (2007)).
which indicates the absence of structural similarity, to 1, which indicates identity. In practice, Tanimoto scores ≤0.2 are so low that they do not represent any meaningful structural similarity (Flower, 1998; Hert et al., 2009).

As a first approach to assessing the diversity of SAM-selected molecules, all pairwise Tanimoto coefficients were calculated for the top-scoring 5% and a random-scoring 5% of compounds from each of the 10K DIVERSet and NINDS libraries (the distributions are shown in Fig. 2-11B). The average pairwise similarity scores for the top-scoring compounds from the DIVERSet and NINDS libraries are 0.138 and 0.112, respectively, with 87% and 94% of the pairwise scores being ≤0.2, respectively. The average pairwise similarity scores for the random-scoring compounds from the DIVERSet and NINDS libraries are 0.110 and 0.089, respectively, with 96% and 97% of the pairwise scores being ≤0.2, respectively. Although the random-scoring molecules are expectedly more diverse, the SAM does not simply reduce the library to a small number of structurally similar compounds.

Next, the diversity of the core scaffolds present in the SAM-selected compounds was investigated. Murcko scaffolds (Bemis and Murcko, 1996), which retain the ring systems and linkers of molecules but eliminate the side chains, were generated for the top-scoring 5% of molecules and three random sets of an equivalent number of molecules from each of the 10K DIVERSet and NINDS libraries. Similarity networks were created for the top-scoring and the random-scoring sets of scaffolds, in which scaffolds are connected if they have a pairwise ECFP_4 or Tanimoto score ≥0.7 (the top-scoring scaffold similarity networks are shown in Fig. 2-11C). Scaffold networks generated in this way have been used to explore the scaffold composition of purchasable screening libraries (Shelat and Guy, 2007). I then counted the number of unique scaffold clusters in each network, including unconnected singletons. The
top-scoring DIVERSet scaffolds are comprised of 290 distinct scaffold clusters (Fig. 2-11C), and the random-scoring DIVERSet scaffolds are comprised of an average of 380 (±3.5) distinct scaffold clusters. The top-scoring NINDS scaffolds are composed of 33 distinct scaffold clusters (Fig. 2-11C), and the random-scoring NINDS scaffolds are composed of an average of 30 (±5.0) distinct scaffold clusters. Hence, this analysis shows that although the application of the SAM can reduce the scaffold diversity of a given library, the top-scoring 5% of molecules (~500 structures) from my analyses still represent hundreds of unique scaffolds.

Finally, the structural diversity of the bioactive molecules in the top-scoring 5% of molecules from the 10K DIVERSet and NINDS libraries was assessed. Similarity networks were built for these bioactive compounds by linking compounds with a Tanimoto coefficient of 0.3 or more, as a high degree of structure-phenotype concordance has been observed in cell-based assays for pairs of molecules that have a Tanimoto coefficient of 0.3 or greater (Hoon et al., 2008; Young et al., 2008). Nine out of 14 (64%) active compounds in the top-scoring DIVERSet subset are singletons in the network, and two out of four (50%) actives in the top-scoring NINDS subset are singletons in the network (Fig. 2-11D). Furthermore, the majority of the bioactive compounds retrieved in the top-scoring 5% of the 10K DIVERSet (11 out of 14) and NINDS (four out of four) libraries have distinct core scaffolds (Fig. 2-11C). Hence, my analyses of the structural diversity of the SAM-selected compounds suggest that the SAM enriches for both accumulating and bioactive molecules without dramatically limiting the structural diversity of the compounds screened or the hits obtained.

The SAM can be used directly by opening the script found in Additional Data File 1 using Pipeline Pilot software (see Materials and Methods 2.4.15). Alternatively, the training set (Additional Data File 2) can be used in conjunction with Open Source software (Additional Data
File 3) to build a similar model. An Excel spreadsheet containing all of the raw data compiled over the course of this project can be found in Additional Data File 4.

### 2.3 Discussion and Conclusion

The general resistance of *C. elegans* to pharmacological perturbation has impeded its utility as a chemical-genetic model system. My systematic analysis of drug-like compound accumulation has shown for the first time that poor bioavailability is a major contributing factor to this resistance. The property-based modeling described here enables the identification of molecules that have an increased likelihood of accumulating in worms and, by extension, have an increased likelihood of affecting a biologically relevant target. Hence, by screening molecules that are more likely to accumulate, the xenobiotic defenses of the worm can be circumvented to identify new biological probes and potential drug leads. The *C. elegans* SAM provides a tool by which commercially available molecules can be prioritized to increase the probability of identifying structurally distinct compounds with diverse bioactivities in worms, resulting in more efficient screens. The approach of generating computer-based models to identify molecules with a higher likelihood of bioactivity can be applied to other systems as well (such as planaria, *Drosophila* and zebrafish) and could serve as a new paradigm for the design of model organism chemical screens.

There are additional important considerations when applying the SAM to novel compound sets in the future. First, the model will work best when applied to libraries with feature distributions that are similar to those of the training set used to generate the model. In the same vein, the structural fragments learned by the model do not represent an exhaustive list, and there are likely other structural features that influence small-molecule bioavailability in...
worms that were not sampled in our analysis. Sampling of more compounds from an increasingly diverse chemical space will undoubtedly improve the prediction coverage of the model. Finally, the enrichment-of-bioactives and diversity analyses were performed on only two libraries—the only two publicly accessible, large-scale screening datasets for *C. elegans*. When applying the SAM to new compound sets, the enrichment rates and structural diversities of top-scoring compounds will depend on the characteristics of the compound sets being ranked and the phenotypes assayed.

The structural scaffolds I have identified provide the first guidelines to better design new small-molecule libraries intended for chemical screens with *C. elegans* and perhaps other nematodes. Notably, the biaryl scaffold, the piperazine scaffold and the chromenone substructure in the 2- or 3-phenyl-chromen-4-one scaffolds have all been previously identified as privileged substructures (Horton et al., 2003; Klekota and Roth, 2008). A privileged substructure is defined as “a single molecular framework able to provide ligands for diverse receptors” (Evans et al., 1988; Horton et al., 2003). Specificity can be achieved by varying the substituents that decorate the privileged scaffold (Chen and Shoichet, 2009; Hajduk et al., 2000; Mason et al., 1999). For example, the biphenyl scaffold, which accounts for one-third of the accumulating ‘unfused’ biaryls in worms, is found in 4.3% of all known drugs, representing molecules from diverse therapeutic classes (Hajduk et al., 2000; Horton et al., 2003). Indeed, statistical analysis of NMR-derived binding data for 10,080 compounds (represented by 104 substructural fragments) and 11 protein targets identified the biphenyl scaffold as a privileged substructure that preferentially binds proteins (Hajduk et al., 2000). These results are encouraging, as molecular scaffolds that promote bioavailability in *C. elegans* also promote physical interactions with diverse protein targets.
My analysis of nematicide accumulation in worms revealed that 40% of these compounds accumulate as metabolites in the worm. This result suggests that compounds may be processed to their bioactive form by the worm, which is analogous to how prodrugs like heroin and levodopa are metabolized to their bioactive form in humans (Garzon-Aburbeh et al., 1986; Inturrisi et al., 1983). Hence, for a given bioactive compound discovered through an in vivo screen, it is not a certainty that the parent structure is the bioactive species. This insight is especially relevant to those who seek the target of a compound using biochemical approaches, which may not be fruitful if the bioactive agent is a metabolite of the parental compound. Thus, accumulation analysis of bioactive compounds identified from whole organism screens should be considered as a first step in the biochemical characterization of target proteins.

There is little information regarding the mechanisms by which nematodes modify and detoxify xenobiotics. It is known that the C. elegans genome encodes a number of conserved detoxification genes (Lindblom and Dodd, 2006), and that the expression of these genes is upregulated in response to treatment with xenobiotics (Laing et al., 2010; Menzel et al., 2001; Menzel et al., 2005), however biochemical characterization of drug metabolites in worms remains scant. My survey of xenobiotic metabolism is the first of its kind for C. elegans, and the modifications I identified are consistent with mammalian phase I and II drug biotransformations. However, in contrast to mammals, worms seem to favour the conjugation of glucose to xenobiotics, as opposed to glucuronic acid, a common modification in mammalian drug metabolism. Interestingly, in a study published subsequent to this work, it was shown that C. elegans glucosylates the anthelmintic albendazole (Laing et al., 2010). Therefore, glucosylation may be a favoured mechanism by which C. elegans, and possibly other nematodes, modify and detoxify xenobiotics.
Relative to the HPLC protocols used with other model systems (Hou et al., 2004), the chromatographic output of my worm HT-HPLC assay yields a high signal-to-noise ratio because the majority of endogenous worm material elutes in the flow through. In combination with the wide spectral range of the diode array detector, the low background of my assay allows for the easy identification and quantification of small molecules and their metabolites from worm lysates. Of course, the method could be improved in a number of ways. First, my assay is restricted to molecules that absorb UV-Vis light. To accommodate a broader structural diversity and to improve detection limits, prospective approaches should consider detection methods that do not rely on the absorbance of light (e.g. evaporative light scattering or mass spectrometry). Second, the method does not provide information regarding the tissue distribution of drugs in *C. elegans*. To that end, molecules with intrinsic fluorescence properties could be used to better understand the relationship between chemical structure and localization.

Machine-learned structure-based models trained on bioactivity data, as opposed to accumulation data, have also been used to enrich compound sets with bioactive molecules (Wallace et al., 2011). Subsequent to the publication of the work reported here, my collaborator Iain Wallace and I constructed a machine-learned structure-based bioactivity model. The 9,740 Chembridge molecules that I previously assayed for bioactivity in *C. elegans*, of which 47 are worm-bioactive, were used as a training set. The model was built using the same methods used to build the SAM. We used this structure-based bioactivity model to rank the same NINDS molecules that were used as a test set for the SAM, and we found that the top 5% of compounds that are structurally distinct from the training set are 2.8-fold enriched for bioactives relative to random. In contrast, the SAM achieved a 3.8-fold enrichment of
bioactives in the top-scoring 5% of the NINDS test set. Thus, in this case, the SAM outperforms the bioactivity model. One reason for the SAM’s superior performance could be that the SAM training set contains 60% more positives than the bioactivity training set (i.e. 74 accumulating molecules versus 47 bioactives). Whatever the case may be, as more C. elegans small-molecule bioactivity data becomes publicly available, relatively more predictive structure-based bioactivity models can be developed to complement future SAMs.

Here, I have established my high-throughput HPLC-based accumulation assay as a powerful approach to investigate the interaction of a whole animal with a constellation of chemical structures in its environment. This method was used to survey the accumulation of over 1,000 small drug-like chemicals in C. elegans, and machine learning was applied to model the features found in accumulating structures. As a result, we now have a better understanding of the chemical space occupied by drug-like compounds that accumulate in C. elegans and I have devised methods to circumvent its xenobiotic defenses.
2.4 Materials and Methods

2.4.1 Worm Culture

Wild-type (N2) worms were used for all experiments, and were originally obtained from the C. elegans Genetics Center (U. Minnesota). Worms were cultured at 20°C using standard techniques (Lewis and Fleming, 1995) unless otherwise indicated.

2.4.2 HPLC-Based Small Molecule Accumulation Assay

Late-stage fourth-larval-stage worms, grown from synchronized hatchlings at 25 °C for 45 h on NA22 Escherichia coli, were used for the accumulation assay. The worms were harvested, washed at least twice and re-suspended in enough M9 buffer (Burns et al., 2006) for a final concentration of ~10 worms per μl. Five hundred microliters of this worm suspension was added to each well of Pall AcropPrep 96-well filter plates (0.45-μm GHP membrane, 1-ml well volume). Chemicals were added to each well to a final concentration of 40 μM (0.4% DMSO, v/v). Worms were incubated in the small-molecule solutions at 20°C for 6 h with aeration, after which the incubation buffer was drained from the wells by vacuum (6 h is the longest time allowed before the filter membranes weaken). The worms were then washed three times with 500 μl of M9 buffer. After washing, the worms were resuspended in 50 μl of M9 buffer, transferred to new 96-well solid-bottom plates and stored frozen at −20 °C. The samples were later lysed by adding 50 μl of a 2× lysis solution (100 mM KCl, 20 mM Tris, pH 8.3, 0.4% SDS, 120 μg ml−1 proteinase K) to each well and incubating the plates at 60 °C for 1 h with agitation. After lysis, the plates were stored frozen at −80 °C for later processing by HPLC.

For HPLC analysis, worm lysate plates were thawed and 100 μl of cold acetonitrile (ACN) was added to each lysate. The samples were mixed by pipetting. The plates were centrifuged
at 796g for 5 minutes. After centrifugation, 75 μl of the lysate was injected onto a 4.6 X 150 mm Zorbax SB-C8 column (5 micron particle size) and eluted with the following solvent and flow rate gradients over 5.2 minutes:

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Solvent</th>
<th>Flow Rate (ml/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>Solvent A: 85%</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Solvent B: 15%</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>Solvent A: 85%</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Solvent B: 15%</td>
<td></td>
</tr>
<tr>
<td>3.20</td>
<td>Solvent A: 30%</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Solvent B: 70%</td>
<td></td>
</tr>
<tr>
<td>4.25</td>
<td>Solvent B: 100%</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Solvent A: 4.9:95:0.1 (ACN/H2O/TFA)  
Solvent B: 95:4.9:0.1 (ACN/H2O/TFA)

A column temperature of 27°C was used. UV-Vis absorbance was measured every 2 nm between 190 and 602 nm. Absorbance intensity data was converted to a three dimensional heat-mapped chromatogram using MATLAB (The MathWorks). All HPLC was performed using an HP 1050 system equipped with an autosampler, vacuum degasser, column heater, and diode-array detector. HP Chemstation software was used for data acquisition.

