UNIVERSITY OF CALIFORNIA
RIVERSIDE

Chemical Genetics With a Synthesized Bioactive Small Molecule Library of Drug-Like Terminal Acetylenes in *Arabidopsis thaliana*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Andrew Joseph Defries

August 2013

Dissertation Committee:

Professor Julia Bailey-Serres, Co-Chairperson
Professor Michael Pirrung, Co-Chairperson
Professor Jason Stajich

The Dissertation of Andrew Joseph Defries is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my partner Anna Maria Huerta. Without you I would not have made it through the folds.

To the Mind's Eye we call Imagination.

ABSTRACT OF THE DISSERTATION

Chemical Genetics With a Synthesized Bioactive Small Molecule Library of Drug-Like Terminal Acetylenes in *Arabidopsis thaliana*

by

Andrew Joseph Defries

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, August 2013
Professor Julia Bailey-Serres, Co-Chairperson
Professor Michael Pirrung, Co-Chairperson

A formidable challenge in the genomic era is the assignment of structure and function to unknown genes in model organisms. Forward chemical genetics seeks to discover and characterize pathways through the identification of novel phenotypes induced by organic small molecules and subsequent isolation of their target gene(s). Downstream genetic and biochemical strategies to dissect the action of bioactive molecules are often hampered by weak target-ligand interactions, low target abundance, genetic factors, and synthetic inaccessibility of analogs. We present a synthetic scaffold that simultaneously permits fluorescence detection and photoaffinity labeling (diazirine) in a combinatorial library of thousands of drug-like small molecules synthesized through click chemistry. Using this library, we identified several bioactives that alter seedling growth of the model plant *Arabidopsis thaliana* and generated candidate ethyl methanesulfonate-induced resistant mutants for downstream simple sequence length polymorphism marker-assisted mapping and single nucleotide polymorphism identification by Illumina next-generation DNA sequencing. Our proof-of-principle study will facilitate target identification of novel bioactives and accelerate chemical genetics discoveries.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Abstract

Biological organisms are complex, and our predecessors have forged the path to characterize how disparate components sum to a living thing. We as a community bask in this knowledge, and may share, explore and expand scientific investigation. Nonetheless, we have achieved our current understanding by reducing the complexity of living systems to strings of DNA and translation products. At present select students of Chemistry and Biology are staring into the breech between these fields and are boldly attempting to join efforts to link genes with function in biological systems [1]. These efforts are often explored as the interactions between a protein and small molecules, and expanded if possible to explain the organism at large.

1.2 The technique and art of chemical genetics

In the field of chemical genetics the investigator utilizes a set of bioactive compounds to influence a biological system by inhibiting protein function on the cellular

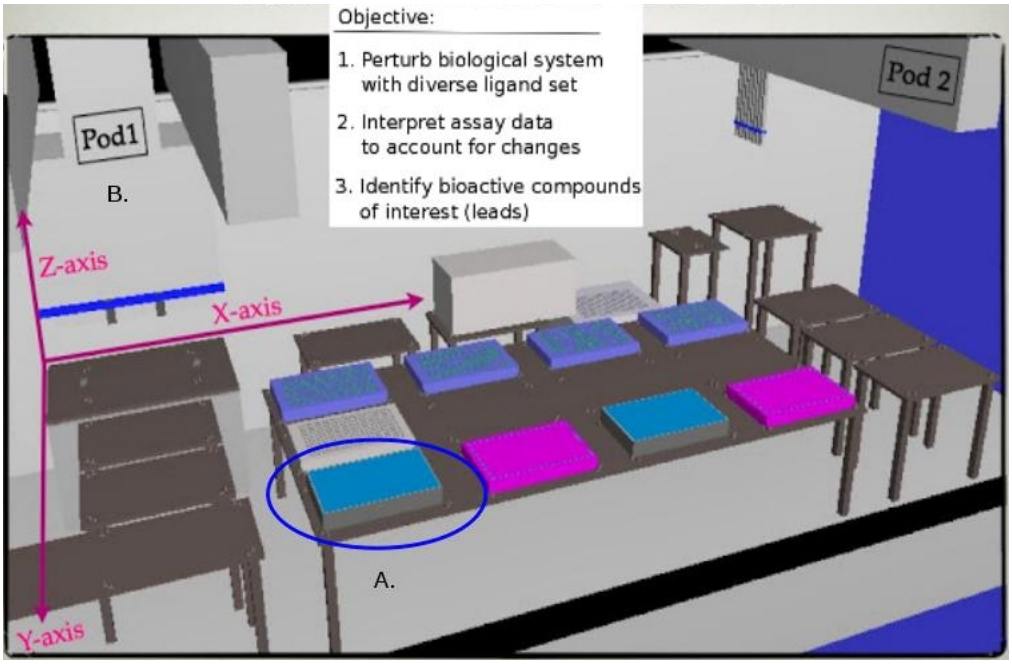

Figure 1.1: High throughput experiments are conducted using microwell plates and an automated robotic platform. A. Microwell plates of 96-well format contain compound libraries and are used to conduct assays. B. AP-96 multichannel fluid dispenser transfers compounds to the destination (shown as Pod1).

level [2]. This approach offers the convenience to tune dosage by titrating the concentration of the administered probe [3, 4], and is analogous to creating a knock out mutation or knock down or leaky mutation of an allele to study protein function [5]. A skillful investigator will use known and unknown probes to observe the phenotype spectrum or range of perturbation to gain insight into the molecular chain of events collectively referred to as mechanism of action (MOA) [6, 7]. High throughput assays were conducted with the use of microwell plates and robotics, and enabled experimental access to thousands of experiments to test novel compounds from large chemical libraries (Figure 1.1). We have used this platform to discover several novel terminal acetylene and fluorescent and tagged diazirine small molecules. In chemical genetics one may add probe at any time to the experimental system, which is form of temporal control [8]. I will demonstrate in Chapter 4, Phenotypes of Bioactives, how one may administer the

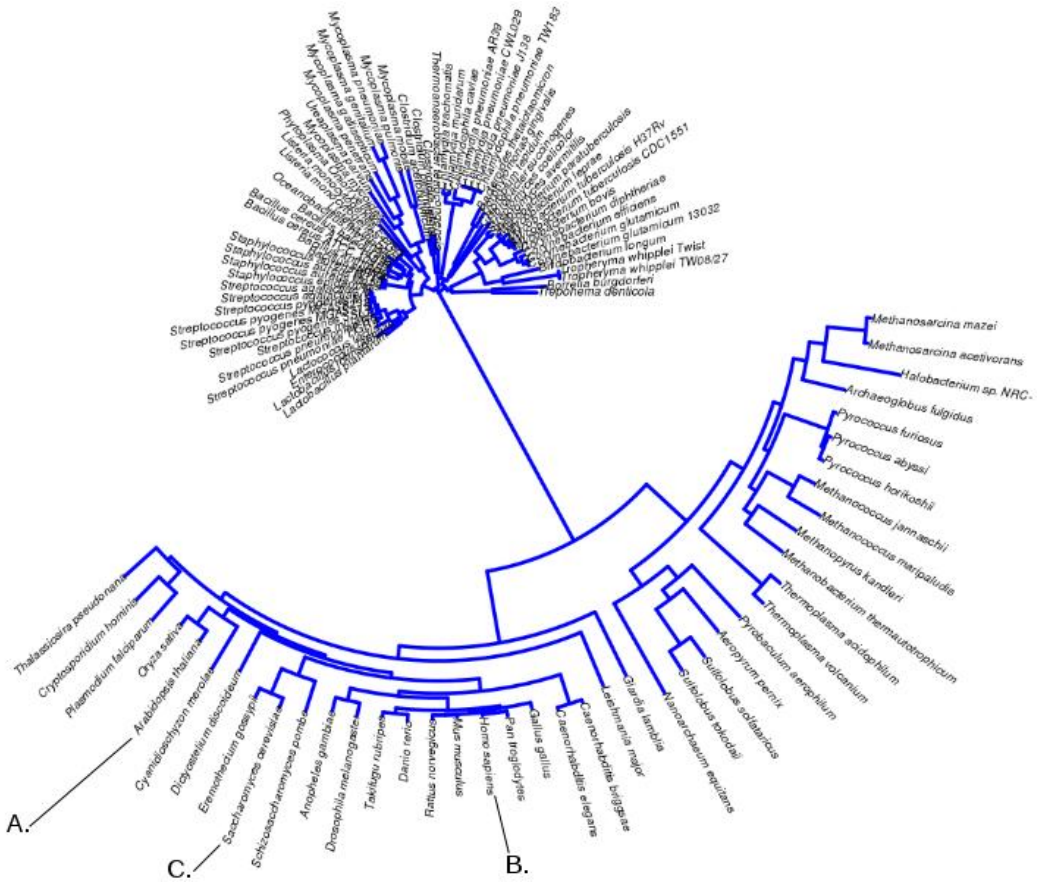

Figure 1.2: The model plant *Arabidopsis* is an experimental link to distant and related Eukarya. A. *Arabidopsis* is shown next to *Oryza sativa* branching from angiosperm plants long after the development of multicellularity. B. *Homo sapiens* has closely related organisms in animalia such as *Gallus gallus*, *Mus musculus*, and *Danio rerio*. C. Fungi has diverged early with *Saccharomyces cerevisiae* shown. Phylogenetic tree shown is unrooted and represent phylogenetic distances.

probe as a volatile without disturbing the experimental system.