2.4.3 Survey of Small Molecule Accumulation (Additional Information)

Before testing their accumulation in worms, a standard 7.5 nmol amount of each small molecule was processed by HPLC to determine if they are detectable by the HPLC-DAD system. Molecules that are visible in the chromatogram but have a column retention time <1.25 minutes were not assayed for accumulation in worms since these compounds would be masked by the peak of worm contents (green arrows in Fig. 2-1A).
The majority of the Spectrum compounds tested for accumulation were assayed in worms three times, except for 13% which were assayed twice, and 3.6% that were assayed only once due to technical issues (these compounds were eliminated from further consideration). All of the DHP analogs and the Fenbufen analogs were assayed in triplicate. The majority of the nematicides tested for accumulation were assayed three times, except for 5 compounds that were assayed twice. Only two of the LOPAC compounds (see Fig. 2-3E) were assayed three times, the remainder were assayed twice. A small-molecule was considered to accumulate in worms if the parent structure, or a metabolite of the parent structure, were visible in the chromatograms of at least 2 replicate lysates after incubation. The parent structure was assumed to accumulate if the peak detected in the worm lysates was within ± 6 seconds (± 2% of the run) of the retention time of the standard peak, and had an invariable absorbance spectrum. Any detectable peak with a retention time that was not within ± 6 seconds of the standard peak, but with an invariable, or very similar, absorbance spectrum compared with the standard was assumed to be a metabolic derivative of the parent structure. The majority of metabolites eluted at dramatically different times than the parent molecule (i.e. 85% of metabolites eluted more than 30 seconds faster or slower than the parent compound).

### 2.4.4 No Worm Sham Trials

To control for false positives due to compounds precipitating in the well during incubation, molecules scored as accumulating were reprocessed one time using the accumulation assay without worms. All molecules that were detectable in the absence of worms were excluded from the dataset as false positives.
2.4.5 SDS Wash Controls

To control for false positives due to compounds sticking to the worm cuticle, molecules scored as accumulating were reprocessed twice using the accumulation assay, and worms were washed with dilute detergent after incubation. The washes consisted of two washes with 500 µL of a 0.1% SDS solution and a third wash with 500 µL of M9 buffer. Structures that were not detectable in both worm lysates after SDS washes were scored as non-accumulating. For technical reasons, SDS washes of 10 of the 144 accumulating molecules were not possible – either the wells would not clear during the wash steps or sufficient quantities of the compound were not available to perform the experiment. These 10 molecules were scored as accumulating.

2.4.6 Dead Worm Controls

75 out of 76 compounds that accumulated in worms as putative metabolites were reprocessed once using the accumulation assay with heat-killed worms. This control was to ensure that the putative metabolites are live-worm dependent and not simply accumulated contaminants or breakdown products of the parent molecule. Synchronized late-stage L4 worms, grown at 25°C, were used for these experiments, and were killed by way of an overnight incubation at 37°C without aeration. After incubating the dead worms in the desired compounds, the worms were washed, collected, and lysed chemically as previously described in section 2.4.2. The samples were stored frozen at -80°C for later processing by HPLC. Any putative metabolite that was detected in the dead worm sample was scored as a false positive for metabolism. The 1 compound for which dead worm controls were not performed (for technical reasons) was assumed to be a live worm-dependent metabolite. The compounds that accumulate as
metabolites in the complete small-molecule accumulation dataset (Fig. 2-3) were determined to be true positives by this method.

A second dead worm control, in which the buffer and the dead worms were collected after incubation, was performed to account for live worm-independent breakdown and/or modified structures that fail to accumulate in dead worm tissue. 68 out of the 75 previously tested compounds were reprocessed three times using the accumulation assay with heat-killed worms. After incubating the dead worms in the desired compounds, the worms and the incubation buffer were collected and stored frozen at -20°C. The samples were later dried using a Savant DNA120 SpeedVac, resuspended in 50 µl of water, and lysed chemically as previously described in Materials and Methods 2.4.2. The samples were stored frozen at -80°C for later processing by HPLC. Any putative metabolite that was detected in at least 1 dead worm sample was scored as a false positive for metabolism. All of the false positives from the first dead worm control experiments were recovered by the second method, and 10 metabolites from 10 new compounds were found to be false positives. However, three of these metabolites were found to be false positives for accumulation by SDS wash control, and only two of the remaining metabolites are the major accumulating structures in worms; therefore the false positives discovered by the second approach do not substantially impact my analysis.

2.4.7 Determining the Volume of Worms in the Sample Lysates

To estimate the number of worms in the sample lysates, known amounts of L4 worms (1000, 2000, 4000, 6000, 10000) were lysed and processed by the HPLC method, and a linear standard curve was generated using the area under the peak of endogenous worm contents (green
arrows in Fig. 2-1A) of the lysates. Chemstation software was used to calculate peak areas at 310 nm. The endogenous peak areas for all sample lysates were then calculated, at 310 nm, and the number of worms in each sample lysate was determined using the standard curve. All worm lysates in the complete small-molecule accumulation dataset have endogenous peak areas that fall within the linear range of the standard curve. The volume of worms in the sample lysates were then calculated assuming that one L4 worm has a volume equal to 3 nl (Watanabe et al., 2005).

2.4.8 Estimating Minimal Detectable Concentration of a Small Molecule in a Lysate

The minimum detectable absorbance intensity for the heat-mapped chromatograms was estimated by conducting a two-fold dilution series of 19 random molecules from the complete Spectrum dataset, until a peak was no longer detectable in the chromatogram. The minimum detectable absorbance intensity should be relatively constant for all molecules in the dataset, barring variability due to peak width. For all 19 molecules, peak heights (at $\lambda_{\text{max}}$) were calculated for the smallest amount with a detectable peak and for the largest amount with an undetectable peak. Peak height is a measure of the absorbance intensity of a peak at its apex. $\lambda_{\text{max}}$ is the wavelength at which maximum absorbance occurs; here, $\lambda_{\text{max}}$ refers to a wavelength $\geq 230\text{nm}$ at which maximum absorbance occurs, since the solvent absorbs light between 190 and 210 nm at all retention times. Calculating the height of a peak that is not detectable in the heat-mapped chromatogram is possible because, in all cases, a peak was visible at $\lambda_{\text{max}}$ in the raw Chemstation plot. Next, peak height cutoffs from 5.0 to 16.0 mAU were used to parse the 38 peak heights derived from the 19 molecules into detectable and
undetectable categories. A peak height cutoff of 8.6 mAU was found to be the most accurate at parsing the data, and was adopted as the minimum detectable absorbance intensity for the heat-mapped chromatograms.

The minimum detectable amount of a molecule was determined using the following equation:

\[
\text{minimum detectable amount} = \frac{(8.6 \text{ mAU})}{(\text{peak height/nmol})}
\]

The 8.6 mAU value is the minimum detectable absorbance intensity, as defined above. The peak height/nmol value is known for each of the 1,132 molecules assayed because a standard 7.5 nmol amount was processed for every compound and peak height was calculated at \( \lambda_{\text{max}} \). This approach assumes that peak height behaves in a linear fashion with the amount of molecule, from the minimum detectable amount up to 7.5 nmol. This is a valid assumption, since the peak heights of 20 out of 22 randomly chosen molecules from the complete Spectrum dataset are linear with the amount of molecule from below the minimum detectable amount up to 7.5 nmol or higher (fitted lines with R-squared values \( \geq 0.95 \) were considered linear).

To determine the minimum detectable concentration of a small-molecule in a given lysate, the minimum detectable amount of the small-molecule was first multiplied by \( \frac{8}{3} \) (because only \( \frac{3}{8} \) of the lysate is injected over the column) and then divided by the volume of worms in the lysate.

### 2.4.9 Quantifying Small Molecule Concentration in Worms

Small-molecule levels in the injected lysates were quantified using peak areas at \( \lambda_{\text{max}} \) and converted to absolute values using the following equation:
small-molecule level in injected lysate = (peak area in injected lysate) ÷ (peak area/nmol)

The peak area/nmol value is known for each of the 1,132 molecules assayed because a standard 7.5 nmol amount was processed for every compound and peak area was calculated at λmax. It is assumed that peak area behaves in a linear fashion with compound levels, from the amount in the injected lysate up to 7.5 nmol. This is a valid assumption, since the peak areas of 21 out of 22 randomly chosen molecules from the complete Spectrum dataset are linear with the amount of molecule, from below the minimum detectable amount up to 7.5 nmol or higher (fitted lines with R-squared values ≥ 0.95 were considered linear). For the small-molecule metabolites it was assumed that UV-Vis absorbance varies with concentration in a manner identical to their respective parent structures. Due to high background intensities, metabolite levels could not be quantified for structures that elute in the flow-through and overlap with the peak of endogenous worm contents (green arrows in Fig. 2-1A).

To determine the small-molecule concentration in a worm lysate, the amount of small-molecule in the injected lysate was first multiplied by 8/3 (because only 3/8 of the lysate is injected over the column) and then divided by the volume of worms in the lysate. An average small-molecule concentration in worms was calculated by averaging the lysate concentrations for a given molecule.

2.4.10 Deriving a Cutoff to Eliminate the Detection Limit Bias

For this analysis, all of the 2,941 replicate lysates of the complete Spectrum dataset were considered individually. All lysates associated with molecules that accumulate as unquantifiable metabolites, due to proximity to the endogenous peak, were eliminated from the analysis. The remaining 2,891 lysates were binned by the minimum detectable
concentration (MDC) of the small-molecule in the lysate. The bins were defined so that every bin has an approximately equal number of lysates. For each bin, an enrichment ratio was calculated as follows:

\[
\log_2 \left( \frac{(\text{# of lysates with detectable molecules in this bin})}{(\text{# of lysates in this bin})} \right) \times \frac{(\text{# of lysates with detectable molecules in the whole dataset})}{(\text{total # of lysates in the whole dataset})}
\]

This ratio was plotted for each bin. The lowest MDC bin (< 6 µM) is enriched for lysates with detectable molecules. The hypergeometric p-value for this enrichment is < 10^{-10}. I define this p-value as the enrichment probability for that bin, i.e. the probability of obtaining such an enrichment of lysates with detectable molecules by chance. Significant enrichment was defined as p < 0.01. The highest MDC bin (> 96 µM) is enriched for lysates without detectable molecules. Thus, there is a detection limit bias in the dataset.

Concentration cutoff values between 10 µM and 25 µM were established to eliminate the bias in the dataset. These cutoffs define the highest MDC of a sample lysate allowed for inclusion in the dataset, and the lowest quantified concentration (QC) of a molecule considered detectable. If, as a result of the cutoff, the enrichment probability for the lowest MDC bin is ≥ 0.01, that cutoff is defined as having eliminated the bias. The enrichment probability for the lowest bin was used because it is the only bin that has a constant number of sample lysates at all cutoffs. I found that all cutoffs between 10 µM and 25 µM eliminate the bias. This is likely because the bias can be eliminated by either low MDC cutoffs, or by high QC cutoffs. A cutoff dominated by low MDC would eliminate many false negatives from the dataset, but many true negatives as well, resulting in an erroneously high estimate of small-molecule accumulation. A cutoff dominated by high QC would retain many false negatives in the dataset, and eliminate many true positives, resulting in an erroneously low estimate of
small molecule accumulation. So, in order to obtain an accurate estimate of the rate of small-molecule accumulation in the worm, and to have a high-confidence dataset for machine learning, it was important to derive a cutoff that is not dominated by either MDC or QC.

To establish a proper concentration cutoff value, a series of MDC cutoffs and QC cutoffs from 10 µM to 25 µM were applied to the dataset, respectively. The MDC cutoffs were applied to the dataset so that lysates having a MDC > cutoff were eliminated from the dataset at that cutoff, but all lysates having a detectable molecule with a QC > 0 were considered “detectable” at all cutoffs. The QC cutoffs were applied in such a way that all lysates having a QC < cutoff were considered “not detectable” at that cutoff, but all lysates were included in the dataset at every cutoff. A log₂ ratio of the MDC and QC enrichment probabilities was plotted for each cutoff value. The cutoff value that had the smallest difference in enrichment probabilities between MDC and QC was chosen. The rationale is that at a given cutoff value, if MDC does not affect the enrichment probability any more or less than QC, the cutoff is not dominated by either MDC or QC. A 19 µM cutoff value satisfied this criterion, and it is the only cutoff value at which both MDC and QC cutoffs do not eliminate the bias on their own. The 19 µM concentration cutoff value eliminates the bias in the dataset, and the enrichment probability for the lowest bin is no longer significant (p > 0.09).

2.4.11 Generating the Unbiased Datasets

To generate the unbiased Spectrum dataset, the 19 µM cutoff was applied to the complete Spectrum dataset. All of the molecules in the dataset were considered individually (not the sample lysates). Before applying the cutoff, all molecules that accumulate only as metabolites that are not quantifiable, due to proximity to the peak of worm contents, were excluded from
the dataset. The remaining molecules in the dataset were analyzed further by applying the 19 
µM cutoff using the following steps:

1. A molecule was included in the dataset only if it has 2 or more replicate lysates with 
MDCs ≤ 19 µM.
2. A small-molecule was considered to accumulate only if the parent structure, or a 
metabolite of the parent structure, has a QC ≥ 19 µM, in at least 2 out of 2 or 2 out of 3 
replicate lysates.
3. All small-molecules that are measurable at a concentration ≥ 19 µM in only 1 out of 2 
replicate lysates were excluded from the dataset.
4. After steps 1-3, any non-accumulating small-molecule that has unquantifiable 
metabolites detectable in 1 out of 2, 2 out of 2, 2 out of 3, or 3 out of 3 replicate lysates 
were excluded from the dataset.

To generate the training set for machine learning, the same procedure used to generate the 
unbiased Spectrum dataset was used, except for two differences:

1. The cutoff was applied to the complete small-molecule accumulation dataset instead of 
the complete Spectrum dataset.
2. All molecules that are detectable as either a parent structure or a metabolite in at least 
2 out of 2, or 2 out of 3 replicate lysates at a concentration ≥ 19 µM were included in the 
dataset, regardless of whether the lysates have MDCs that are ≤ 19 µM, or not.

2.4.12 Analysis of Metabolites by LC-MS

Metabolites were HPLC-purified from worm lysates and dried using a Savant DNA120 
SpeedVac (acid was not added to the HPLC solvents). Chromatographic separations of the
purified metabolites for LC-MS were performed using a nano-AQUITY Ultra Performance Liquid Chromatography (UPLC) system (Waters Corp.). Separation was carried out in a 75 um × 200 mm BEH column, packed with 1.7 μm C18 particles. The flow rate was set at 0.3 μL/min with the following solvent gradient:

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Solvent</th>
<th>Flow Rate (μL/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>Solvent A:100%</td>
<td>0.3</td>
</tr>
<tr>
<td>2.00</td>
<td>Solvent A:100%</td>
<td>0.3</td>
</tr>
<tr>
<td>3.00</td>
<td>Solvent A: 97% Solvent B: 3%</td>
<td>0.3</td>
</tr>
<tr>
<td>10.00</td>
<td>Solvent A: 60% Solvent B: 40%</td>
<td>0.3</td>
</tr>
<tr>
<td>40.00</td>
<td>Solvent A: 10% Solvent B: 90%</td>
<td>0.3</td>
</tr>
<tr>
<td>45.00</td>
<td>Solvent A: 10% Solvent B: 90%</td>
<td>0.3</td>
</tr>
<tr>
<td>50.00</td>
<td>Solvent A: 97% Solvent B: 3%</td>
<td>0.3</td>
</tr>
<tr>
<td>51.00</td>
<td>Solvent A:100%</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Solvent A: 0.2% Formic Acid / H2O
Solvent B: 0.2% Formic Acid / ACN

Mass spectrometry was performed using a Micromass Quadrupole-Time-of-Flight Premiere instrument (Waters Corp.). The data acquisition software used was MassLynx NT, version 4.0. Mass spectra were acquired in positive ion mode using a nano-ESI with capillary voltage and sample cone voltage set to 3,000 V and 20 V, respectively. The MS acquisition rate was set to 1.0 s, with a 0.1-s interscan delay. Ninety-eight percent argon gas was employed as the collision gas with collision energy varying from 13–46 V for the mass range of 100–1,000 m/z. Ions selected for LC-MS-MS were identified after manual analysis of original LC-MS runs, and a corresponding inclusion list was generated for targeted data-dependent acquisition experiments.
2.4.13 Metabolite Accurate Mass Measurements

Metabolites selected for accurate mass measurements were HPLC-purified from worm lysates and dissolved in a 1:1 mixture of methanol and water with 0.1% formic acid added. Approximately 1 μL of this solution was injected via a loop into a liquid chromatography (LC) flow (Agilent 1100 capillary LC) of 1:1 MeOH/H2O (0.1% formic acid), flowing at 100μL/min into the ESI source of a Sciex QStar mass spectrometer, operating in the time-of-flight mode at a resolution of ~8,000. Spectra were collected every second over the m/z range 90 to 1,000u and an external calibrant was used to calibrate the instrument over the mass range.

2.4.14 β-glucosidase Digestions

Putative O-linked β-glucosides were HPLC-purified from worm lysates and digested with 5 units/ml almond β-glucosidase (Sigma) in 80mM sodium acetate buffer (pH 5.0) at 37°C for 12 hours. To control for compound stability and spontaneous hydrolysis, equivalent amounts of purified metabolites were incubated under identical conditions, but without β-glucosidase.

2.4.15 Cheminformatics and Machine Learning

The cheminformatic package in Pipeline Pilot version 6.1 (Scitegic Inc. Accelrys) was used to standardize the representation of all compounds studied, including removing inorganic compounds, salts and duplicates. Pipeline Pilot was also used for all Naive Bayes statistical model building. The Naive Bayesian structure-based accumulation model (SAM) was built using the Extended Connectivity Fingerprints (ECFP_4) method (Rogers et al., 2005), in which the compounds are represented by overlapping fragments of a diameter of up to four bond
lengths. The SAM was validated using a five-fold cross validation procedure, where four-fifths of the data are used to train the model and the remaining one-fifth are used to test the model. This procedure is run five times, and each compound appears in the test set once and the training set four times. The accuracy of the SAM was measured using the enrichment rate. The enrichment rate is calculated by ranking all compounds in the test set using the model, and then comparing the number of actives found in the top \( n \% \) to the number of expected actives for \( n \% \). The final model was built using all of the compounds in the training set. When ranking other datasets with the SAM, compounds were filtered from the dataset before ranking if they had >85% structural similarity, using the Tanimoto score with ECFP_4 fingerprints (see below), to any compound in the training set used to build the model.

All similarity calculations were carried out using the Tanimoto coefficient with the ECFP_4 fingerprinting method. The Tanimoto coefficient is the number of fragments in common between two compounds, divided by the total number of fragments present in both compounds. Murcko scaffolds (Bemis and Murcko, 1996) were generated using the ‘Generate Fragments’ protocol in Pipeline Pilot.

### 2.4.16 Building a Model of Small Molecule UV-Vis Absorbance

The UV-Vis absorbance model was built using Naïve Bayesian statistical modeling and ECFP_4 fingerprinting. The model was trained on HPLC-DAD absorbance data obtained for 1,770 compounds. A molecule was considered to absorb UV-Vis light if it would have a minimum detectable concentration of \( \leq 19 \, \mu M \) in a worm lysate processed by the HT-HPLC method described above, assuming the lysate was obtained from 5000 worms at the fourth larval stage. The same 19 \( \mu M \) cutoff value was used to generate the “unbiased Spectrum dataset”
and the training set for machine-learning. The UV-Vis absorbance model was used to rank the 50K DIVERSet library before biological validation of the SAM, to increase the probability that the test compounds could be detected by our HPLC-DAD method, and the top 10% were selected for testing.

2.4.17 Determining the Structural Features that Influence Accumulation

There is redundancy in the top-scoring and bottom-scoring 30 fragments learned by the structure-based accumulation model, since the ECFP_4 fingerprinting method captures all fragments that are 0, 2, or 4 bond lengths from each heavy atom. Also, some fragments are likely present in the top and bottom-scoring sets only because they are found in compounds alongside other more prominent features. The top 5 and bottom 5 non-overlapping and structurally non-redundant fragments (Fig. 2-8A,B) were compiled by considering the most abundant fragment first, eliminating all other fragments that overlap, considering the next most abundant fragment, and so on. If the majority (>70%) of the compounds with a given fragment contain one or more other more prominent fragments then that fragment was considered redundant. The top 5 most prominent non-redundant and non-overlapping fragments were compiled in this way. There is some residual redundancy in the top and bottom-scoring 5 fragments, but this approach minimizes the overlap to provide relatively independent features.

2.4.18 Fenbufen Analog Phenotypic Analysis

For the Fenbufen analog lethality assays, two L4 worms were seeded in the wells of 24-well plates containing Fenbufen analogs at the desired concentrations or DMSO for control.
5 days after depositing the L4 parents, each well was photographed each day with a Leica MZ7.5 microscope and a PL-A661 PixeLINK camera. The number of viable adults in the wells was counted from the pictures. Each Fenbufen analog experiment was done in triplicate. DMSO control experiments were done nine times.
Chapter 3

General Discussion

Chapter 3. General Discussion

3.1 Summary

In the preceding chapter, I described my systematic investigation of the accumulation and metabolism of hundreds of small drug-like molecules in the tissues of the nematode *C. elegans*. This was accomplished using a high-throughput HPLC assay that I developed. I discovered that the majority of molecules fail to accumulate in worms, despite most of them having properties that promote oral bioavailability in humans. In addition, I found that accumulation is generally required for bioactivity. These results suggest that only a small fraction of “human orally bioavailable” chemical space is bioavailable to worms and has access to worm target proteins during screening (Fig. 3-1).

Given that most purchasable chemical libraries are designed to be orally active in humans, I reasoned that higher hit rates could be achieved for *C. elegans* chemical screens if the molecules were selected from a “worm-bioavailable” subset of this chemical space (grey area in Fig. 3-1). To define a worm-focused chemistry, my drug-like molecule accumulation data was used as a training set to model the structural signatures of accumulating compounds using computer-based chemical fingerprinting and machine-learning techniques. The resultant structure-based accumulation model (SAM) scores substructural features positively if they have a high likelihood of being found in accumulating molecules, whereas features unlikely to be found in accumulating compounds are scored negatively. A key utility of the SAM is that it can be used to generate accumulation scores for any molecule (or set of molecules) of interest, and defines high-scoring molecules as more likely to be worm-bioavailable. When applied to
Figure 3-1. Theoretical Venn diagrams of human- and worm-bioavailable chemical spaces. 
(A) One possibility is that human- and worm-bioavailable chemical spaces are comparable in size, and that there is only a small fraction of overlap between the two. (B) Another possibility is that worm-bioavailable chemical space is much smaller than human-bioavailable chemical space, and it is completely (or almost completely) contained within human-bioavailable chemical space. In both (A) and (B), the grey area represents the chemical space that is bioavailable to both worms and humans. The structure-based accumulation model scores molecules from the grey area as being more likely to accumulate in worms, relative to the molecules found in the white area. The black area of chemical space (if it exists) was not learned by my model. The Venn diagrams are not to scale.
naive chemical libraries, the SAM enriches for compounds that accumulate and that are bioactive in whole worms.

To my knowledge, this work represents the first description of a worm-bioavailable chemical space, and will hopefully provide a framework for the chemical matter used in future *C. elegans* chemical-genetic experiments. Given the enormous size of purchasable chemical space (> 13 million compounds), the SAM will likely serve as a valuable tool to select and prioritize the molecules that are included in future *C. elegans* screening ventures. For the benefit of the chemical biology and drug discovery communities, the SAM, and the accumulation data used to generate it, have been made publicly available.

### 3.2 Implications of a Worm-Bioavailable Chemical Space

#### 3.2.1 Worm-Focused Chemical Libraries

Currently, the Prestwick *C. elegans* library (Prestwick Chemical) is the only chemical library marketed towards screens in worms. It is a relatively small library of only 240 small molecules, the majority of which are FDA approved drugs that are non-toxic to *C. elegans*. Since these compounds are FDA approved drugs they likely have good oral bioavailability in humans; however, as I have established, human oral bioavailability does not guarantee worm bioavailability. Though the Prestwick library may have some utility for *C. elegans* screens, its small size, lack of novel chemical matter, and ignorance of worm bioavailability properties highlight the need for an improved commercially available worm-focused library of small molecules.
A number of chemical vendors have collections of compounds whose structures are biased towards different subsets of chemical space. For example, ChemBridge Corp. (www.chembridge.com) has a number of “focused” chemical libraries such as their kinase-biased and NHR-biased collections. Our lab is currently working with ChemBridge to create the first purchasable “worm-focused” chemical library designed specifically for screens in C. elegans. The SAM is being used to rank one million unique ChemBridge structures, and the top 5% of these molecules (50,000 compounds) will be included in the library. This collection of worm-bioavailable small molecules will likely improve the efficiency by which biologically active compounds are identified from C. elegans chemical screens, and will be a critical tool in the discovery of novel reagents to probe the biology of animals.

3.2.2 Worm-Bioavailable Structural Analogs for SAR Analyses

A useful first step in the characterization of any biologically active compound identified from a chemical screen is the examination of structurally similar derivatives of the initial “hit” compound for structures with altered potencies, with the aim of relating structure to activity. This type of structure-activity relationship (SAR) can identify structural analogs that have improved potency and selectivity relative to the parent compound, as well as highlight the structural features that are required or dispensable for the observed bioactivity. For example, medicinal chemists use SARs to inform the modifications they make to initial “hit” compounds, so as to achieve greater potency and selectivity in vitro and in vivo, while accommodating metabolic stability, bioavailability, and tolerability in vivo (Lombardino and Lowe, 2004).

In vivo SAR analysis of a biologically active molecule discovered through a phenotype-based screen can inform biochemical approaches to target identification. For example,
knowing the features that are dispensable for activity facilitates the construction of affinity-tagged versions of the molecule that will not disrupt target protein binding during affinity purification. The information obtained from an in vivo SAR analysis depends on whether or not the structural analogs used are bioavailable to the organism and have access to the target protein. If the majority of the analogs used for the analysis are bioavailable, then the SARs will provide information about the molecular features that are dispensable for target protein binding. Conversely, if the majority of the analogs are not bioavailable to the organism, then the SARs identified will relate primarily to bioavailability and will provide little insight regarding the features that are dispensable for binding. Therefore, the capacity to select bioavailable structural analogs of initial “hit” compounds is of critical importance for in vivo SAR analyses.

Given the tendency of C. elegans to resist the accumulation of exogenous drug-like molecules in its tissues, it is of particular importance to carefully select bioavailable analogs for in vivo SAR analyses with this organism, and to measure the accumulation of the analogs tested. For example, our lab previously analysed twelve structural analogs of the 1,4-dihydropyridine (DHP) family of L-type calcium channel antagonists and found that only three were bioactive in whole worms (Kwok et al., 2006). Using my HT-HPLC assay, I subsequently tested the accumulation of these DHP analogs in worms and found that only the three bioactive structures, and one inactive structure, accumulate (Burns et al., 2010). Therefore, our DHP SAR analysis provided limited information regarding the features that are required for target binding, since the majority of compounds could not access the target. My exploration of worm-bioavailable chemical space has revealed that compounds are more likely to accumulate in worms if they are relatively hydrophobic (i.e. clogP greater than roughly 4). Consistent with this finding, the three bioactive DHP analogs have clogP values larger than 4, whereas the
majority of the inactive analogs have clogP values that are less than 4. Thus, in vivo SAR analyses with C. elegans would benefit from the exclusion of structural analogs that have clogP values below 4, by enriching the SAR space with worm-bioavailable molecules, as well as the confirmation of their accumulation using my HT-HPLC assay.

3.2.3 Other Organism-Specific Chemical Spaces

The bioavailability spaces of small molecules are almost certainly species-specific. A comparison of the physicochemical properties between human orally bioavailable drugs and antibiotics against four distinct pathogenic species of bacteria reveals that these molecules occupy organism-dependent areas of chemical space (Hopkins and Bickerton, 2010). Furthermore, there are differences in the absorption and metabolism of drugs between humans and rats (Lombardino and Lowe, 2004), two species that are relatively close on an evolutionary time scale. Drug pharmacokinetics can even vary between two individuals of the same species (Willmann et al., 2009).