1.3 The model plant *Arabidopsis thaliana*

Arabidopsis thaliana, a member of the *Brassicacea* is a model organism with a small sequences nuclear genome of 150 MB (Figure 1.2A)[9, 10]. A number of qualities deem *Arabidopsis* amenable to genetic mechanism of action studies: small size and short life cycle for ease of propagation, high progeny number to follow allele assortment, small genome to perform genetic studies approaching saturation rates, amenability to transformation [11], mutant resources [12], ease of controlled pollination, self-pollinator, widespread geographic natural variation [13]. *Arabidopsis* is a member of the large plant family *Brassicacea* [14], closely related to crops such as *Brassica oleracea* [15], *Brassica napus*, and *Brassica rapa* [16]. The *Brassica* or mustard family are angiosperms that provide an agricultural staple.

1.4 *Arabidopsis* for mechanism of action studies

Genetic redundancy that accumulated during the evolution of *Arabidopsis* was lost during the appearance of it A. thaliana [17]. *Arabidopsis* has a small genome with half the ploidy of the ancestor *Arabidopsis lyrata* [18], *Capsella* [19], and other *Brassicacea*. Functional redundancy [20] has been a fundamental problem [21] for the uncovering of conserved gene networks in a number of model organisms [22]. Therefore, *Arabidopsis thaliana* has been chosen, in this thesis, as the model organism for experimentation.

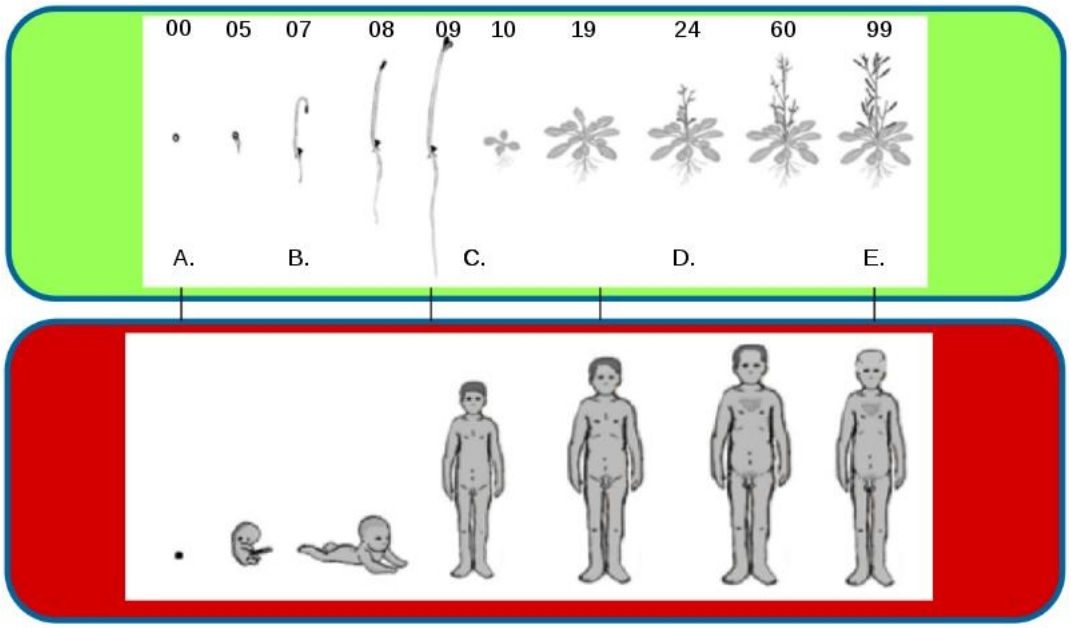

Figure 1.3: The life cycle of the model organisms enables us to translate biological understanding to more complex organisms. A-E. Fertilization, embryonic development juvenille development, and senescence are common developmental milestones between *Arabidopsis thaliana* and *Homo sapiens*. BBCH codes for the developmental stage (explained in Chapter 4: Phenotypes of Bioactives) are shown for *Arabidopsis* from 00-99.

1.5 *Arabidopsis*, an experimental link to distant and related Eukarya

The life cycle of *Arabidopsis* can be used to study developmental processes that have parallels in *Homo sapiens* (Figure 1.3A-E) [23]. For instance, both organisms undergo meiosis, sexual reproduction, fertilization, embryonic, juvenile, adult development, and senescence. Beyond these developmental similarities a comparison of the functional categories of *Arabidopsis* proteins compared to other sequenced genomes revealed 48-60% *Arabidopsis* proteins involved in proteins synthesis with counterparts in other species [24]. This is a reflection of highly conserved gene functions between plants and related organisms.

1.6 Chemical genetics in the model plant *Arabidopsis thaliana*

In the elucidation of genetic networks essential genes can cause embryonic lethal phenotypes. The technique of chemical genetics offers a means to ablate gene function after the passage of developmental milestones to expose a phenotype spectrum for examination [25] evidenced by recent findings [26]. A number of investigators have successfully employed simple target identification strategies for small molecules using classical genetic techniques. Thus, *Arabidopsis* is poised for use in chemical genetics research.

1.7 Vision of tagged chemical libraries

During the infancy of chemical genetics, what was considered by Mitchison and Schreiber as pharmacological genetics [27], it was a new milieu to use lead compounds in the same manner as genetic mutations to identify previously unknown targets. This was envisioned to be accomplished efficiently through covalent bond formation between ligand and receptor, an atypical situation. The concept was to produce combinatorial amine-like libraries that possessed functional groups to enhance downstream investigation of bioactives via affinity enrichment or covalent capture or both [27]. These methods were born out of the problems that arise in forward chemical genetics from novel small molecule probes that produce interesting phenotypes but require a high concentration to achieve 50% inhibition of the observed phenotype (IC50) in the range of 0.5 nM to 25 μM.

1.8 Activity based proteomic profiling, a method to create more tractable small molecule probes

Chemical genetics has a numerous problems that can complicate downstream investigation of bioactive leads [28]. Activity Based Proteomic Profiling (ABPP) has enhanced our odds of successful biochemical isolation of targets through the exploitation of irreversible inhibitors [1]. In ABPP the desired probe is bioactive either *in vivo* or *in vitro* and possesses reactive functional group(s) for a specific family of enzymes [29]. These probes are also thoughtfully designed to be tri-functional with a bioactivity group, a fluorescent or other detection group, and a group for biotin-avidin affinity chromatography techniques [30].

1.9 Comparison of probes in ABPP and forward chemical genetics

The type of probes used in ABPP are not of interest to the forward chemical genetics investigator unless they can be synthesized in large diverse libraries [31], screened *en masse* [32, 33] and stored. Small molecule probes containing reactive groups are not the end goal of forward genetics. This is why we seek, in this thesis, a probe that can be activated by ultraviolet (UV) light to produce a carbene *in situ*. It is our hope that through diazirine photolysis a carbene functionalized probe can be utilized for covalent labeling and facile downstream target investigation [34, 35]. We have devised a strategy to prepare and screen a combinatorial library to find this probe to promote the vision of pharmacological genetics. The design of the amine building block and use for generation of combinatorial libraries will be detailed (Chapter 2: Organic synthesis),

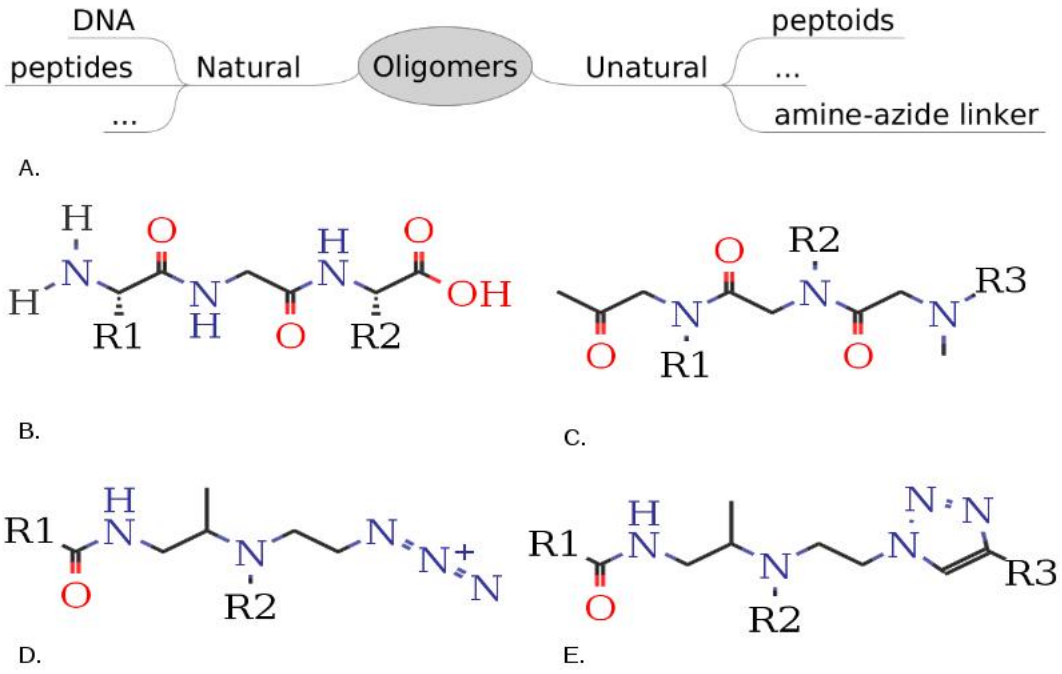

Figure 1.4: Chemical diversity in nature has inspired synthetic chemistry and combinatorial chemistry. A. A rich chemical diversity is possible from the oligomers DNA, peptides and proteins. B. The generic di-peptide shown here was formed from reaction with the amino or N-terminal and carboxy or C-terminal of two amino acids to produce the characteristic peptide backbone. C. Peptoids are a synthetic attempt to mimic a diverse peptide backbone and cap the reactive N and C-terminal ends. D. Our amine azide linker strategy is amenable to click chemistry. E. Introduction of diverse substituents to the backbone (Shown as R2 and R3) forms a peptide bond isostere, the triazole ring.

and biological use will also be detailed (Chapter 3: Chemical genetic screen).