Purchasable chemical libraries are predominantly designed to contain molecules with properties that favour human oral bioavailability. This approach to library design likely does not optimize probe discovery rates in non-human organisms. Indeed, I have shown that selecting molecules from a worm-bioavailable chemical space improves the efficiency by which bioactive compounds are identified from chemical screens in worms. Bioactive discovery rates in other organisms that are amenable to high-throughput chemical screens would likely also benefit from the compilation of species-specific collections of small molecules. For microorganisms such as bacteria or yeast, the modeling of small molecule bioavailability and/or bioactivity spaces would be relatively straightforward. Indeed, subsequent to the
publication of my work, Wallace et al. (2011) constructed a machine-learned model of yeast-bioactive chemical space for the budding yeast *Saccharomyces cerevisiae* (Wallace et al., 2011). The authors ranked the 2,000 compounds of the Spectrum library (Microsource Inc.) and obtained a ~3.5-fold enrichment rate of yeast bioactives in the top-scoring 10% relative to a random-scoring 10%.

Selecting molecules from organism-specific chemical spaces can increase bioactive discovery rates, saving time and money for academic chemical screening ventures. The successful application of worm-bioavailable and yeast-bioactive chemical spaces provides evidence for this assertion, and could signal a paradigm shift in the way chemical libraries are constructed for model organism chemical-genetic screens.

### 3.3 Anthelmintic Discovery in Worm-Bioavailable Space

Parasitic nematodes infect over 1 billion people worldwide (Keiser and Utzinger, 2010), and impose significant agricultural losses through infected livestock and crops (Besier, 2007; Fuller et al., 2008). Resistance to commonly used anthelmintics among animals is escalating and the recent increase in the use of mass drug administration for the treatment of human helminth infections raises the risk that clinically relevant resistance will soon develop (Besier, 2007; Sutherland and Leathwick, 2011). Despite the obvious need for new anthelmintic drugs, the last 30 years has seen only three new anthelmintic classes marketed for veterinary use (monepantel and derquantel for sheep (Kaminsky et al., 2008; Kaminsky et al., 2009; Little et al., 2010), and emodepside for cats and dogs (Keiser and Utzinger, 2010)) and only one new drug approved for use in humans (tribendimidine in China (Hu et al., 2009; Keiser and Utzinger, 2010)). Clearly, there is a pressing need to discover new anthelmintic drug classes.
The relatively large size and complex life cycles of many parasitic nematode species present significant challenges for their manipulation in a laboratory setting, particularly when considering their use in high-throughput chemical screens. *C. elegans* provides an experimentally tractable alternative to parasitic worms for the discovery and characterization of novel anthelmintic drugs, and enables the high-throughput screening of small molecules for bioactivity in the context of the whole animal over its entire life cycle. Recent technological advances in liquid handling (Byrne et al., 2007), worm sorting (Pulak, 2006), and automated imaging (Buckingham and Sattelle, 2008, 2009; Burns et al., 2006; Gosai et al., 2010) facilitate a screening throughput in *C. elegans* that is high even by industrial standards (i.e. tens of thousands of chemicals per day (Gosai et al., 2010)), transforming *C. elegans* into a viable platform for industrial anthelmintic discovery.

Several lines of evidence support the idea that *C. elegans* is a suitable model for anthelmintic drug discovery. (i) Comparative genomic analyses of *C. elegans* and parasitic nematodes suggest that *C. elegans* is no more distinct from any parasitic nematode than any two parasitic nematodes are from each other (Holden-Dye and Walker, 2007; Mitreva et al., 2005). (ii) The majority of anthelmintic drugs used to treat parasitic worm infections of humans and livestock are effective against *C. elegans*, and most have conserved modes of action (Holden-Dye and Walker, 2007; Kaminsky et al., 2008). (iii) Both established and putative anthelmintic drug targets are conserved in *C. elegans* (Hetherington et al., 2011; Holden-Dye and Walker, 2007; McVeigh et al., 2005; Ogawa et al., 2009; Stepek et al., 2010; Wang et al., 2009). (iv) It has recently been shown that the long-lived nonreproductive *C. elegans* dauer stage is analogous to the infective larval stage (iL3) of parasitic nematodes, and that a conserved endocrine signaling mechanism controls entry into and recovery from both dauer
and iL3 (Ogawa et al., 2009; Wang et al., 2009). It therefore follows that anthelmintics developed against *C. elegans* might be equally effective against parasitic nematodes based on the genomic and biological relationships between the nematodes.

Encouragingly, the molecular properties that influence bioavailability in *C. elegans* are similar to those for parasitic worms, suggesting that my structure-based accumulation model may have some utility in predicting drug bioavailability for parasitic worm species as well. For example, lipophilicity is a major determinant of drug entry into parasitic worms (Alvarez et al., 2007) and I find that molecules that accumulate in *C. elegans* are generally more hydrophobic than those that fail to accumulate. In addition, two of the lowest scoring features defined by our SAM are sulfoxide and sulfonyl groups, and these features have been shown to reduce anthelmintic bioavailability in parasitic worms (Alvarez et al., 2007) and references therein. For instance, the sulfoxide metabolite of albendazole has poorer bioavailability in *Ascaris suum*, *Moniezia spp.*, and *Fasciola hepatica* compared to the unmodified structure. Furthermore, *F. hepatica* can sulfoxidate and sulfonate the anthelmintic triclabendazole (TCBZ), and it has been shown that the rate of TCBZ sulfoxidation is significantly higher in TCBZ-resistant isolates of *F. hepatica* relative to susceptible worms, suggesting that sulfoxidation plays a role in TCBZ detoxification. Of the 34 compounds in our accumulation dataset that contain a sulfonyl or sulfoxide group, zero accumulate in *C. elegans*.

Despite these similarities, there are documented differences in the pharmacokinetics of established anthelmintics between *C. elegans* and parasitic worms. For example, amidantel and derquantel are potent anthelmintics in animal models of nematode infection, but have minimum effective concentrations in *C. elegans* at or above 350 and 100 μM, respectively (Ruiz-Lancheros et al., 2011; Tomlinson et al., 1985). Nevertheless, as most established
anthelmintics are effective against *C. elegans*, the “bioavailability spaces” of parasitic worms and *C. elegans* certainly overlap, even if they are not identical.

### 3.4 *C. elegans* Drug Delivery

The worm-bioavailable chemical space I have described will undoubtedly improve the efficiency by which small molecule probes are identified from *C. elegans* chemical screens. However, it has been shown that distinct target proteins bind small molecule modulators that occupy distinct areas of chemical space (Paolini et al., 2006). Thus, employing worm-bioavailable compounds exclusively for screens in *C. elegans* may bias the hits obtained towards a select group of protein targets, at the expense of others. For example, the majority of drugs that bind and modulate the activity of human metalloproteases and serine proteases have clogP values that place them in a chemical space unlikely to accumulate in worms (Burns et al., 2010; Paolini et al., 2006). Therefore, to complement my *C. elegans* structure-based accumulation model, there is a need to develop and apply methods that sensitize worms to a broader chemical space. In this section I will describe a number of approaches that could be used to achieve this goal. See Fig. 3-2 for a summary of these approaches in the context of a high-throughput *C. elegans* chemical screening platform.

#### 3.4.1 Modify the Screening Media

*C. elegans* can be cultured in both liquid media and on solid substrate, and chemical screens performed in both media types have been described (Kwok et al., 2006; Petrascheck et al., 2007). I have observed relatively higher hit rates for liquid-based versus solid-based screens, likely because the worms are bathing in the chemical solution, which provides greater
Figure 3-2. Methods to maximize bioavailability in the context of *C. elegans* high-throughput chemical screening. The workflow for a high-throughput chemical screen in *C. elegans* is shown. Liquid handling, worm sorting, and imaging can be fully automated with the use of robots and automated imaging platforms. The worm strains, media, and chemicals used for screening can be optimized to maximize the bioavailability of the compounds being screened (see text for details). SMoC - small molecule carrier.
opportunity for the chemicals to penetrate the animals and access target proteins. Aside from bioavailability considerations, liquid-based screens are more amenable to miniaturization and high-throughput screening using, for example, 384-well microtiter plates and microfluidic chips. Thus, for maximal throughput and bioavailability, liquid media is likely the best choice for *C. elegans* chemical screens. The simplest method to achieve higher internal xenobiotic concentrations in the worm is to increase the compound’s external concentration. However, increasing the screening concentration can be cost-prohibitive, and result in the precipitation of the xenobiotic and consequently reduce its bioavailability. Also, the screening concentration is limited by the concentration of the stock solution and by the amount of solvent that the organism can reasonably tolerate. Chemicals in purchasable libraries are typically dissolved in dimethylsulfoxide (DMSO) at a concentration of 10 mM. DMSO concentrations greater than 1% can adversely affect the development and behavior of *C. elegans* (unpublished observations). Thus, the upper concentration limit for a screen of purchasable chemicals in *C. elegans* is fixed at 100 mM and published chemical screens employing *C. elegans* have typically used concentrations equivalent to or lower than this value (Gosai et al., 2010; Kwok et al., 2006; Lemieux et al., 2011; Min et al., 2007; Petrascheck et al., 2007; Samara et al., 2010).

A second method to increase target accessibility in the worm is to spike the screening media with a small amount of mild detergent such as Triton X-100 or Tween-20. For example, wild-type *C. elegans* treated with 0.1% Triton X-100 arrests at the first larval stage when exposed to 200 mg/ml of the mushroom toxin α-amanitin, whereas detergent-free worms are resistant to α-amanitin at concentrations as high as 800 mg/ml (Rogalski et al., 1990). Furthermore, Semple et al. found that wild-type worms treated with 0.1% Triton X-100 showed more than 100-fold greater sensitivity to 0.5 mg/ml puromycin after 2 days of exposure,
relative to detergent-free animals (Semple et al., 2010). Triton X-100 also sensitized worms to
two other antibiotics, G418 and phleomycin. Presumably the detergent permeabilizes the
cuticle and/or the intestine; however, the mechanism of sensitization is unclear. In addition to
serving as permeability agents, mild detergents could also solubilize lipophilic molecules in the
screening media that would otherwise precipitate out of solution, thereby enhancing
compound availability to the worm. To my knowledge, mild detergents have never been used
as permeability agents in C. elegans high-throughput chemical screens, but they could easily be
incorporated into the media preparation for this type of screening platform. Before universally
adopting detergent-spiked media for C. elegans chemical screens, it is critical to assess how
general the detergent-sensitizing effects are for a broad range of pharmacologically active
molecules. Purchasable collections of pharmacologically active compounds are available that
can be used for this purpose (Irwin and Shoichet, 2005).

Another approach to sensitize worms to xenobiotics is to screen them in the background
of a low dose of a characterized compound that elicits the desired phenotype at higher
concentrations. This type of “combination chemical screen” has the potential to sensitize
worms to new bioactive compounds that would otherwise elicit no phenotype when
administered alone. In addition to sensitizing worms to novel bioactives, combination chemical
screens also have the potential to uncover therapeutically useful synergistic chemical
combinations that would be missed in a “single-chemical screen”. A caveat to combination
chemical screens is that chemical combinations can act antagonistically (i.e. one compound can
reduce the efficacy of another) (Lehar et al., 2008). However, this limitation is easily
circumvented by screening single molecules in parallel to any combinatorial screen. Also, our
lab has found that pooling large numbers of chemicals generally reduces hit rates. In previous
chemical screens it was found that pooling eight small molecules per well, as opposed to
screening the same chemicals one molecule per well, resulted in the identification of 10-fold
fewer actives ((Burns et al., 2006; Kwok et al., 2006) and unpublished observations). The
reason for the reduction in hit rate is not clear. One possibility may be that the worm’s
xenobiotic defenses are upregulated in response to the large number of different structures
and that the upregulated defenses limit the accessibility of what would otherwise be a genuine
hit. Regardless, combination chemical screens have the potential to sensitize worms to new
bioactive compounds that would otherwise elicit no phenotype when administered alone.

3.4.2 Liposomal Delivery

*C. elegans* is a filter feeder, meaning that it draws bacterial cells suspended in liquid into its
pharynx, traps the bacteria, and spits out the liquid (Fang-Yen et al., 2009). This feeding
strategy likely selects against the oral uptake of hydrophilic small molecules that are not taken
up by the bacteria, and remain dissolved in the aqueous culture solution. To circumvent this
barrier to hydrophilic small molecule accumulation in *C. elegans*, Shibamura et al. made use of
a liposomal delivery system in which hydrophilic small molecules are incorporated into
liposomes made of phospholipid bilayers, and these liposomes are then orally administered to
the worms by mixing them with the bacterial food source (Shibamura et al., 2009). To test their
approach, they fed worms liposomes loaded with the fluorescent dye fluorescein, and
measured the amount of dye present in the worms after 3 hours of incubation. They found
that dissolving fluorescein in the culture medium without liposomes required 100-fold more
dye, compared with the liposomal delivery method, to achieve similar fluorescein amounts in
worm tissue. In addition, they tested the ability of four water soluble antioxidant chemicals to
extend the lifespan of worms using their liposomal delivery method. They found that all four antioxidants significantly extended the life of worms, relative to control liposomes, and that similar concentrations of the antioxidants dosed to the worms in the absence of liposomes had no effect on lifespan. The authors suggest that the liposomes mimic bacterial cells and are not filtered out by the pharynx, permitting delivery of the liposomes (and their contents) to the lumen of the worm intestine. One drawback of the liposomal delivery method is that it is low-throughput, and would be difficult to adapt to a high-throughput screen of thousands of chemicals. Nevertheless, this approach holds promise for small-scale C. elegans chemical-genetics experiments employing water soluble compounds.