1.10 Diverse chemical libraries from combinatorial reactions

Combinatorial libraries for chemical genetics have been inspired by the simplicity and diversity of biomolecules derived from natural amino acids (Figure 1.4A) [36]. Peptide combinatorial libraries innovated a number of high throughput solid phase techniques and paved the way for the synthesis of large libraries of peptide-like small molecules [37, 38]. Peptoids, unnatural oligomers from N-substituted glycine, have a similar side chain spacing as peptides but lack stereochemistry (Figure 1.4C)[37]. Our rationale was motivated by the fact that the formation of a 1,4 substituted 1,2,3-triazole (Figure 1.4E) was considered a peptide isostere [6] and enabled a wide range of reactants using a polytriazole catalyst [39] and reduced copper [40]. Therefore, we sought to design a small azide with a secondary amine for fluorophore conjugation, and an amine with a capped acetamide for combinatorial library generation using a reaction in water that is high yielding and regiospecific known as click chemistry (Figure 1.4D). Later bioactive leads may be functionalized into carbene generating Photo Affinity Ligands (PALs) using a boc-protected amine azide building block (synthesis discussed in detail in Chapter 2: Organic Synthesis and use discussed in detail in Chapter 6: Target identification efforts using covalent capture).

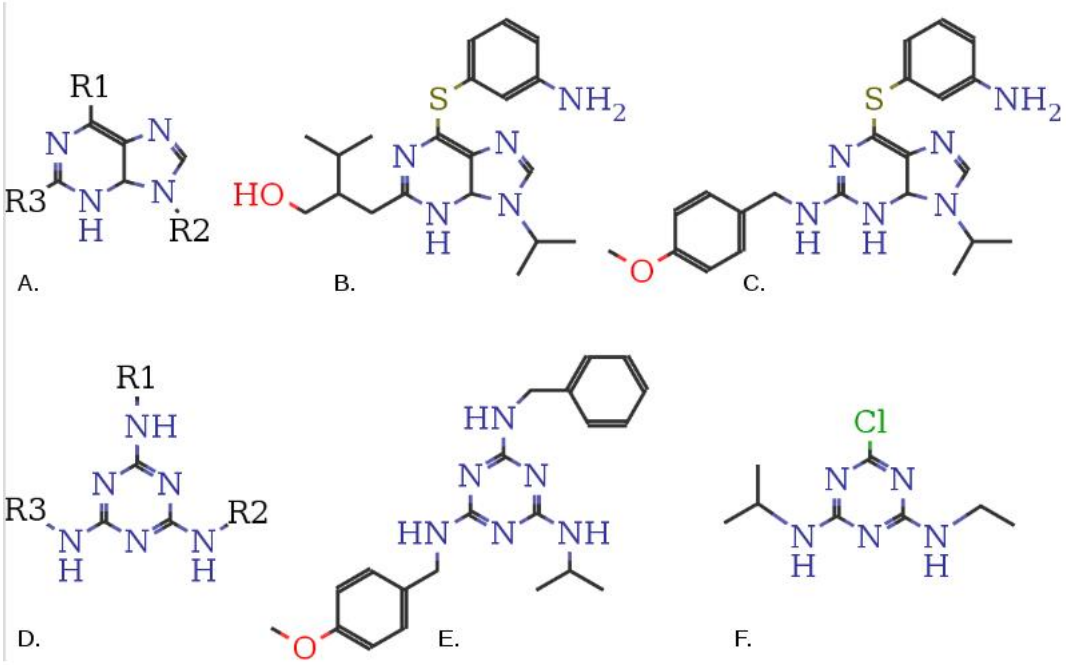

Figure 1.5: The purine scaffold was used to develop several combinatorial libraries. A. A purine ring scaffold shown with a number of diversity generation sites. B. Diminutol is a bioactive microtubule dynamics regulator. C. Myoseverin is also a microtubule binding compound. D. A triazine scaffold was used for the generation of related combinatorial libraries. E. Encephalazine is a bioactive in *Danio rerio* embryos. F. Atrazine is a popular herbicide.

1.11 Triazine tagged combinatorial libraries

The specific design of our amine-azide takes into account previous studies in forward chemical genetics with libraries from a purine scaffold (Figure 1.5A) [41]. Investigators made successive advancements with the purine scaffold to identify more novel bioactives (Figure 1.5A-C), and later learned the switch to a trizine core would reveal novel bioactives. These triazine compounds possessed groups for immobilization on an agarose support and were used to identify protein targets in the model organism *Danio rerio*, *in vitro* human cell lines [42, 43, 44], and in *Xenopus laevis* extracts [41] (Figure 1.2B, clade tip not shown).

The genius of the approach was to synthesize and screen libraries on the same scaffold with and without linkers. Therefore, investigators could clearly identify leads amenable to modification. The triazine scaffold was not sufficiently diverse or amenable for use in plant chemical genetics since it resembles the popular herbicide atrazine. Our chemistry platform capitalizes on commercially available precursors for amine block synthesis and widely available drug-like terminal acetylenes for library synthesis.

1.12 Conclusion

Our work was accomplished through the union of automation tools, cheminformatics, image analysis, spectroscopy, spectrometry, organic synthesis, pharmacology and genetics. These tools, examples, and bundled R code can be used to perform chemical genetics in *Arabidopsis* and offer inspiration for work in other model organisms.

Chapter 2

Organic Synthesis

2.1 Abstract

We envisioned a combinatorial chemistry strategy that enabled efficient attachment of diverse pharmacophores such as drug-like terminal acetylenes with fluorescent building block azides using an amine azide linker. The acetamide half of the amine azide linker was synthesized from glycine, and subjected to reductive amination to produce building block A. The attachment of fluorophores was achieved using standard amine coupling techniques with several flourophores. A second building block was synthesized that enables the derivitization of the hit molecule to a diazirine functionalized fluorescent probe. Furthermore, we synthesized a unique tris-[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) catalyst that enabled direct biological screening of combinatorial libraries synthesized in high throughput via the click reaction with the clickable library and fluorescent amine azide building blocks.

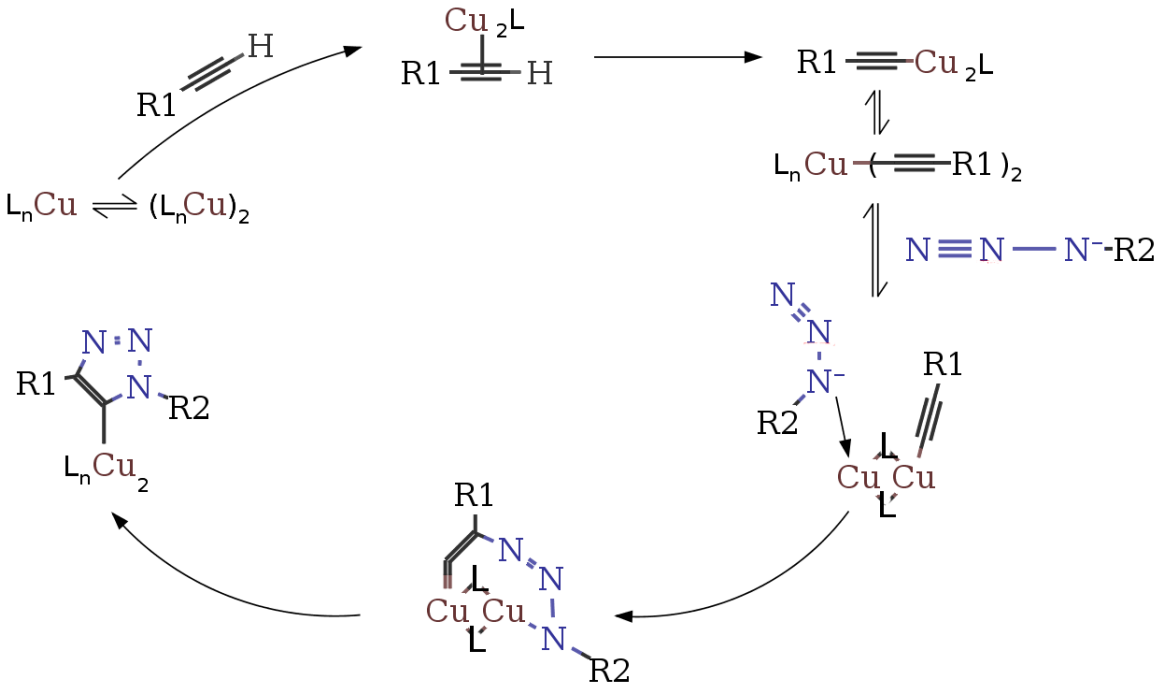

Figure 2.1: Copper catalyzed ligand mediated click chemistry.

2.2 Introduction

The azide-alkyne copper catalyzed ligand-mediated [39] Huisgen 1,3-dipolar cycloaddition known as click chemistry [45] creates products in high yield [40], with high regioselectivity, can be performed in water [45], and is a general reaction with azides and alkynes (Figure 2.1). In this variant of click chemistry a ligand and reduced copper coordinate with the acetylene and azide to catalyze the formation of the 1,2,3-triazole (Figure 2.1). Therefore, we chose click chemistry as our platform for high throughput synthesis. Our approach is feasible since reactions with acetylenes and azides were characterized as bioorthogonal [46], amenable to high throughput, and commercial availability of approximately 80,000 drug-like compounds with terminal acetylenes [47].