3.4.3 Drug-Sensitive Mutants

As I previously described, C. elegans has evolved extensive physical and enzymatic defenses to resist the uptake and accumulation of xenobiotics in its tissues. Thus, mutant worms that are defective in xenobiotic defense have the potential to exhibit multidrug sensitivity, and mutants of this class would be invaluable for biological probe and drug discovery efforts. For example, C. elegans pgp-3 mutants are hypersensitive to the growth defects induced by colchicine and chloroquine – two structurally distinct molecules with differing modes of action (Broeks et al., 1995). The C. elegans PGP-3 protein is expressed in the apical membranes of the intestinal cells and the excretory cell, consistent with a role in xenobiotic efflux. A C. elegans nhr-8 deletion mutant is also sensitive to colchicine and chloroquine (Lindblom et al., 2001). The nhr-8 promoter drives expression in the intestine, consistent with a role in xenobiotic defense in this tissue. Both the pgp-3 and nhr-8 mutants develop normally and are morphologically indistinguishable from wild-type worms.
Mutations in a number of genes required for *C. elegans* cuticle integrity also result in drug hypersensitivity (Bounoutas et al., 2009; Gravato-Nobre et al., 2005; Partridge et al., 2008). Partial loss-of-function mutants of one of these genes, *bus-8*, are hypersensitive to nicotine, 1-phenoxypropan-2-ol, and the anthelmintic ivermectin (Partridge et al., 2008). These molecules are structurally disparate and have distinct modes of action. *bus-8* mutants also show increased uptake of Hoechst 33258 dye, suggesting that these mutants are generally more permeable to exogenous molecules. Finally, a *C. elegans* mutant called *dal-1*, which was identified from a forward genetic screen for mutant worms with abnormal intestinal morphologies, exhibits multidrug sensitivity (Paulson and Waddle, personal communication). Despite having altered gut morphology, characterized by membrane invaginations that span the length of the intestine, *dal-1* worms develop normally and are grossly indistinguishable from wild-type worms. *dal-1* mutants are hypersensitive to the effects of colchicine and chloroquine, the neurotransmitters serotonin and octopamine, as well the L-type calcium channel antagonists nemadipine A, felodipine, and verapamil. Presumably the altered intestinal morphology of *dal-1* mutants improves the absorption of solutes dissolved in the intestinal lumen; however, the mechanism of *dal-1* drug hypersensitivity has not been determined.

My survey of *C. elegans* xenobiotic metabolism suggests that the majority of lipophilic molecules that penetrate the worm, and are metabolized, are rendered more hydrophilic via polar group functionalization and/or glucose conjugation; thereby promoting excretion from the animal and preventing subsequent uptake. These biotransformations are likely carried out by one or more of the 86 cytochrome p450 (CYP) and 72 UDP-glucosyltransferase (UGT) enzymes encoded by the worm genome (Lindblom and Dodd, 2006). Knocking down the
function of the enzymes that carry out these modifications would likely sensitize worms to a broader range of xenobiotics. The responsible enzymes can be identified by using RNAi targeted against the worm’s CYP and UGT genes, and monitoring the metabolism of the five molecules I have shown to be polar group functionalized and/or glucosidated using my HT-HPLC method. Future chemical screens could then employ worms that are mutant for the identified genes, with the goal of sensitizing worms to a broader chemical space. It is possible that many CYP or UGT genes act to modify xenobiotics in C. elegans, and that these enzymes act redundantly. In this case, the RNAi knockdown of one gene may not reveal its role in drug biotransformation. However, as is the case with human CYP3A4/5 and CYP2D6 enzymes, which metabolize the majority of clinically used drugs (Eddershaw and Dickins, 2004), it is possible that only one (or a few) of the worm CYPs and UGTs act to modify xenobiotics.

The use of multidrug-sensitive mutants in high-throughput chemical screens will likely uncover novel bioactives that would otherwise elicit no phenotype in wild-type worms. However, the utility of the above-mentioned mutants in large-scale screening ventures has, to my knowledge, not been demonstrated. Before employing these mutants in screens for novel anthelmintics, it will be necessary to determine if they are generally sensitive to a broad range of pharmacologically active molecules, and to assess the well-to-well variability in their growth and behavior. Nonetheless, multidrug-sensitive mutants have great potential to improve the hit rates of C. elegans chemical screens.

3.4.4 Small Molecule Carriers (SMoCs)

A relatively new approach for delivering molecules into cells is the use of cell-penetrating peptides (CPPs) (Joliot and Prochiantz, 2004; Okuyama et al., 2007; Schwarze et al., 1999;
CPPs can facilitate the cellular uptake of a variety of cargo molecules that range in size from nano-particles to full-length proteins, both in vitro and in vivo. In general, CPPs contain less than 30 amino acids, with multiple basic residues, and often adopt an amphipathic alpha-helical structure when exposed to cell membranes. CPPs are linked to their cargo molecules via covalent bonds or through non-covalent interactions, depending on the CPP. The HIV-1 TAT-derived peptide is well-known example of a CPP (Joliot and Prochiantz, 2004) and references therein). Kim et al. have shown that wild-type *C. elegans* orally dosed with 500μM of platinum nanoparticles (nano-Pt, a superoxide dismutase mimetic) live 24% longer than untreated worms; whereas worms treated with only 5μM of TAT peptide-conjugated nano-Pt live 25% longer than both untreated worms and worms treated only with the TAT peptide (Kim et al., 2010). In this case, TAT peptide-conjugation resulted in a 100-fold greater sensitivity to the biological effects of nano-Pt, suggesting that the TAT peptide can be used to deliver exogenous molecular cargo to the tissues of whole worms.

Recently, a small molecule mimic of an amphipathic alpha-helical CPP domain was synthesized that can deliver both small molecule and protein cargos into cultured cells (Okuyama et al., 2007). This small molecule carrier (SMoC) consists of a biphenyl core scaffold with guanidine side chains emanating from the 2, 3, 5’, and 6’ positions (Fig. 3-3). A free primary amine reactive group emanates from the 4’ position of the biphenyl scaffold, which allows a variety of conjugation chemistries depending on the need. The SMoC has two important advantages over CPPs for the delivery of small molecules into cells and living organisms: 1. Reduced susceptibility to biodegradation, and 2. Amenable to simple attachment chemistries that can be carried out in multi-well plates. The first advantage is particularly important in the context of a living organism, such as *C. elegans*, that can metabolically
Figure 3-3. Reaction scheme for the generation of SMoC-tagged chemical libraries. (A) The synthesis of a SMoC amenable to downstream “click” chemistry is shown. The “clickable” SMoC is synthesized via an amide bond formation reaction, between the SMoC and azide-PEG4-acid, using the carbodiimide EDC as a coupling reagent (Montalbetti et al. (2005)). Azide-PEG4-acid is commercially available. The “clickable” SMoC contains a terminal azide group linked to the SMoC via a polyethylene glycol (PEG) spacer. The tert-butyloxycarbonyl (Boc) protecting groups can be removed by treatment with concentrated acid. (B) Synthesis of a theoretical SMoC-tagged small molecule via an azide-alkyne Huisgen 1,3-dipolar cycloaddition “click” reaction (Kolb et al. (2001), Srinivasan et al. (2007)). The click reaction can be carried out in water at room temperature in multi-well plates, making it amenable to the arrayed synthesis of SMoC-tagged chemical libraries. Collections of small molecules comprised solely of compounds with terminal acetylenes have been assembled and arrayed into 96-well plates, for similar purposes.
degrade CPPs. The second advantage is key for the construction of large libraries of carrier-tagged small molecules for high-throughput screening. Libraries of SMoC-tagged compounds could easily be generated using a simple “click” chemistry (Kolb et al., 2001; Srinivasan et al., 2007) reaction that is performed in water in multi-well plates (see Fig. 3-3 for a reaction scheme).

The enhanced bioavailability of SMoC-tagged chemical libraries would distinguish them as valuable reagents for the identification of novel probes from high-throughput chemical screens in C. elegans, and other organisms as well. To my knowledge, SMoCs are not yet commercially available, though combinatorial methods for the synthesis of large amounts of these carriers have been described (Rebstock et al., 2008). Before adopting SMoC-tagged libraries as a C. elegans screening tool it will be necessary to perform small-scale experiments with perhaps a few hundred compounds to determine if SMoC-tagging improves the rate at which bioactives are identified, and to assess any SMoC-related toxicity. Nonetheless, SMoC-tagged chemical libraries have great potential as a means to better deliver exogenous small molecules into the tissues of worms.

3.5 Towards a More Complete Map of Worm-Bioavailable Chemical Space

The C. elegans structure-based accumulation model was built using accumulation data from a training set of less than 500 molecules, taken primarily from the Spectrum library (Microsource) and a DiverSet library (Chembridge). Despite the relatively small size of the training set, the SAM performs quite well at predicting compounds that will accumulate in
worms based on their structures. However, a more complete map of worm-bioavailable chemical space could be generated by training the SAM on a larger dataset that better reflects the structural space of purchasable compounds. To that end, a library of one to ten thousand compounds, selected from multiple chemical vendors, could be assembled and assayed for accumulation in worms. This data could then be used to train a SAM that is more broadly applicable across purchasable chemical space.

The worm-bioavailable chemical space I have described so far is contained within human orally bioavailable chemical space (Fig. 3-1); a consequence of training the SAM with purchasable compounds, the majority of which are biased towards being orally bioavailable to humans. This raises an interesting question: How extensive is the worm-bioavailable chemical space that falls outside of that which is bioavailable to humans? One possibility is that worm-bioavailable chemical space is almost completely contained within human bioavailable chemical space, and only a small fraction can accumulate in worms (Fig. 3-1B). Another possibility is that worm- and human-bioavailable chemical spaces are similar in size, but only a small portion of these spaces overlap (Fig. 3-1A). The latter possibility raises the prospect of a vast worm-specific chemical space that could be mined for bioactives that are available to C. elegans and possibly other nematodes. It is difficult to speculate on the type of chemical matter that would be contained in such a worm-specific chemical space, as it has never been described.

3.6 Conclusion

The resistance of C. elegans to bioactive drug-like compounds has impeded its use as a model organism for chemical genetics and drug discovery. The work I have presented here sheds light
on the area of chemical space that is bioavailable to worms, and provides a roadmap to the structures that can access target proteins in whole worms and induce phenotypes. As the coverage of worm-bioavailable chemical space improves, so too will our ability to identify bioactive compounds from *C. elegans* screens. In combination with the drug delivery methods I described, the structure-based accumulation model of worm-bioavailable chemical space allows the circumvention of the worm’s xenobiotic defenses, and positions *C. elegans* to emerge as a premier model system for the discovery and characterization of novel small molecule probes and drugs to treat human disease.
Chapter 4

Small Molecule Screening in C. elegans Uncovers Putative Modulators of the Metabotropic Glutamate Receptor MGL-1

The work in this chapter was performed by me with two exceptions. Rachel Puckrin (Roy Lab, University of Toronto) provided technical assistance with the C. elegans liquid-based chemical screen. All cell-based mGluR functional assays were performed by Sujeenthar Tharmalingam in David Hampson’s lab at the University of Toronto. In addition William Ryu (University of Toronto) provided the camera, software, and technical assistance for the worm motility experiments.
Chapter 4. Small Molecule Screening in *C. elegans* Uncovers Putative Modulators of the Metabotropic Glutamate Receptor MGL-1

Abstract

Central nervous system (CNS) disorders are a worldwide health burden. There is a pressing need for new drugs to treat disorders of the CNS, yet few novel neuroactive compounds have been discovered in the last fifty years. Of late, metabotropic glutamate receptors (mGluRs) have surfaced as promising CNS drug targets. mGluRs are glutamate-regulated G-protein coupled receptors that modulate neuronal excitability and synaptic transmission throughout the CNS. The human genome encodes for eight mGluR subtypes, and preclinical evidence suggests that subtype-selective small-molecule modulators have the potential to treat a variety of CNS disorders including depression and Alzheimer’s disease. In spite of their therapeutic potential, only a handful of subtype selective mGluR modulators exist.

In this chapter, I describe a novel class of heterobiaryl amide compounds (HBACs), identified from a phenotypic screen in *C. elegans*, that are potential mGluR modulators. The HBACs are structurally similar to known small-molecule modulators of mGluR activity, and chemical-genetic analyses in the worm suggest that the HBACs may target the *C. elegans* mGluR MGL-1. It has recently been shown that mGluR pharmacology is conserved from worms to mammals. Thus, the HBACs have potential as tools to probe mGluR biology, and as leads for the development of CNS drugs.
4.1 Introduction

Disorders of the central nervous system (CNS), such as depression, schizophrenia, and bipolar disorder, are a significant global health burden and are among the leading causes of human years lost to disability (WHO, 2008). Hundreds of millions of people worldwide suffer from one or more neuropsychiatric disorder. There is an apparent need for improved drug therapies to treat CNS disorders, however very few novel neuroactive compounds have been discovered in the last fifty years (Agid et al., 2007).

In recent years, metabotropic glutamate receptors (mGluRs) have emerged as attractive targets for the development of drugs to treat CNS disorders ((Niswender and Conn, 2010) and references therein). mGluRs are family C G-protein coupled receptors that modulate neuronal excitability and synaptic transmission throughout the CNS in response to glutamate binding (see Fig. 4-1 for a schematic diagram of the mGluR structure). The human genome encodes eight mGluR subtypes that are classified into three groups based on sequence homology, G-protein coupling, and pharmacology. Group I includes mGluRs 1 and 5, group II includes mGluRs 2 and 3, and group III includes mGluRs 4, 6, 7 and 8. Group I receptors typically couple to the G12/G13 G-alpha subunit in response to glutamate binding leading to the downstream mobilization of intracellular calcium stores and PKC activation. Conversely, Group II and III receptors typically couple to the G12/13 G-alpha subunit, resulting in downstream inhibition of adenyl cyclase and activation of K⁺ channels. The different mGluR subtypes are distributed heterogeneously throughout the CNS, allowing for the treatment of distinct disorders with subtype-selective drugs. Indeed, preclinical evidence suggests that small-molecule modulators of specific mGluR subtypes have the potential to treat a number of CNS disorders including
Figure 4-1. Schematic diagram of a metabotropic glutamate receptor dimer. Metabotropic glutamate receptors (mGluRs) are glutamate-activated G-protein coupled receptors. mGluRs contain a large N-terminal extracellular domain, called the “venus flytrap” domain (VFD), that binds glutamate and other orthosteric modulators. Conformational changes induced by ligand binding are transduced from the VFD to the heptahelical and intracellular C-terminal domains via a cysteine-rich region. Ligand-induced conformational changes lead to the activation of associated G-proteins. Activated G-proteins can then modulate the activity of downstream effector proteins such as enzymes, ion channels, and transcription factors. mGluRs are constitutive dimers, and evidence suggests that full receptor activation is achieved only when glutamate occupies both binding sites in the VFDs. The binding sites for allosteric modulators, such as PAMs and allosteric agonists, are found in the heptahelical domain. This figure is an adaptation of a similar figure found in Niswender and Conn, 2010.
depression (Pilc et al., 2008), anxiety (Swanson et al., 2005), schizophrenia (Moghaddam, 2004), Parkinson’s disease (Conn et al., 2005), and Alzheimer’s disease (Lee et al., 2004).