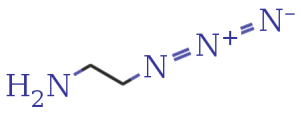

Figure 2.2: 2-azidoethan-1-amine.

2.3 Materials and Methods

2.3.1 2-azidoethan-1-amine

2-azidoethan-1-amine (Figure 2.2) was synthesized by dissolving sodium azide (16.8 g, 258 mmol) and 2-chloroethan-1-amine hydrochloride (10 g, 86 mmol) into water (86 ml) and heating for 15 hours at 80° C. The reaction was cooled to room temperature and KOH was added until the pH was 8.5. Thereafter, the product was extracted into ice cold diethyl ether (200 ml) from the aqueous solution, the organic was dried over anhydrous Na2SO4. Solvent was removed *in vacuo* to obtain a pure light yellow oil (3.7 g, yield 50%).

1HNMR (300 MHz, CDCl3): delta = 1.35-1.50 (br s, 2H), 2.89 (t, $J = 6$ Hz, 2H), 3.37 (t, it J = 6 Hz, 2H).

IR(neat):2090 cm-1 (azide stretching).

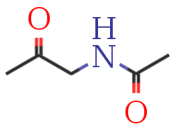

Figure 2.3: N-acetyl-N–(2-oxopropyl)acetamide.

2.3.2 N-acetyl-N-(2-oxopropyl)acetamide (Dakin-West).

N-acetyl-N–(2-oxopropyl)acetamide (Figure 2.3) was synthesized from glycine (75 g, 1.0 mol), acetic anhydride (1.1 L, 12 mol) and pyridine (485 ml, 6.0 mol) was refluxed overnight under positive pressure with a gentle stream of N2 bubbling from a needle at the bottom of the vessel or with a N2 balloon. Solvents were removed *in vacuo* to obtain a viscous black crude. Deacetylation of N-acetyl-N-(2-oxopropyl) acetamide to N-(2-oxopropyl)acetamide was accomplished through gentle reflux in excess methanol overnight. Volatiles were removed *in vacuo* to obtain a brown-red oil with a sweet smell. The residue was further purified via Kugelrorh distillation at 114° C under reduced pressure (0.7 Torr). The resulting white oil was purified by silica chromatography on a gradient elution of DCM:isopropanol from 98:2 to 90:10 in five steps (1 g, yeld 1%).

1HNMR (400 MHz, CDCl3): delta = 2.05 (s, 3H), 2.22 (s, 3H), 4.17 (d, $J = 4.2$ Hz, 2H), 6.19 (br s, 1H).

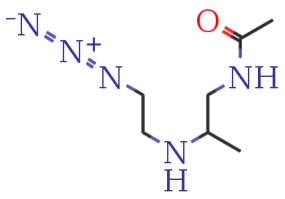

Figure 2.4: N-N-2[(2-azidoethyl)amino]propylacetamide.

2.3.3 N-N-2[(2-azidoethyl)amino]propylacetamide (block A)

N-N-2[(2-azidoethyl)amino]propylacetamide (block A) (Figure 2.4) was synthesized from 2-azidoethan-1-amine 1.6 g, 20 mmol), N-(2-oxopropyl)acetamide (2.3 g, 20 mmol) and LiCl (1.2 g, 30 mmol). The reactants were added to methanol (25 ml) and stirred for 1h before the addition of sodium cyanoborohyrdide (1.3 g, 20 mmol), and the reaction was left to proceed overnight. Solvent was removed *in vacuo* and the crude was washed with saturated aqueous NaHCO3, extracted into EtOAc, filtered over a cotton plug, dried with anhydrous Na2SO4, and evaporated *in vacuo*. The resultant crude product was dissolved in chloroform and filtered over a cotton plug to remove insoluble contaminants to obtain the crude product (1.4 g, yield 39%).

1HNMR (300 MHz, CDCl3): delta = 1.08 (d, J = 6.3 Hz, 3H), 1.99 (s, 3H), 2.69-2.84 (m, 2H), 2.91-3.05 (m, 2H), 3.33-3.43 (m, 3H), 6.08 (br s, 1H).

13CNMR (75 MHz, CDCl3): delta = 18.4. 23.0, 44.5, 45.9, 51.5, 52.4, 171.0.

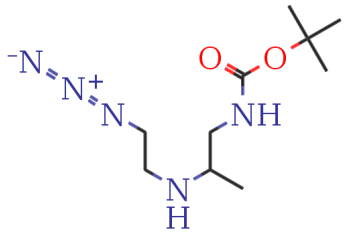

Figure 2.5: tert-butyl N-{2-[(2-azidoethyl amino]propryl}carbamate (block B).

2.3.4 Synthesis of tert-butyl N-{2-[(2-azidoethyl)amino]propyl)carbamate block B

Suvadeep Nath synthesized tert-butyl N-(2-oxopropyl)carbamate from an amino alcohol precursor and this was subjected by him to reductive amination similar to block A to produce tert-butyl N-{2-[(2-azidoethyl amino]propryl}carbamate (Figure 2.5), otherwise referred to as boc-protected block B (Figure 2.7). NMR analysis was performed by Andrew Defries.

1HNMR (300 MHz, CDCl3): delta = 1.07 (d, J = 6.3 Hz, 3H), 1.46 (s, 9H), 2.71-3.03 (m, 4H), 3.12-3.23 (m, 1H), 3.38-3.43 (m, 2H), 4.96 (br s, 1H).

13CNMR (75 MHz, CDCl3): delta = 18.6, 28.5, 45.6, 46.0, 51.9, 52.8, 79.2, 156.4.

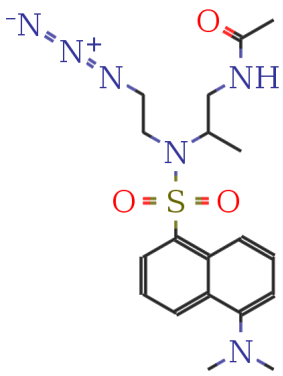

Figure 2.6: Dansyl block A.

2.3.5 Dansyl block A

Dansyl block A was synthesized ((Figure 2.6) by adding dansyl-Cl (5.0 g, 18 mmol), DIPEA (7.2 g, 56 mmol), a catalytic amount of DMAP, and block A (3.3 g, 18 mmol) in dry DCM under an N2 atmosphere. Volatiles were removed *in vacuo* and the residue was purified by SiO2 chromatography using a DCM:isopropanol gradient elution. Alternatively, the crude may be dissolved in 50:50 ACN:H2O and purified on a CombiHT-C8 preparative Reverse-Phase column and eluted using an isocratic method with 50:50 (solvent A 95% water, 5% ACN, 0.05 % formic acid and solvent B 95% ACN, 5% water, 0.05% formic acid) (300mg, yield 1%).

1HNMR (500 MHz, CDCl3) delta = 0.96 (d, $J = 6.5$ Hz, 3H), 1.64 (s, 3H), 2.89 (s, 6H), 3.01 (m, 1H), 3.22 (m, 1H), 3.38 (m, 1H), 3.54 (m, 2H), 3.69 (m, 1H), 3.89 (m, 1H), 5.94 (br s, 1H), 7.21 (d, $J = 7$ Hz, 1H), 7.56 (dd, $J = 1$ Hz, $J = 7.5$ Hz, 1H), 7.62 (dd, $J = 1$ Hz, $J = 7.5$ Hz, 1H), 8.27 (d, $J = 9$ Hz, 1H), 8.30 (dd, $J = 1$ Hz, $J = 7.5$ Hz, 1H), 8.59 (d, $J = 8.5$ Hz, 1H).

m/z of 419.1866 was obtained for dansyl block A corresponding to (M+H) for the expected molecular formula C19H26N6O3S.

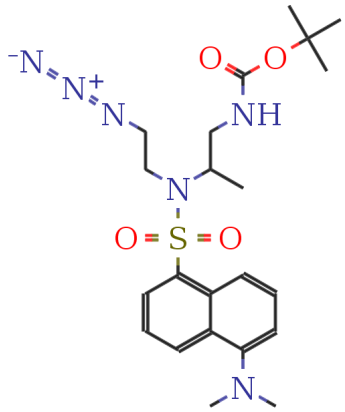

Figure 2.7: Dansyl boc block B.

2.3.6 Dansyl boc block B

Dansyl block B (Figure 2.7) was synthesized by adding dansyl-Cl (5.0 g, 18 mmol), DIPEA (7.2 g, 56 mmol), a catalytic amount of DMAP, and block B (4.4 g, 18 mmol) in dry DCM under N2 atmosphere (Figure S14). Volatiles were removed *in vacuo* and the residue was purified similarly to block A dansyl on an isocratic method with 30:70 (solvent A: solvent B) as eluent.

1HNMR (400 MHz, CDCl3): delta = 0.92 (d, J = 6.8 Hz, 3H), 1.42 (s, 9H), 2.90 (s, 6H), 3.0-3.06 (m, 2H), 3.10-3.30 (m, 2H), 3.44-3.52 (m, 2H), 3.90-4.1 (m, 1H), 4.99 (br s, 1H), 7.20 (d, J = 8 Hz, 1H), 7.54 (t, J = 7.6 Hz, 1H), 7.60 (t, it J = 8.4 Hz, 1H), 8.27-8.31 (m, 2H), 8.57 (d, J = 8.4 Hz, 1H).

m/z of 477.2268 was obtained corresponding to (M+H)+ for the expected molecular formula C22H32N6O4S.

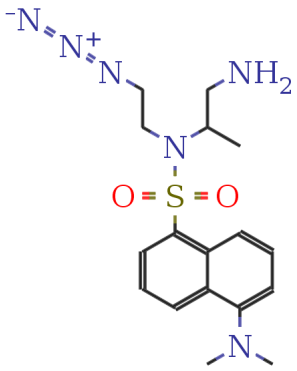

Figure 2.8: Dansyl free amine block B.

2.3.7 Dansyl free amine block B

The resulting boc precursor which was de-protected (Figure 2.8) using 10% TFA in DCM at room temp for 1h. Volatiles were removed *in vacuo* and the free amine was isolated by preparative RP-HPLC.

1HNMR (400 MHz, D6-DMSO): delta = 0.84 (d $J = 6.4$ Hz, 3H), 2.85 (s, 6H), 2.92-2.95 (m, 1H), 3.29-3.34 (m, 1H), 3.47-3.61 (m, 4H), 4.08-4.14 (m, 1H), 7.32 (d, $J = 8$ Hz, 1H), 7.63-7.70 (m, 2H), 8.22 (d, $J = 6.4$ Hz, 1H), 8.30 (d, $J = 8.4$ Hz, 1H), 8.54 (d, $J = 8$ Hz, 1H).

13CNMR (100 MHz, D6-DMSO): delta = 16.30, 22.42, 38.16, 42.90, 45.10, 45.75, 49.94, 50.40, 52.69, 114.07, 116.97, 119.05, 123.75, 128.25, 129.18, 130.00, 135.05, 151.08, 158.30, 169.40.

m/z of 377.1757 was obtained corresponding to (M+H)+ for the molecular formula C17H24N6O2S.

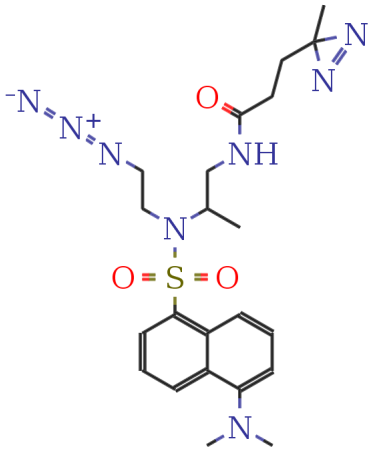

Figure 2.9: Dansyl diazirine.

2.3.8 Synthesis of diazirine dansyl block B (DDA)

Equimolar amounts of (0.22 mmol) SDA-Diazirine (Pierce) and dansyl block B free amine (83 mg, 0.22 mmol) were dissolved in a minimum of DMSO and diluted with DCM/MeOH containing DIPEA. The reaction was conduced overnight and the product was purified by RP-HPLC using an 80% isocratic method using ACN/H2O) as eluent to give the product dansyl diazirine block B (Figure 2.9)(40 mg, 50% yield).

1HNMR (500 MHz, D6-DMSO): delta = 0.93 (d, $J = 7$ Hz, 3H), 1.56 (s, 3H), 2.84 (s, 6H), 3.08 (m, 2H), 3.3-3.5 (m, 9H), 3.95 (m, 1H), 7.27 (d, $J = 7.5$ Hz, 1H), 7.61-7.66 (m, 2H), 8.15 (d, $J = 9$ Hz, 1H), 8.18 (dd, $J = 1$ Hz, $J = 7.5$ Hz, 1H), 8.50 (d, $J = 8.5$ Hz, 1H).

13CNMR (126 MHz, D6-DMSO): delta = 16.26, 22.32, 39.99, 41.71, 42.05, 45.03, 50.38, 52.65, 115.21, 118.74, 123.65, 128.23, 129.21, 129.76, 130.16, 134.96, 151.43, 169.21.

m/z of 487.2234 was obtained corresponding to (M+H)+ for the expected molecular formula C22H30N8O3S.

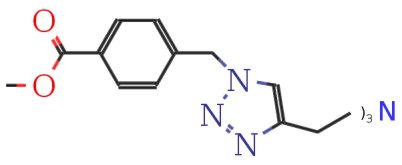

Figure 2.10: TBTA-(CO2Me)3.

2.3.9 TBTA-(CO2Me)3

Synthesis of TBTA-(CO2Me)3 was performed by Suvadeep Nath (Figure 2.10). NMR analysis was performed by Andrew Defries.

1HNMR (400 MHz, D6-DMSO): delta = 3.32 (s, 9H), 3.65 (s, 6H), 5.70 (s, 6H), 7.37 (d, J = 8.8 Hz, 6H), 7.94 (d, J = 7.6 Hz, 6H), 8.13 (s, 3H).

13CNMR (100 MHz, CDCl3): delta = 47.68, 52.8, 52.9, 125.2, 128.5, 129.9, 130.3, 142.2, 144.5, 166.5.

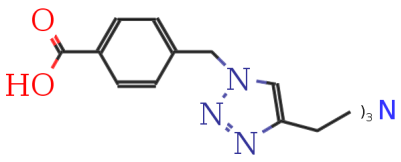

Figure 2.11: TBTA-(CO2H)3.

2.3.10 TBTA-(CO2H)3

Synthesis of TBTA-(CO2H)3 was performed by Suvadeep Nath (Figure 2.11). NMR analysis was performed by Andrew Defries.

1HNMR (300 MHz, D6-DMSO): delta = 3.31 (br s, 3H), 3.65 (s, 6H), 5.67 (s, 6H), 7.34 (d, $J = 8.4$ Hz, 6H), 7.91 (d, $J = 7.8$ Hz, 6H(, 8.12 (s, 3H).

13CNMR (100 MHz, CDCl3): delta = 47.7, 53.0, 125.15, 128.4, 130.4, 131.3, 141.6, 144.4, 167.6.

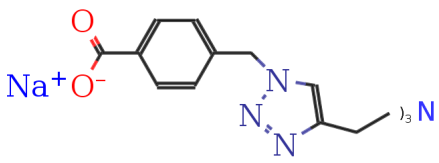

Figure 2.12: TBTA-(CO2Na)3.

2.3.11 TBTA-(CO2Na)3

Synthesis of TBTA-(CO2Na)3 was performed by Suvadeep Nath (Figure 2.12). NMR analysis was performed by Andrew Defries.

1HNMR (400 MHz, D20): delta = 3.67, (s, 6H), 5.33 (s, 6H), 7.08 (d, $J = 8.0$ Hz, 6H), 7.64 (s, 3H), 7.67 (d, $J = 8.4$ Hz, 6H).

13CNMR (100 MHz, D20): delta = 48.2, 53.4, 125.3, 127.7, 129.6, 136.6, 137.7, 144.1, 174.8.

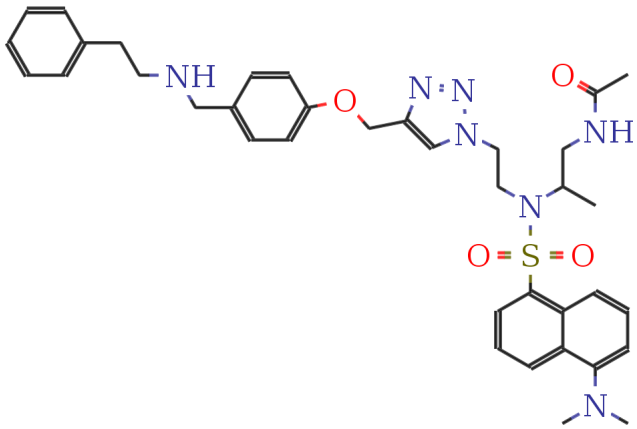

Figure 2.13: CLK021C05 dansyl block A

2.3.12 CLK021C05 dansyl block A

Synthesis of CLK021C05 dansyl bock A (Figure 2.13) and the subsequent compounds were synthesized using standard click chemistry conditions using reduced copper and TBTA-(CO2Na)3 with the respective azide building block and terminal acetylene from the clickable collection.

1HNMR (400 MHz, CD3OD): delta = 0.87 (d, J = 6.8 Hz, 3H), 1.66 (s, 3H), 2.63 (t, 8H), 2.84 (s, 6H), 2.92-3.19 (m, 6H), 5.12 (s, 2H), 7.06 (d, J = 9.2 Hz, 1H), 7.23-7.31 (m, 7H), 7.37 (d, J = 8.4 Hz, 2H), 7.55-7.62 (m, 2H), 7.85 (s, 1H), 8.23 (d, J = 8.8 Hz, 1H), 8.26 (dd, J = 7.6 Hz, 1H), 8.52 (br s, 1H), 8.59 (d, J = 8.4 Hz, 1H).

m/z of 706.3146 was obtained corresponding to the (M+Na)+ for the expected molecular formula C37H45N7O4S.

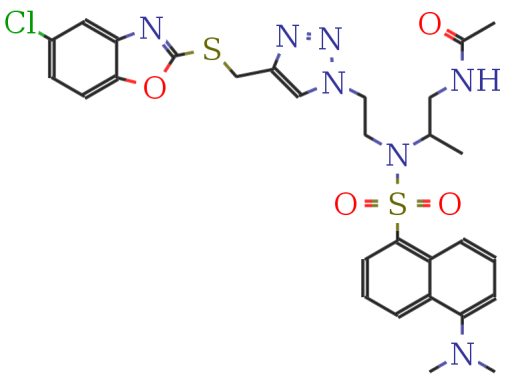

Figure 2.14: CLK024F02 dansyl block A.