Small-molecule modulators of mGluRs fall into two classes: 1. Orthosteric modulators that bind the glutamate binding site, and 2. Allosteric modulators that bind an allosteric site in the heptahelical domain of the receptor (Niswender and Conn, 2010) (Fig. 4-1). Both orthosteric and allosteric modulators can act as agonists (activators), inverse agonists (inhibitors), or antagonists of mGluR activity. Allosteric modulators can also act as positive allosteric modulators (PAMs) or negative allosteric modulators (NAMs), sensitizing or desensitizing the receptors to glutamate, respectively. Despite the therapeutic potential of mGluR modulators for treating neurological disorders only a handful of subtype specific modulators have been reported (Table 4-1). For example, the mGluR7 allosteric agonist AMN082 is the only subtype selective allosteric agonist available, and there are no subtype specific modulators for mGluR3, mGluR6, or mGluR8 (Niswender and Conn, 2010).

*C. elegans* is a potential platform for the discovery of novel small molecule mGluR modulators. The *C. elegans* genome encodes three mGlu genes: *mgl-1, mgl-2,* and *mgl-3* (Dillon et al., 2006; Tharmalingam et al., 2012). Based on sequence homology, the MGL-1, MGL-2, and MGL-3 proteins are thought to be orthologs of the mammalian group II, group I, and group III receptors, respectively (Dillon et al., 2006; Tharmalingam et al., 2012). Indeed, using a fluorescence-based calcium mobilization assay in HEK cells, Tharmalingam *et al.* (2012) showed that MGL-2 couples to G_{q}/G_{11} in response to glutamate treatment, consistent with the homology-based functional classification of this receptor (Tharmalingam et al., 2012). Furthermore, it has been shown that established orthosteric and allosteric modulators of mammalian group I mGluRs can modulate the activity of *C. elegans* MGL-2 in a mammalian
Table 4-1. mGluRs with reported subtype-selective modulators.

<table>
<thead>
<tr>
<th>Group</th>
<th>Receptor</th>
<th>Orthosteric agonist</th>
<th>Orthosteric antagonist</th>
<th>Allosteric agonist</th>
<th>NAM</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mGluR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mGluR5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>mGluR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mGluR3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>mGluR4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mGluR6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mGluR7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mGluR8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Subtype-selective modulator has been reported**
- **No subtype-selective modulator has been reported**
cell-based assay (Tharmalingam et al., 2012), suggesting that the pharmacology of mammalian mGluRs is at least partially conserved in worms.

In this chapter, I describe my characterization of a novel class of heterobiaryl amide compounds (HBACs) that may be acting to modulate the activity of *C. elegans* MGL-1. The HBACs were discovered through an unbiased phenotypic screen of 4,000 small organic molecules in worms. Despite having a unique core scaffold, the HBACs bear a structural resemblance to known mGluR modulators, and chemical-genetic analyses suggest that the HBACs may be acting as PAMs or allosteric agonists of MGL-1. Given that mammalian mGluR pharmacology is conserved in worms, it is possible that the HBACs modulate the activity of one or more mammalian receptors as well. Thus, the HBACs could provide a novel scaffold for the development of subtype specific mGluR modulators to be used as biological probes or as drugs to treat disorders of the CNS.

### 4.2 Results

#### 4.2.1 A Phenotype-Based Chemical Screen in Worms Uncovers a Heterobiaryl Amide Compound (HBAC) that Induces Motility Defects

As part of our lab’s ongoing efforts to identify small molecule probes of animal biology, I screened 4,000 small molecules for the induction of any obvious developmental or behavioural abnormalities in wild-type *C. elegans* (Materials and Methods 4.4.1 and 4.4.2). From this screen I identified a previously uncharacterized compound that induces motility defects in liquid (Fig. 4-2, Materials and Methods 4.4.3 and 4.4.4). Wild-type worms treated with the chemical have reduced swimming activity in liquid culture relative to untreated worms. This
Figure 4-2. A newly identified heterobiaryl amide compound induces motility defects in *C. elegans*. (A) The chemical structure of heterobiaryl amide compound A (HBAC-A), identified from a phenotypic screen in worms. (B) Still frames taken at 10-second intervals from 30-second videos of adult worms grown in liquid culture with or without HBAC-A.
newly identified molecule is a heterobiaryl amide compound (HBAC) that I refer to as HBAC-A (Fig. 4-2).

The motility defects induced by HBAC-A are dependent on a number of environmental conditions such as light exposure, temperature, and salt concentration, as well as other unidentified variables, resulting in a phenotype that is inconsistent from one experiment to another, and difficult to quantify. I therefore sought other approaches to elucidate the biological target of HBAC-A.

4.2.2 The HBACs are Structurally Similar to Known mGluR Modulators

To gain insight into the mode-of-action of HBAC-A, I searched the published literature for structurally similar molecules that have reported biological activity. I found two classes of structurally related molecules that have been shown to modulate the activity of mammalian metabotropic glutamate receptors (mGluRs) (see Fig. 4-3 for their structures) (Engers et al., 2009; Kulkarni and Newman, 2007). One structural class is comprised of mGluR5 NAMs (Kulkarni and Newman, 2007), while the other class comprises mGluR4 PAMs (Engers et al., 2009). Both classes share a common heterobiaryl amide substructure with HBAC-A. Thus, I hypothesized that the HBACs may be modulating the activity of one or more of the *C. elegans* mGluRs to affect motility, and possibly other processes as well.

Given the variability of the motility phenotype induced by HBAC-A, I was interested to find *C. elegans mgl* mutant phenotypes that could be used to test my hypothesis that HBAC-A is an mGluR modulator. So far, worm *mgl* genes have been implicated in the regulation of three distinct processes: 1. Reversal (or “backing up”) behaviour [unpublished results, worm meeting abstract], 2. Starvation-induced fat metabolism (Greer et al., 2008), and 3. Autophagy to
Figure 4-3. The heterobiaryl amide compounds (HBACs) are structurally similar to established mGluR modulators. (A) The molecular scaffold of the HBACs is composed of a quinoline ring and a phenyl ring linked together by an amide group. (B) Heterobiaryl amide mGluR4 PAM chemical structures (Engers et al. (2009)). (C) Heterobiaryl amide mGluR5 NAM core structure (Kulkarni and Newman (2007)).
promote survival during starvation (Kang and Avery, 2009). Of these three processes, starvation survival proved to be the most experimentally tractable option to test my hypothesis, since assays for the other two processes are technically more complicated and are prone to experimental variability (Chao et al., 2005; Yen et al., 2010).

4.2.3 HBAC-A Causes Premature Starvation-Induced Death (SID) in Worms, Consistent with it Modulating MGL-1 or MGL-2 Activity

Autophagy is a highly conserved lysosomal pathway used to degrade cytoplasmic organelles and long-lived proteins (Levine and Klionsky, 2004). In C. elegans, physiological levels of autophagy during starvation promote the survival of the organism; however, insufficient or excessive levels of autophagy cause premature starvation-induced death (SID) (Fig. 4-4A) (Kang and Avery, 2009; Kang et al., 2007; You et al., 2006). In the C. elegans pharyngeal muscle, the MAP kinase protein MPK-1 promotes autophagy, and is activated by a muscarinic acetylcholine receptor (mAChR) pathway in response to starvation (Fig. 4-4B) (Kang et al., 2007; You et al., 2006). The G-protein beta subunit GPB-2 negatively regulates muscarinic signaling, and acts to inhibit autophagy in the pharynx (Kang and Avery, 2009; Kang et al., 2007). gpb-2 mutants exhibit a premature SID phenotype that is suppressed by mutations in mpk-1 and by the small-molecule mAChR antagonist atropine (Kang et al., 2007).

It has been shown that mgl-1 loss-of-function and mgl-2 null mutations reduce levels of autophagy in gpb-2 mutants and rescue the premature SID in this background (Kang and Avery, 2009), suggesting that mgl-1 and mgl-2 normally act to promote autophagy in response to starvation. mgl-3 null mutations have no effect on the early SID of gpb-2 mutants. Intriguingly, MGL-1 and MGL-2 act in the nervous system of worms, in the AIY and AIB neurons
Figure 4-4. The HBACs cause premature starvation-induced death by engaging an excessive starvation response. (A) In *C. elegans*, physiological levels of autophagy promote survival during starvation, whereas insufficient or excessive levels lead to premature death (You et al. (2006), Kang et al. (2007), Kang and Avery (2009)). (B) In the *C. elegans* pharynx, a MAPK pathway acts to promote autophagy in response to starvation, and the type 5 G protein beta subunit (Gβ(5)) GPB-2 inhibits this pathway. A pathway suggested by and adapted from You et al. (2006) is shown. (C) Starvation survival assay for wild-type worms treated with vehicle, HBAC-A, or HBAC-B. (D) Starvation survival assay for wild-type and *gpb-2* null mutants treated with vehicle or HBAC-B. (E) Starvation survival assay for worms heterozygous or homozygous for an *mpk-1* null allele, treated with vehicle or HBAC-B. All experiments in C to E were performed in triplicate. Error bars show the standard error of the mean. A single asterisk indicates a p-value < 0.05 relative to vehicle control. A triple asterisk indicates a p-value < 0.0001 relative to vehicle control.
respectively, to modulate a systemic starvation response, possibly linking the perception of starvation with a physiological response in target tissues (Kang and Avery, 2009). Therefore, if HBAC-A alters the activity of MGL-1 or MGL-2 I would expect HBAC-A-treated wild-type worms to die prematurely when challenged with starvation. To test this hypothesis I performed starvation survival assays using wild-type worms treated with either HBAC-A or solvent as a control (Materials and Methods 4.4.5). HBAC-A-treated worms exhibit premature SID (Fig. 4-4C), consistent with this chemical modulating the activity of MGL-1 or MGL-2.

4.2.4 A SAM-Selected Structural Analog of HBAC-A, called HBAC-B, is a More Potent Inducer of Premature SID

The premature SID induced by HBAC-A is incompletely penetrant after fourteen days of starvation (Fig. 4-4C), and requires that the chemical be used at a concentration approaching its limit of solubility. In addition, more than nine days of starvation must transpire before a difference in survival is observed between HBAC-A-treated and untreated worms (Fig. 4-4C). Therefore, I was interested to find a structural analog of HBAC-A which is relatively more potent, to facilitate my characterization of the early SID phenotype.

HBAC-A has a sulfonamide-containing side chain that emanates from the 2’ position of the phenyl ring (Fig. 4-3). My structure-based accumulation model (see Chapter 2) identified the sulfonyl group as a common feature in molecules that fail to robustly accumulate in worm tissue. I therefore tested a structural analog of HBAC-A which lacks the sulfonamide side chain, but retains the HBAC core scaffold, with the hope that it would be relatively more potent. This new analog, which I call HBAC-B (Fig. 4-3), induces a completely penetrant early SID phenotype after only ~6 days of starvation, and at a 4-fold reduced concentration relative to HBAC-A (Fig.
4.4C). Given the improved potency of HBAC-B relative to HBAC-A, I performed the majority of my subsequent experiments with the HBAC-B analog, and refer to this family of structures as the HBACs.

4.2.5 HBAC Treatment Induces Premature SID by Engaging an Excessive Starvation Response, Consistent with MGL-1 or MGL-2 Agonism

Premature SID can result from either excessive or insufficient autophagy during starvation (Fig. 4-4A). I was therefore interested to determine the mechanism by which the HBACs cause early SID. As a first step towards addressing this question, I tested the HBAC sensitivity of gpb-2 null mutants, which engage an excessive starvation response and exhibit premature SID as a result. Using the starvation survival assay, I found that gpb-2 null mutants are hypersensitive to the premature SID caused by HBAC-B, relative to wild-type worms (Fig. 4-4D). This result suggests that the HBACs induce early SID by engaging an excessive starvation response, and that the HBAC target acts in parallel to gpb-2. Furthermore, since MGL-1 and MGL-2 normally act to promote autophagy (and the starvation response), this result is consistent with the HBACs acting as PAMs or agonists of one or both of these receptors. The observation that the HBACs share structural similarity with known mGluR modulators further supports this model.

If the HBACs are agonists of MGL-1 or MGL-2, and cause premature SID by engaging an excessive starvation response, I would expect a null mutation in mpk-1 to suppress the early SID of the HBACs. Consistent with this expectation, I found that an mpk-1 null allele suppresses the premature SID of HBAC-B (Fig. 4-4E). This result further substantiates the conclusion that the HBACs engage an excessive starvation response to cause premature SID, and supports a model whereby the HBACs act as agonists or PAMs of MGL-1 or MGL-2.
4.2.6 The HBAC Sensitivity Profile of mgl Mutants Suggests MGL-1 Agonism as a Possible Mode-of-Action

If the HBACs agonize MGL-1 or MGL-2, I would expect that deleting mgl-1 and/or mgl-2 from the worm genome would cause HBAC resistance. There are publicly available strains that carry deletion alleles for all three worm mgl genes (Materials and Methods 4.4.1). The mgl-2 and mgl-3 deletions occur early in the sequence and are out-of-frame, which implies that these alleles do not produce functional proteins (Fig. 4-5A). Thus, if MGL-2 is the sole target of the HBACs, the mgl-2 deletion strain should be HBAC-resistant. In contrast, the mgl-1 deletion is in-frame and occurs in the glutamate-binding domain, leaving the cysteine-rich, heptahelical, and C-terminal domains intact (Fig. 4-5A). mGluRs that lack the glutamate-binding domain, but are otherwise whole, can respond to treatment with allosteric modulators (Goudet et al., 2004). Indeed, small molecules that act as PAMs on whole receptors, can act as allosteric agonists on mGluRs missing their glutamate binding domain (Goudet et al., 2004). Thus, if the HBACs are MGL-1 allosteric agonists or PAMs, the mgl-1 deletion strain may be HBAC-sensitive.

I tested the starvation survival of all three mgl deletion strains, as well as strains carrying all possible double mutant combinations, in the presence or absence of HBAC-B. I found that all six strains have wild-type HBAC sensitivities (Fig. 4-5B). This result suggests that neither MGL-2 nor MGL-3 are the sole HBAC target protein, but does not eliminate the possibility that the HBACs act as PAMs or allosteric agonists of MGL-1.
Figure 4-5. The *mgl* mutant HBAC sensitivity profile. (A) Domain compositions of proteins produced from a wild-type *mgl* gene, and from *mgl-1* (tm1811), *mgl-2* (tm355), and *mgl-3* (tm1766) deletion alleles. The glutamate binding domain (GBD), cysteine-rich domain (CRD), heptahelical domain (HD), and C-terminal domain (CTD) are indicated. (B) Starvation survival assays for single and double *mgl* mutants treated with vehicle or 75μM HBAC-B. All experiments were performed in triplicate. Error bars show the standard error of the mean.
4.2.7 MGL-2 Activity is Not Modulated by the HBACs

The HBAC-sensitivity of the mgl-2 deletion strain rules out MGL-2 as the sole HBAC target protein; however, this result does not exclude the possibility that the HBACs are allosteric agonists or PAMs of both MGL-1 and MGL-2. A fluorescence-based calcium mobilization assay in HEK cells has previously been used to test the ability of glutamate, and other small-molecule mGluR modulators, to affect the activity of heterologously expressed MGL-2 (Tharmalingam et al., 2012). Using this assay (Materials and Methods 4.4.6 and 4.4.7), my collaborator Sujeenthar Tharmalingam and I tested the effects of HBAC treatment on the activity of MGL-2 and the mammalian mGluR5 (Fig. 4-6). For both receptors, we found no change in intracellular calcium concentration in response to HBAC treatment, before or after the addition of glutamate, suggesting that MGL-2 and mGluR5 are not HBAC target proteins.

The HEK cell-based calcium mobilization assay described above works primarily for group I mGluRs, since these receptors couple with Gq/G11 in response to agonist binding. Conversely, group II and III receptors typically couple with Gi/o, and produce no response in this assay. Based on sequence homology, MGL-1 is thought to be a group II receptor, and will likely not couple glutamate binding with intracellular calcium release. Consistent with its homology-based functional classification, we find that MGL-1 does not cause a change in intracellular calcium levels in response to glutamate treatment, despite the protein having conserved residues required for glutamate binding (Fig. 4-7), and despite abundant protein expression in HEK cells as determined by western blot. Previous studies have rendered group II and III mGluRs amenable to the calcium mobilization assay by co-transfecting cells with chimeric and promiscuous G-alpha proteins that couple group II and III receptor activation with intracellular calcium release (Corti et al., 1998; Gomez et al., 1996). We find that cells co-transfected with
**Figure 4-6. MGL-2 and mGluR5 are not HBAC target proteins.** The effect of HBAC treatment on MGL-2 and mGluR5 activity was tested using a fluorescence-based calcium mobilization assay in HEK cells. Both N-terminally and C-terminally FLAG-tagged versions of MGL-2 were used for these experiments, as indicated. The top and bottom graphs show mGluR activity before and after glutamate addition, respectively. R.F.U (relative fluorescent units) are proportional to the intracellular calcium concentration, which serves as a proxy for receptor activation (see Materials and Methods 4.4.6). The experiments were performed in duplicate.
The National Center for Biotechnology Information protein database. The highlighted in grey have similar biochemical properties as the mGluR1 residue. Mutations in the Of these residues, those highlighted in black are identical to the mGluR1 residue, and those

<table>
<thead>
<tr>
<th>Rat_mGlu1</th>
<th>Rat_mGlu2</th>
<th>Rat_mGlu3</th>
<th>Rat_mGlu4</th>
<th>Rat_mGlu5</th>
<th>Rat_mGlu6</th>
<th>Rat_mGlu7</th>
<th>Rat_mGlu8</th>
<th>Worm_MGL</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLQSP</td>
<td>RIAK</td>
<td>NMMFLRQLMKDFGEK</td>
<td>MSLGHPWAVWPPS</td>
<td>MFLVHKLLEWLRG</td>
<td>NMVLRFLQQK</td>
<td>MFLVHKLLEWLRG</td>
<td>NMVLRFLQQK</td>
<td>MFLVHKLLEWLRG</td>
</tr>
<tr>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
</tr>
<tr>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
</tr>
</tbody>
</table>

Rat_mGlu1 | Rat_mGlu2 | Rat_mGlu3 | Rat_mGlu4 | Rat_mGlu5 | Rat_mGlu6 | Rat_mGlu7 | Rat_mGlu8 | Worm_MGL |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KLQSP</td>
<td>RIAK</td>
<td>NMMFLRQLMKDFGEK</td>
<td>MSLGHPWAVWPPS</td>
<td>MFLVHKLLEWLRG</td>
<td>NMVLRFLQQK</td>
<td>MFLVHKLLEWLRG</td>
<td>NMVLRFLQQK</td>
<td>MFLVHKLLEWLRG</td>
</tr>
<tr>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
<td>RIGKQ</td>
</tr>
<tr>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
<td>VFD</td>
</tr>
</tbody>
</table>

Figure 4-7. Sequence alignment of the amino-terminal domains of rat mGlur1-8 and worm MGL-1. The alignment was performed using ClustalW 2.1. All sequences begin at the initiator methionine. Residues found in the ligand binding pocket of the mGlur1 crystal structure that are identical across all eight mammalian mGlur1s are denoted with an asterisk (Kuang et al., 2003). Of these residues, those highlighted in black are identical to the mGlur1 residue, and those highlighted in grey have similar biochemical properties as the mGlur1 residue. Mutations in the residues indicated with an asterisk have been shown to reduce or abolish ligand binding (Kuang et al., 2003, Muto et al., 2007, Hampson et al., 1999). Rat mGlur sequences were obtained from the National Center for Biotechnology Information protein database. The C. elegans MGL-1 sequence was obtained from WormBase.
the MGL-1 receptor and either chimeric or promiscuous G-alpha proteins are still unable to mobilize calcium in response to glutamate treatment. Thus, the calcium mobilization assay is not a useful tool for assessing MGL-1 receptor pharmacology, and alternative methods for testing HBAC activity at the MGL-1 receptor are needed.

Clearly, more direct evidence that the HBACs target MGL-1 is required before any strong conclusions about the HBAC mode-of-action can be made. Nevertheless, my chemical-genetic analyses support a model that the HBACs act as MGL-1 allosteric agonists or PAMs, and I will discuss alternative means of target identification below.

4.3 Discussion

In the previous section, I described a chemical screen in *C. elegans* that uncovered a novel class of heterobiaryl amide compounds (HBACs) that may be acting as allosteric agonists or PAMs of the worm mGluR MGL-1. Two lines of evidence support this model: 1. The HBACs are structurally similar to known mGluR modulators, and 2. The HBACs cause premature starvation-induced death in a manner consistent with MGL-1 allosteric agonism or positive allosteric modulation. A small-molecule agonist of MGL-1 would be a useful reagent to probe MGL-1 function *in vivo*, since strains carrying hypermorphic *mgl-1* alleles are not currently available. For example, MGL-1 acts in the AIY interneuron to regulate the starvation response, but the mechanisms by which the AIY signals to remote tissues, such as the pharyngeal muscles, to modulate the starvation response is not known (Kang and Avery, 2009). Forward genetic and RNAi screens for mutations and gene knockdowns that cause resistance to the premature HBAC-induced SID could reveal these mechanisms. Furthermore, testing the HBAC-resistance of worms in which candidate “starvation response-regulatory” neurons have been
ablated could reveal the neural pathway by which the nervous system communicates the starvation state of the animal to more peripheral tissues.

The utility of a small-molecule MGL-1 agonist is clear, however more direct evidence that the HBACs are MGL-1 agonists is required before they can be used confidently as probes. As I described in the previous section, the calcium mobilization assay used to investigate group I mGluR pharmacology is not amenable to MGL-1. The likely reason for this is that MGL-1 is most similar phylogenetically to the group II mGluRs (Dillon et al., 2006; Tharmalingam et al., 2012), which do not couple glutamate binding with intracellular calcium release. In contrast to group I receptors, group II and III receptors typically associate with the $G_{i/o}$ G-alpha protein and glutamate binding to these receptors generally causes adenylyl cyclase inhibition and $K^+$ channel activation (Niswender and Conn, 2010). Recently, a cell-based functional assay was developed that can be used to interrogate the pharmacology of $G_{i/o}$-coupled GPCRs (Niswender et al., 2008). This novel assay capitalizes on the ability of the liberated G-beta-gamma subunits of $G_i$ and $G_o$ heterotrimers to interact with G-protein-coupled inwardly rectifying potassium channels (GIRKs) (Knoflach and Kemp, 1998; Saugstad et al., 1996), and the ability of potassium channels to conduct the ion thallium (Weaver et al., 2004). In brief, HEK cells co-expressing the GIRK channel and the mGluR of interest are incubated in buffer containing thallium, the cells are treated with the test ligand, and thallium flux is measured using a fluorescent indicator of the thallium ion. Using this method, agonists cause a receptor-dependent increase in thallium flux. The GIRK assay may have utility for assessing MGL-1 pharmacology, and could be used to test my hypothesis that the HBACs are MGL-1 PAMs or allosteric agonists. The GIRK assay could also be used to test the ability of the HBACs to modulate the activity of mammalian group II and III mGluRs.
The utility of the GIRK assay for HBAC target identification relies on the assumption that the HBACs are MGL-1 modulators. The chemical-genetic data I reported in this chapter support a model whereby the HBACs act as allosteric agonists or PAMs of \textit{C. elegans} MGL-1; however, it is possible that the HBACs have a distinct protein target. Relatively less biased methods would also be useful for identifying the HBAC target protein. For example, a forward genetic screen for worms that are resistant to the premature SID caused by HBAC-B, and the subsequent mapping of genetic loci, could identify potential target genes. Day 6 of starvation is a good time point to assess the HBAC-B-resistance of mutant worms, since 100% of HBAC-B-treated worms fail to recover from this amount of starvation and no SID is observed for solvent-treated worms at this time point. If the HBACs are allosteric agonists or PAMs of the MGL-1 receptor, a screen of the second filial generation of mutagenized parent worms could uncover loss-of-function mutations in the \textit{mgl-1} gene that confer HBAC-resistance. Over and above the target gene, this approach could identify other genes that act as positive regulators of the starvation response pathway.

In addition to forward genetic screens, chemical proteomics methods can be used to profile the protein targets of small molecules (\cite{Rix:2009} and references therein). One approach involves the immobilization of the small molecule on a solid substrate to create an affinity matrix for the purification of target proteins from tissue lysates, and their subsequent identification by mass spectrometry (MS). For this approach, small molecules must often be chemically modified to accommodate coupling to the matrix. These modifications, as well as matrix coupling, can render the small molecules unable to bind their target proteins. It is therefore crucial to determine whether the functionalization of small molecules for coupling disrupts bioactivity. With these concerns in mind, Min \textit{et al.} (2007) constructed a library of
small molecules pre-tagged with a side chain amenable to downstream coupling reactions (Min et al., 2007). The authors screened this library in *C. elegans* for molecules that suppress the constitutive dauer phenotype of *daf-2* insulin receptor mutants. They discovered a compound which they call GAPDS that suppresses the *daf-2* dauer phenotype and promotes reproductive development in this background. For target identification, they coupled the GAPDS molecule to an activated carboxylic acid affinity resin, and performed protein affinity purification using worm extracts. Subsequent analysis of the affinity purified proteins by SDS-PAGE and MS revealed glyceraldehyde-3-phosphate dehydrogenase as a GAPDS target protein. This type of approach could be used for HBAC target identification as well. The HBAC-A molecule contains a terminal acetylene functional group that can be easily reacted with molecules containing terminal azide groups using a simple “click chemistry” reaction (see section 3.3.4 of Chapter 3) (Kolb et al., 2001; Srinivasan et al., 2007). Polyethylene glycol (PEG) molecules with an azide functional group at one terminus and a free amine at another are commercially available. Thus, HBAC-A can be conjugated to such an azide-PEG-amine via click chemistry, and the free amine of this conjugate can then be coupled to activated carboxylic acid resins to create an HBAC affinity matrix. Furthermore, I have shown that the acetylene functional group and the sulfonamide-containing side chain of HBAC-A are dispensable for activity, so affinity coupling of HBAC-A should not disrupt target protein binding. This chemical proteomic approach to target identification complements forward genetic screens as an unbiased method for the elucidation of the HBAC target protein.

Here, I have described a novel family of heterobiaryl amide compounds (HBACs) that promote the starvation response in *C. elegans*. These compounds were identified from a phenotype-based chemical screen in worms that I performed. Chemical-genetic analyses
suggest that these compounds may modulate the activity of the worm metabotropic glutamate receptor MGL-1, although more direct evidence in support of this hypothesis is needed. I have discussed three distinct target identification methods that can be used in the future to help elucidate the target protein of the HBAC molecules. If the HBACs are found to be bona fide MGL-1 agonists or PAMs, they will serve as useful reagents to probe the role of group II mGluRs in the biology of animals. Furthermore, since mGluR receptor pharmacology is conserved from mammals to worms, it is possible that the HBACs I have discovered could modulate the activity of mammalian mGluRs, positioning them as probes of mammalian mGluR biology as well as templates for the design of mGluR subtype-selective drugs to treat CNS disorders.