2.3.13 CLK024F02 dansyl block A

(Figure 2.14)

1HNMR (400 MHz, D6-DMSO): delta = 0.83 (d, $J = 6.8$ Hz, 3H), 1.53 (s, 3H), 2.54 (s, 1H), 2.83 (s, 6H), 3.00 (t, $J = 7.6$ Hz, 2H), 3.60-3.64 (m, 2H), 3.90 (m, 1H), 4.50 (t, $J = 7.2$ Hz, 2H), 4.64 (s, 2H), 7.26 (d, $J = 7.6$ Hz, 1H), 7.37 (dd, $J = 2.0$ Hz, $J = 8.8$ Hz, 1H), 7.57-7.65 (m, 2H), 7.70 (d, $J = 8.8$ Hz, 1H), 7.77 (d, $J = 2$ Hz, 1H), 8.10 (s, 1H), 8.14 (d, $J = 9.2$ Hz, 1H), 8.17 (dd, $J = 1.2$ Hz, $J = 7.6$ Hz, 1H), 8.50 (d, $J = 8.4$ Hz, 1H).

m/z of 642.1698 was obtained corresponding to the (M+H) for the expected molecular formula C29H32ClN7O4S2.

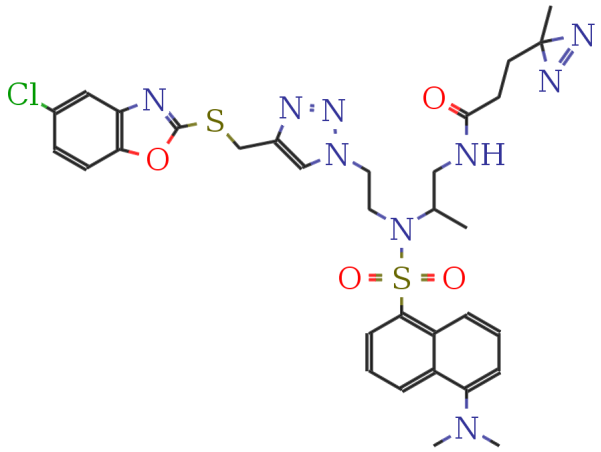

Figure 2.15: CLK024F02 dansyl diazirine.

2.3.14 CLK024F02 dansyl diazirine

(Figure 2.15)

1HNMR (400 MHz, CDCl3): delta = 0.84 (dm J = 7.2 Hz, 3H), 1.55-1.80 (m, 7H), 2.91 (s, 6H), 2.95-3.10 (m, 1H), 3.18-3.25 (m, 1H), 3.37-3.48 (m, 1H), 3.49-3.60 (m, 2H), 3.60-3.76 (m, 1H), 3.85-4.00 (m, 1H), 4.59 (s, 2H), 7.20 (d, J = 7.2 Hz, 1H), 7.24 (dd, J = 6.4 Hz, J = 2.0 Hz, 1H), 7.37 (d, J = 4.0 Hz, 1H), 7.40 (d, J = 3.6 Hz, 1H), 7. 52-7.60 (m, 3H), 7.64 (d, J = 2 Hz, 1H), 8.14 (d, J = 8.4 Hz, 1H), 8.25 (dd, J = 7.6 Hz, J = 0.8 Hz, 1H), 8.58 (d, J = 8.4 Hz, 1H).

m/z of 709.2020 was obtained corresponding to a loss of N2, CH3, and 2H for the expected molecular formula C32H36ClN9O4S2.

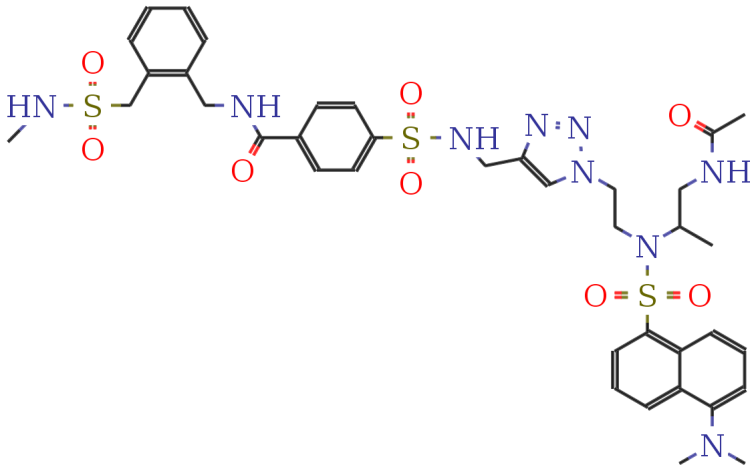

Figure 2.16: CLK042A09 dansyl block A.

2.3.15 CLK042A09 dansyl block A

(Figure 2.16)

1HNMR (400 MHz, D6-DMSO): delta = 0.92 (d, J = 6.8 Hz, 3H), 1.55 (s, 3H), 2.62 (d, J = 4.8 Hz, 2H), 2.82 (s, 6H), 3.01-3.05 (m, 2H), 3.17 (s, 3H), 3.33 (s, 2H), 3.56-3.61 (m, 2H), 3.92-3.97 (m, 1H), 4.57 (s, 2H), 4.64 (d, J = 5.6 Hz, 2H), 7.65 (q, J = 9.6 Hz, J = 4.4 Hz, 3H), 7.24-7.41 (m, 5H), 7.58-7.69 (m, 2H), 7.96 (dd, J = 66 Hz, J = 8.8 Hz, 4H), 8.14 (d, J = 8.8 Hz, 1H), 8.18 (dd, J = 7.2 Hz, J = 1.2 Hz, 1H), 8.26 (br s, 1H), 8.50 (d, J = 8.4 Hz, 1H), 9.20 (t, J = 11.6 Hz, 1H).

m/z of 854.2782 was obtained corresponding to (M+H)+ for the expected molecular formula C38H47N9O8S3.

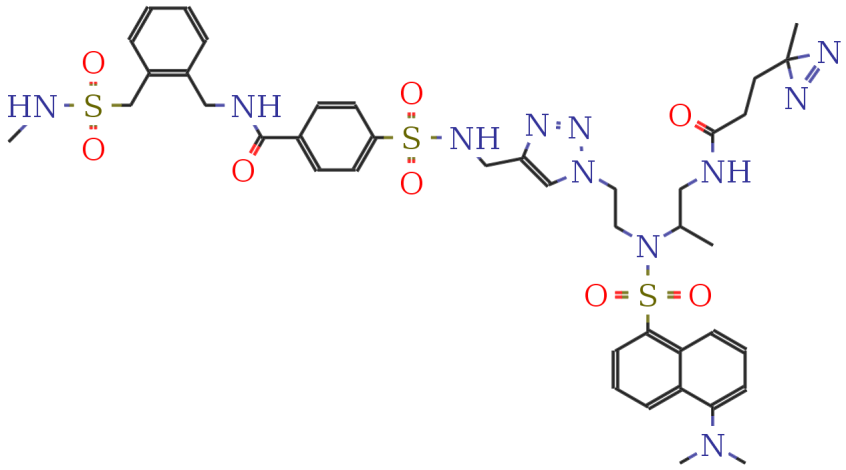

Figure 2.17: CLK042A09 dansyl diazirine.

2.3.16 CLK042A09 dansyl diazirine

(Figure 2.17)

1HNMR (300 MHz, CDCl3): delta = 0.83 (d, $J = 6.6$ Hz, 3H), 1.57 (s, 3H), 1.60-1.90 (m, 4H), 2.84 (d, $J = 5.4$ Hz, 2H), 2.91 (s, 6H), 3.00-3.10 (m, 2H), 3.45-3.58 (m, 2H), 3.62-3.75 (m, 2H), 3.84-4.23 (m, 2H), 4.49 (s, 2H), 4.73 (d, $J = 5.1$ Hz, 3H), 5.81 (s, 1H), 6.07 (t, $J = 9.3$ Hz, 1H), 7.11 (s, 1H), 7.23 (d, $J = 7.8$ Hz, 1H), 7.32-7.41 (m, 4H), 7.51-7.64 (m, 4H), 7.60 (t, $J = 11$ Hz, 1H), 7.88 (dd, $J = 22.5$ Hz, $J = 8.4$ Hz, 4H), 8.18 (d, $J = 8.7$ Hz, 1H), 8.24 (d, $J = 7.2$ Hz, 1H), 8.60 (d, $J = 8.7$ Hz, 1H).

m/z of 876.2633 was obtained corresponding to a loss of N2, CH3, and 2H for the expected molecular formula C41H51N11O8S3.

2.3.17 *In sillico* selection of the clickable library

Commercially available compounds from emolecules.com were downloaded and searched using ChemmineR for compounds containing a connection matrix (conMa) indicating the presence of a terminal acetylene functional group, this subset was further reduced using a drug-like filter.

The emolecules database is pay per use, therefore for demonstration purposes we will detail equivalent operations on the freely accessible ZINC collection of 2.0 million compounds. The ZINC collection was batch downloaded using wget and uncompressed from sdf.gz and searched using the sdfstream function of ChemmineR. The sdfstream was performed to prepare a descriptor matrix of compounds in the ZINC collection, including the field indicating the presences of a terminal acetylene. Compounds containing terminal acetylenes were subset and denoted as ZINC80K in the figures. This subset can be filtered further using physiochemical features provided by ChemmineR, with extra features provided by Open Babel such as logP and Total Polar Surface Area (TPSA).