4.4 Materials and Methods

4.4.1 Worm Culture and Strains

Worms were cultured at 20°C using standard techniques (Lewis and Fleming, 1995) unless otherwise indicated. The following strains were obtained from the C. elegans Genetics Center (U. Minnesota): N2 (wild-type), JT603 (gpb-2(sa603)I), and MT8186 (mpk-1(oz140)/dpy-17(e164) unc-79(e1068)III). The following strains were obtained from S. Mitani (National Bioresource Project, Japan) and were out-crossed four times before use in experiments: FX01811 (mgl-1(tm1811)X), FX00355 (mgl-2(tm355)I), and FX01766 (mgl-3(tm1766)IV). The following strains were made using standard genetic methods: RP2322 (mgl-1(tm1811)X; mgl-2(tm355)I), RP2323 (mgl-1(tm1811)X; mgl-3(tm1766)IV), and RP2324 (mgl-2(tm355)I; mgl-3(tm1766)IV).
4.4.2 *C. elegans* Liquid-Based Chemical Screen

The liquid-based chemical screening protocol used here is an adaptation of an established method for performing *C. elegans* RNA interference screens in liquid culture in 96-well plates (Lehner et al., 2006). 160 mL of a saturated HB101 *E. coli* culture in LB broth was two-fold concentrated in sterile liquid nematode growth media (3 mg/mL NaCl, 2.5 mg/mL Bactopeptone, 5 μg/mL cholesterol, 1 mM CaCl₂, 1 mM MgSO₄, 25 mM KH₂PO₄). Using a multichannel pipette, 40 μl of the NGM+HB101 media was added to the wells of a 96-well flat-bottom culture plate. Using a 96-well pinning tool (V & P Scientific) with a 300 (± 20) nL slot volume, chemicals were transferred from the stock library plates to the culture plates. Using a multichannel pipette with siliconized tips, 10 μL of M9 buffer (Burns et al., 2006) containing ~20 developmentally synchronized first larval stage worms (obtained from a bleach prep (Burns et al., 2006) of gravid adults performed the previous day) was added to each well of the 96-well culture plate. The culture plates were sealed with parafilm, covered with wet paper towels, and stored in a Tupperware box. The boxed plates were incubated for 4 days at 20°C in a shaking incubator set at 200 rpm. After 3 and 4 days of incubation, each well was screened for worms exhibiting any obvious developmental or behavioural defects. The chemicals in the stock library plates are at a concentration of 10 mM in DMSO; thus, the final screening concentration is ~60 μM (0.6% DMSO (v/v)). 160 mL of a saturated HB101 *E. coli* culture provides enough worm food to assay twenty library plates (or 1,600 different chemicals).

The library of 4,000 small molecules screened using the liquid-based assay was provided by Sean Cutler (UC Riverside). All molecules in this library contain a terminal acetylene group, and are therefore amenable to the azide-alkyne Huisgen 1,3-dipolar cycloaddition “click” reaction (Kolb et al., 2001; Srinivasan et al., 2007).
4.4.3 Purchased Chemicals

HBAC-A (3-Prop-2-ylnylsulfamoyl-N-quinolin-2-yl-benzamide) and HBAC-B (N-Quinolin-2-yl-benzamide) were purchased as powder from Enamine Ltd. and Ryan Scientific Inc., respectively.

4.4.4 Swimming Videos

Using the 96-well liquid-based chemical screening method described in section 4.4.2, worms synchronized at the first larval stage were treated with 50μM HBAC-A, or DMSO (0.6% v/v) as a control, and three days later 30-second videos were captured of adult worms swimming. A Leica MZ7.5 stereomicroscope equipped with a Basler A602f-2 camera was used for imaging. Videos were captured at a rate of 10 frames per second using LabVIEW software (National Instruments™).

4.4.5 Starvation Survival Assay

Starvation survival assays were performed as described (You et al., 2006), with minor modifications.

Worm embryos were harvested from a mixed-stage population by bleach preparation (Burns et al., 2006), and allowed to hatch overnight in 6 mL of sterile M9 buffer (Burns et al., 2006) on a nutating shaker at 20°C. The following day (i.e. day 1 of starvation), the developmentally synchronized first larval stage hatchlings were washed once with 12mL of sterile M9 buffer, and then re-suspended in enough buffer to achieve a worm concentration of ~8400 worms per milliliter. For each starvation assay, 2.37 mL of the worm suspension was transferred to a new
15 mL conical tube. 600 μL of sterile ddH2O was added to each suspension, as well as 30 μL of HBAC dissolved in DMSO (for the experimental assay) or 30 μL of DMSO alone (for the solvent control assay). The worms were incubated on a nutating shaker at 20°C for the times indicated in the figures. At each time point, a 15 μL aliquot from each sample tube was deposited on a 6 cm MYOB plate (Burns et al., 2006) seeded with OP50 E. coli. After 3 days of recovery on the MYOB plate at 20°C, the number of worms that reached the second larval stage or older was counted. The number from day 1 of starvation was used as a control and as the denominator to calculate the fraction of surviving animals after starvation. 4 days of recovery on the MYOB plate was allowed for the gpb-2(sa603) null mutants.

For the MT8186 mpk-1(oz140) null strain, survival assays were performed as described above but with two differences: (1) At each time point, three 15ul aliquots from each sample tube was deposited onto three separate 6 cm MYOB plates seeded with OP50 E. coli. (2) Worms were allowed to recover on the MYOB plates for 4 days at 20°C, at which point only the worms that grew to adulthood were counted. mpk-1(oz140) null mutants are 100% penetrant for germline pachytene arrest (Church et al., 1995; Lackner and Kim, 1998). As a result, these worms do not produce mature oocytes and are completely sterile. The MT8186 strain contains the mpk-1(oz140) null allele balanced over dpy-17(e164);unc-79(e1068). This strain throws mpk-1(oz140) homozygotes that are sterile and mpk-1(oz140) heterozygotes that are phenotypically wild-type. Progeny that are homozygous wild-type for mpk-1 are phenotypically dumpy and uncoordinated. For the starvation survival assay, the number of non-dumpy, non-uncoordinated adult worms was counted after 4 days of recovery to determine how many mpk-1(oz140) heterozygotes (with mature oocytes) and homozygotes (without mature
oocytes) survived starvation. Thus, the heterozygotes and homozygotes were sampled from the same tube, providing an excellent internal control for the experiment.

### 4.4.6 Molecular Cloning and Expression Construct Preparation

pPRARB749 (pCMV6::mgl-2::FLAG) and pPRARB750 (pCMV6::FLAG::mgl-2) were made by PCR-amplifying the full length mgl-2 ORF from the plasmid pBluescript::mgl-2 (kind gift from T. Ishihara, Kyushu University) using primer pairs 749 and 750 respectively (see below), and inserting the resulting PCR amplicons into the mammalian expression vector pCMV6 (kind gift from C. Cummins, University of Toronto).

**Primer pair 749:**

5' ‑ GATTGAGATCTATGCTCTACAGTCATGTCACACTTC-3';
5' ‑ GACCGGCGGCCTATTTATCGTCAATGGTGTAGCTTGAGACAGC-3'.

**Primer pair 750:**

5' ‑ GATTGAGATCTATGGACTACAAAGACGATGACGATAAAATGCTCTACAGTCATGTCACACTTC-3';
5' ‑ GACCGGCGGCCTATTTATCGTCAATGGTGTAGCTTGAGACAGC-3'.

The *C. elegans* mgl-1 gene is toxic to *E. coli* at medium to high copy number (T. Ishihara, personal communication) and was cloned into our pPRAB898 vector that contains the pCMV6 mammalian expression elements along with the low copy pSC101 bacterial origin of replication. pPRAB898 was made by PCR-amplifying the region of pMW118::mgl-1 (kind gift from T. Ishihara, Kyushu University) that contains the AmpR gene and pSC101 origin of replication using primer pair 898 (see below), and blunt-end ligating this amplicon with the blunted PciI and NgoMIV pCMV6 digestion product containing the mammalian expression elements and multiple cloning site. pPRAB899 (pPRAB898::mgl-1::FLAG) and pPRAB900 (pPRAB898::FLAG::mgl-1) were made by PCR-amplifying the full length mgl-1 ORF from the
pMW118::mgl-1 plasmid using primer pairs 899 and 900 respectively (see below), and inserting the resulting PCR amplicons into the pPRAB898 expression vector.

Primer pair 898:
5'--TTGCAAGCAGCAGATTACGC--3'
5'--TTTACACCCTTTTCATCTGTGC--3'

Primer pair 899:
5'--GATTGGAATTCGCCACCATGTTCTCGGTGCTTGCACTTG--3'
5'--GACCGGCGGCCGCTATTTATCGTCATCGTCTTTGTAGTCTAACAGAAGTACGTGAGCAGAGC--3'

Primer pair 900:
5'--GATTGGAATTCGCCACCATGGACTACAAAGACGATGACGATAAAATGTTCTCGGTGCTTGCACTTG--3'
5'--GACCGGCGGCCGCTCATAAGAAGTACGTGAGCAGAGC--3'

4.4.7 mGluR Cell-Based Functional Assay

The mGluR cell-based functional assays were performed as described (Tharmalingam et al., 2012). HEK-293-MSR cells were grown in DMEM supplemented with 10% fetal bovine serum and 1% glutamax in 5% CO₂ at 37°C. HEK cells grown to 80-90% confluency on 6 well-plates were transiently transfected with the mGluR expression constructs at 2 μg/well using Lipofectamine2000 as described by the manufacturers. The cells were trypsinized 20 hours post-transfection and plated on poly-l-ornithine treated 96-well microtiter plates (Costar #3603) at 100,000 cells/well in normal culture media. After 24 h of culture, the cells were equilibrated with Ca²⁺ assay buffer (20 mM HEPES, 146 mM NaCl, 5 mM KCl, 1 mM MgCl₂, 1 mM CaCl₂, 1 mg/ml bovine serum albumin, and 1 mg/ml glucose; pH 7.4), loaded with 6 μM Ca²⁺-sensitive fluorescent dye Fluo-4 AM (Invitrogen) and incubated in the dark for 1 h at room temperature. The cells were washed three times and incubated with 150 μl assay buffer for 30 min in the dark at room temperature. Fluorescence (measured as relative fluorescent units
(RFU) was recorded at room temperature on a FlexStation scanning fluorometer (Molecular Devices Inc.) at 485 nm excitation and 525 nm emission, with a 515 nm cut-off.

To test for mGluR agonism, baseline fluorescence was measured for 20 seconds, after which buffer (control), solvent (control), or HBACs were added and fluorescence was measured for an additional 5 min. RFU values were calculated by subtracting the mean baseline fluorescence from the mean fluorescence after treatment (see top graph in Fig. 4-6). To test for mGluR inverse agonism, antagonism, or positive allosteric modulation, baseline fluorescence was measured for 5 min following cell treatment with buffer (control), solvent (control), or HBACs, after which the receptors were stimulated with 100 μM l-glutamate and fluorescence was measured for an additional 2 min. RFU values were calculated by subtracting the mean baseline fluorescence from the mean fluorescence after glutamate treatment (see bottom graph in Fig. 4-6). GraphPad Prism was used to plot the RFU measurements. All experiments were performed in duplicate.
Appendix

Characterization of *C. elegans* Xenobiotic Metabolites by Mass Spectrometry
Compound 1

```
| ms and ms/ms of parent compound was not performed |
```

Parent (P)
240 Da

```
| Metabolite 1 (M1)
402 Da |
```

This metabolite was not digested by β-glucosidase, suggesting that it may be another hexoside such as an α-glucoside or mannoside.

```
| MS and MS-MS of parent compound was not performed |
```

Compound 2

```
| ms and ms/ms of parent compound was not performed |
```

Parent (P)
284 Da

```
| Metabolite 1 (M1)
446 Da |
```

This metabolite was verified to be a β-glucoside by β-glucosidase digestion.

**Figure A-1. Mass spectrometry analysis of xenobiotic metabolites.** For each compound, the structure of the parent molecule and the structure(s) of the metabolite(s), as determined from the MS analysis, are illustrated on the left-hand side of the diagrams. m/z plots are shown for those peaks that are specific to the metabolite fraction and are absent from the DMSO control. These m/z plots are shown for each metabolite, and the corresponding retention time (RT) for the peak is shown at the top of each plot. Relevant masses were selected for fragmentation by tandem MS-MS, and m/z plots of the fragment masses, and corresponding structures, are shown. For some compounds, MS and MS-MS analyses were performed for the parent molecule, and the corresponding m/z plots are shown. This figure is continued on the following four pages.
Compound 3

ms and ms/ms of parent compound was not performed

Compound 4

ms and ms/ms of parent compound was not performed

Figure A-1. MS analysis of xenobiotic metabolites continued...
Compound 5

Parent (P) 225 Da

Metabolite 1 (M1) 211 Da

Metabolite 2 (M2) 373 Da

This metabolite was verified to be a β-glucoside by β-glucosidase digestion.

*212 likely results from in-source fragmentation of 374

Figure A-1. MS analysis of xenobiotic metabolites continued...
The demethylation may occur at either methoxy group. An abundant mass of 269 (M₁₂₆₈ + H), specific to the metabolite fraction, was found at two different retention times, 30.730 and 31.541 minutes. The MS data for RT = 31.540 min is shown.

Two different retention times, 30.000 minutes. The MS data for RT = 27.947 min is shown.

The demethylation may occur at either methoxy group. An abundant mass of 447 (M₂₂₈₄ + H), specific to the metabolite fraction, was found at two different retention times, 27.947 and 29.281 minutes. The MS data for RT = 29.281, TOF MS ES+ Control MS was verified to be a β-glucoside by β-glucosidase digestion. The position of the glucose is not known. An abundant mass of 431 (M₂₂₈₄ + H), specific to the metabolite fraction, was found at two different retention times, 26.391 and 32.600 minutes. The MS data for RT = 26.391 min is shown.

This metabolite (and M₂₂₈₄, see above) was verified to be a β-glucoside by β-glucosidase digestion. The position of the glucose is not known. An abundant mass of 285 (P + 2) could also result from a demethylation plus a hydroxylation, however a fragment mass of 257 (see MS-MS data) would not be obtained for such a metabolite.

This metabolite (and M₂₂₈₄, see above) was verified to be a β-glucoside by β-glucosidase digestion. The position of the glucose is not known. An abundant mass of 447 (M₂₂₈₄ + H), specific to the metabolite fraction, was found at two different retention times, 27.947 and 30.000 minutes. The MS data for RT = 27.947 min is shown.

Figure A-1. MS analysis of xenobiotic metabolites continued...
Figure A-1. MS analysis of xenobiotic metabolites continued...
Additional Files

**Additional Data File 1.** XML file of the structure-based accumulation model script to be used with Pipeline Pilot software.

**Additional Data File 2.** SDF file of the small-molecule training set used to build the structure-based accumulation model.

**Additional Data File 3.** PDF file containing a list of open-source software that could be used to generate a similar structure-based accumulation model if Pipeline Pilot is not available.

**Additional Data File 4.** XLS file of the raw HT-HPLC small molecule accumulation data.
References


Click2Drug (http://www.click2drug.org/directory_ADMET.html).


Keiser, J., and Utzinger, J. (2010). The drugs we have and the drugs we need against major helminth infections. Adv Parasitol 73, 197-230.


WormBase (http://www.wormbase.org/).