2.3.17.1 R code for selection of terminal acetylene subset from ZINC purchasible collection

```
Bash code:
##########
wget -i list
# where the list was provided for server path of
# *.sdf.gz from zinc.docking.org/subsets/all-purchasable
# http://zinc.docking.org/db/bysubset/6/*.sdf.gz

R code:
##################
library(ChemmineR)
##################

#define function to collect descriptors
```

```
desc <- function(sdfset) {
        cbind(SDFID=sdfid(sdfset),
              datablock2ma(datablocklist=datablock(sdfset)),
              MW=MW(sdfset),
              groups(sdfset),
              AP=sdf2ap(sdfset, type="character"),
              rings(sdfset, type="count", upper=6, arom=TRUE)
        )
}

#execute sdfstream
sdfStream(input="ZINC_Purchasible.sdf", output="ZINC_Purchasible.xls", ...
append=FALSE, fct=desc, Nlines=1000)

#subset the sdf for compounds with terminal acetylenes
indexDF <- read.delim("ZINC_Purchasible.xls", row.names=1)
indexDFsub <- indexDF[indexDF$RCCH >= 1, ]

#subset source sdf with indexDFsub containing only the rows for compounds ...
with RCCH>=1

read.SDFindex(file="ZINC_Purchasible.sdf", index=indexDFsub, type="file", ...
outfile="ZINC_80K.sdf")

#This will write all molecules with RCCH>=1 to an SD file called "ZINC_80K.sdf".
#Help and comments provided by Thomas Girke
```

2.3.18 Subsetting ZINC purchasible

The clickable collection was selected from ZINC80K by applying a combination of unique rule of five filters, shown using Open Babel, below. Alternatively, one may use physicochemical properties provided by ChemmineR shown in the next sections.

```
Bash code:
##########
babel ZINC_80K.sdf -osmi --filter "MW<450 logP > 5"
```

2.3.19 Clustering and 3D-multidimensional scaling (3D-MDS) visualization of the clickable collection

```
R code:
##################
library(ChemmineR)
##################
#Clickable.sdf provided as a record for the 4,002 compounds in the ...
#clickable collection of terminal acetylenes

#load sdfset and create apset for clustering
sdfset <- read.SDFset("Clickable.sdf")
apset <- sdf2ap(sdfset)

#cluster apset
clusters <- cmp.cluster(apset, cutoff = c(0.7))

#embed clusters in a 3-dimensional space based on apset
coord <- cluster.visualize(apset, clusters, size.cutoff=1, dimensions=3,...
quiet=TRUE)

#set ranges for colors
coord_alpha <- coord[1:2769,1:3]
coord_beta <- coord[2770:4002,1:3]

#############
library(rgl)
#############
rgl.open(); offset <- 50; par3d(windowRect=c(offset, offset, 640+offset, ...
640+offset))
rm(offset); rgl.clear(); rgl.viewpoint(theta=45, phi=30, fov=60, zoom=1)
spheres3d(coord_alpha[,1], coord_alpha[,2], coord_alpha[,3], radius=0.005, ...
color="black", alpha=1, shininess=20); aspect3d(1, 1, 1)
spheres3d(coord_beta[,1], coord_beta[,2], coord_beta[,3], radius=0.005, ...
color="red", alpha=1, shininess=20); aspect3d(1, 1, 1)
```

```
axes3d(col='black'); title3d("", "", "", "", "", col='black'); bg3d("white")
############
rgl.snapshot("coord_alphaVsbeta.png")
############
rgl.close()
############
```

2.3.20 Generation of phyicochemical descriptors for SDF with Open Babel and ChemmineR

```
Bash code:
##########
babel ZINC_80K.sdf -otxt --append TPSA > ZINC_80K.TPSA
babel ZINC_80K.sdf -otxt --append logP > ZINC_80K.logP

R code:
##################
library(ChemmineR)
##################

propma <- data.frame(MF=MF(sdfset, addH=FALSE), MW=MW(sdfset, addH=FALSE), ...
Ncharges=sapply(bonds(sdfset, type="charge"), length), atomcountMA(sdfset, ...
addH=FALSE), groups(sdfset, type="countMA"), rings(sdfset, upper=6, ...
type="count", arom=TRUE))

#add data from GetMoreProps.sh
More<-read.table(props[a])
names(More)<-c("TPSA","logP")
saved<-names(propma)
propma<-cbind(propma,More$TPSA, More$logP)
names(propma)<-c(saved,c("TPSA","logP"))
write.csv(propma, file=csv_name)

# propma is then sent for second treatment
######################
files <- list.files(recursive=TRUE, pattern = ".csv")
######################
par(mfrow=c(2,3))
######################
DoThis <- function(a){
######################
png_name<-paste(gsub(".csv","",files[a]),".png",sep="")
######################
png(file=png_name)
######################
MW <- read.csv(files[a])$MW
TPSA <- read.csv(files[a])$TPSA
logP <- read.csv(files[a])$logP
######################
library(scatterplot3d)
######################
scatterplot3d(MW,TPSA,logP, xlim=c(0,2000), ylim=c(0,800), ...
zlim=c(0,20), main=gsub(".csv","",files[a]))
######################
dev.off()
```

```
######################
}
######################
#a<-1:length(files)
a<-c(1,2,6,4,3,5)
lapply(a,DoThis)
######################
```

2.3.21 Physical assembly of the clickable library

Click chemistry was chosen as the platform for our tagged library generation. Therefore, we assembled a diverse drug-like terminal acetylene library from vendors Asinex, Chembridge, Life Chemicals, Vitas-M, and Enamine. In total, 4,002 compounds were purchased for the clickable library. The unique molecular weight of each compound was used to add a corresponding volume of diluent, dimethylsulfoxide (DMSO), to each vile to a final concentration of 10 mM or 20 mM (for high molecular weight compounds). Dilution of the neat powders was automated using a Beckman -Coulter Biomex FXp liquid handler with a span-8 pod. Volume information was provided by using a comma separated value (csv) spreadsheet which determined the unique span-8 probe volume.

2.3.22 *In sillico* synthesis of combinatorial libraries from azides and terminal acetylenes

The program Reactor by ChemAxon was used to perform *in sillico* click reactions with terminal acetylenes and azides in .sdf or .mrv formats. A windows batch script written by Daniel Swank was made to perform the click reaction in a nested for loop, thereby creating a combinatorial library from all reactants. Two reaction files specifying the TBTA or Ruthenium catalyzed click reaction.

Batch script:

```
@echo off

echo ==========================
echo = Directory Reactor v0.1 =
echo ==========================


set reactcnt=0
set azidcnt=0
set alkcnt=0
set current=1
for %%A in (Reaction/*) do set /a reactcnt+=1
for %%A in (Azide/*) do set /a azidcnt+=1
for %%A in (Alkyne/*) do set /a alkcnt+=1
set /a count=%reactcnt% * %azidcnt% * %alkcnt%

FOR /f %%1 IN ('dir /b Reaction') DO (
FOR /f %%2 IN ('dir /b Azide') DO (
FOR /f %%3 IN ('dir /b Alkyne') DO (
set /a current+=1
echo Combining: %%~n2 + %%~n3 - %%~n1 [%current% of %count%]
call react -r Reaction/%%1 -m comb Azide/%%2 Alkyne/%%3 -f sdf -o ...
Output/%%~n2-%%~n3-%%~n1.sdf -i R1 -R Clk_ID -P PID -k
)
)
)
pause
```

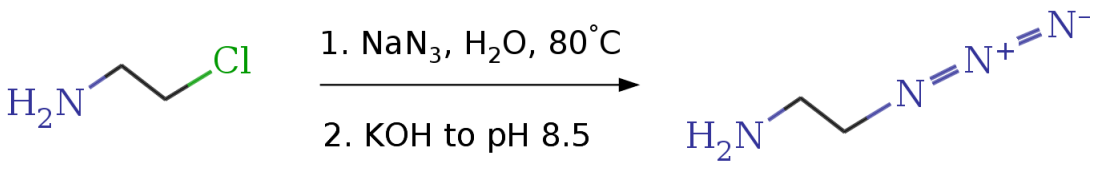

Figure 2.18: Synthesis of 2-azidoethan-1-amine from 2-chloroethan-1-amine

2.3.23 Generation of fluorescent tagged libraries through click chemistry

10 μl of 10 mM terminal acetylene was aliquotted to an empty 96-well polypropylene plate with 1.3 μl of 100 mM dye-azide. Thereafter, a 5 μl mixture derived from a 1:1:1 ratio of 10 mM TBTA-(CO2Na)3, 10 mM CuSO4, and 100 mM Na-ascorbate was added. Mixing was achieved by centrifugation followed by shaking overnight. Plates were allowed to incubate for three days prior to screening.

2.4 Results

2.4.1 Synthesis of 2-azidoethan-1-amine

The synthesis of 2-azidoethan-1-amine was straightforward as described by literature, accomplished through azidolysis of 2-chloroethan-1-amine in water with some heat (Figure 2.18). 1HNMR data confirmed the presence of expected groups on the amine, with broad peaks at 1.43 ppm, and two expected triplets for alkyl CH2 groups. Note the CH2 group adjacent to the installed azide shifted downfield when compared to the precursor [48].

Extraction of the product from the aqueous reaction mixture was optimized by titrating HCl produced during azidolysis of 2-chloroethan-1-amine with equimolar KOH, or until the pH was 8.5. The product was a volatile oil of high toxicity, and

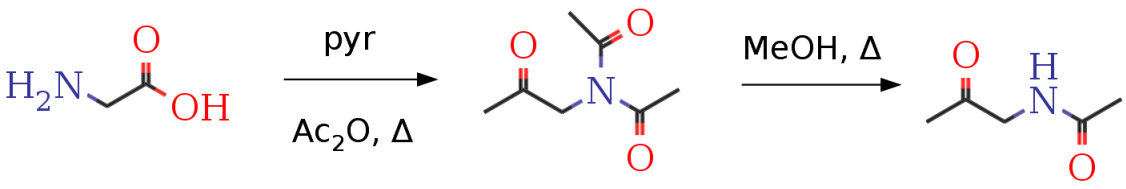

Figure 2.19: Conducting the Dakin-West reaction with glycine.

required extraction from the aqueous layer with ice cold diethyl ether; in addition, ice was used in the rotovap bath. A number of failed or low yield reactions resulted from the volatility of 2-azidoethan-1-amine. This product was used directly when possible, and decomposed readily. One may extend the shelf life of the oil by storing under a wrapped stopcock at 4 degrees in a vessel with a large head space.

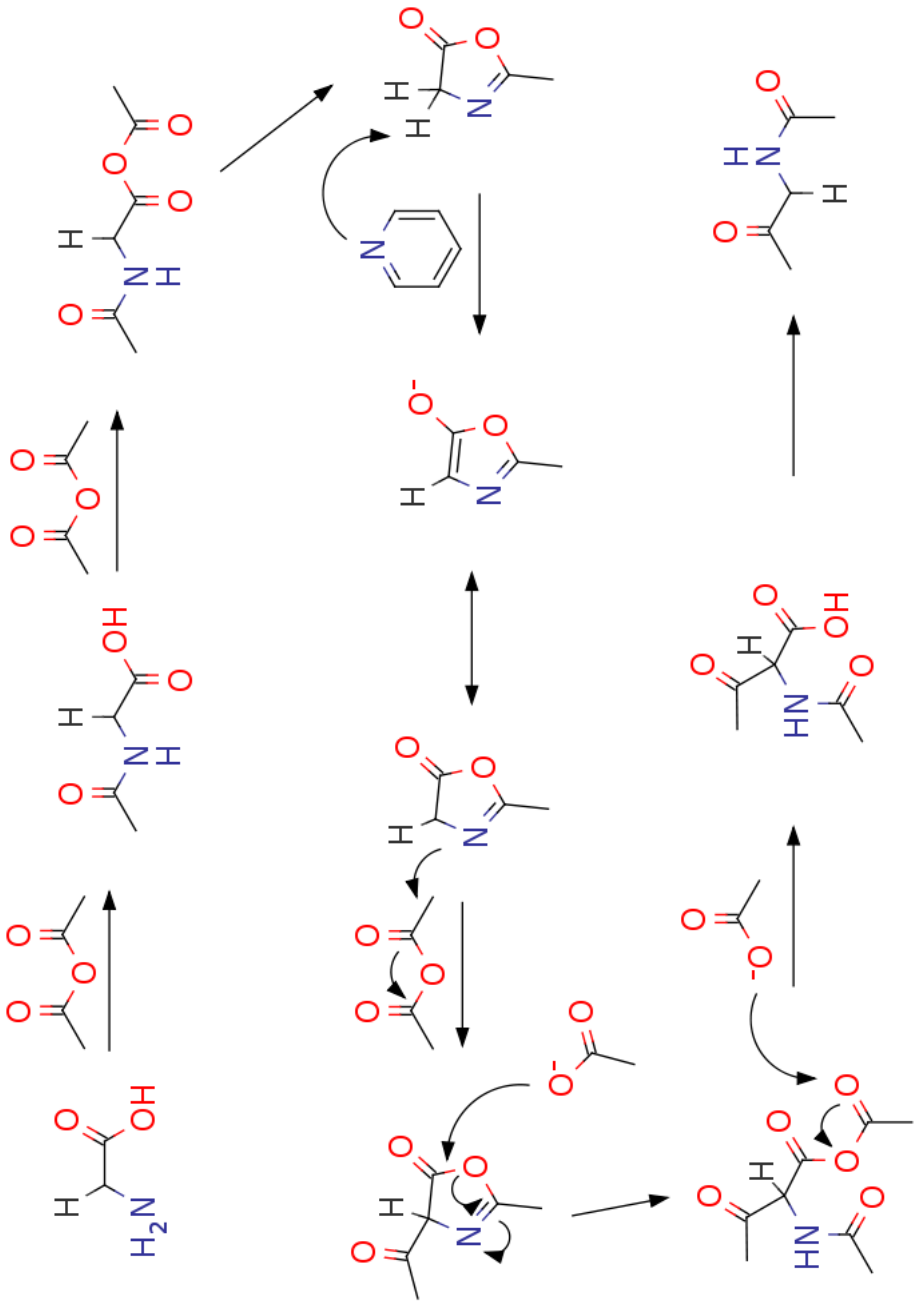

Figure 2.20: Mechanism of the Dakin-West reaction

2.4.2 Synthesis of N-2-(oxopropyl)acetamide

Our first building block N-2-(oxopropyl)acetamide (Figure 2.19) was synthesized in two steps from glycine starting material, as first described by Dakin and West [49], and later by others [50]. The reaction was first developed by Suvadeep Nath who conducted the Dakin-West reaction under reported conditions by boiling glycine and an excess of acetic anhydride and pyridine overnight under a nitrogen atmosphere. Under these conditions glycine is acetylated and cyclizes into a mesoionic heterocycle intermediate 1,3-oxazolium-5-olate or Münchnone which reacts with excess acetic anhydride to give the desired product (Figure 2.20).

The resulting black crude product is isolated by removing the volatiles *in vacuo*, and the final product was isolated by hydrolyzing one acetyl group of the imide by refluxing in water (Figure 2.20). The extraction of the reaction mixture should be performed using a polar organic solvent such as dichloromethane, but the procedure is prolematic with high risk of emulsions and solubility problems exacerbated with residual pyridine. We later modified the reaction and boiled the Dakin-West product in methanol prior to workup (Figure 2.19).

The overall yield of the product is low from two to four percent. Considering the modest safety risks of exposure to refluxing acetyic anhydride and pyridine as well as the extraction procedure using large volumes of dichloromethane, the initial Dakin-West reaction is best done overnight as a very large volume and rarely as possible. If this approach is chosen, the crude product can be stockpiled and purified in batches as needed with no change in the product. An alternative to the workup has been devised that is more laborious, but produces a clear white oil that is difficult to obtain using silica purification alone.

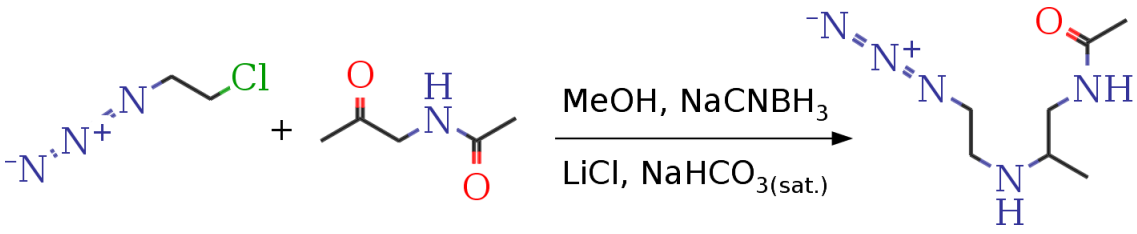

Figure 2.21: Synthesis of organic amine azide linker building block A

We followed the doublet at 4.17 ppm in the 1HNMR to identify fractions containing the pure Dakin-West product. The impurities were removed through short path Kugelrohr distillation under reduced pressure, and remaining polar impurities were removed using silica gel and flash chromatography with a gradient elution in five steps from 98:2 to 90:10 using dichloromethane:isopropanol as eluent. The pure Dakin-West product shows a doublet at 4.17 with a coupling constant of 4.15 Hz for the CH2 group. The acetyl groups are seen as adjacent singlets at 2.05 and 2.22 ppm which represent the two amide rotamers. The amide shows a broad peak at 6.19 ppm. There is often a contaminant that has a doublet at 4.07 ppm that should not be mistaken for the product. Every effort to isolate a pure product including Reverse Phase High Pressure Liquid Chromatography (RP-HPLC) using a 50:50 acetonitrile:water isocratic method.

2.4.3 Synthesis of Block A

The product of the reductive amination of N-acetyl-N-(2-oxopropyl)acetamide with 2-azidoethan-1-amine resulted in the expected building block azide product N-N-2-[(2-azidoethyl)amino]propylacetamide (Figure 2.21). It should be noted that this compound is sparingly solubule in chloroform and produces a poor 1HNMR spectrum at high concentrations. Therefore, ethyl acetate was chosen to workup the product. The stability of the azide under fluorescent lighting during handling is remarkable.

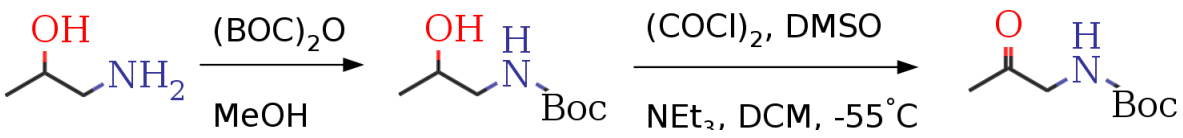

Figure 2.22: Synthesis of building block B precursor

1HNMR revealed that block A or N-N-2-[(2-azidoethyl)amino]propyl acetamide, hereafter Block A, has a methyl group with a chemical shift at 1.09 ppm. The amine is less de-shielded and is broad and upfield at 1.46 ppm. The acetyl group on the acetamide is unchanged from the initial Dakin-West product at 2.0 ppm. The two multiplets from 2.77-2.99 ppm are from similar CH2 groups next to the azide and acetamide, both with comparable electronegative neighbors.

2.4.4 Synthesis of building block B

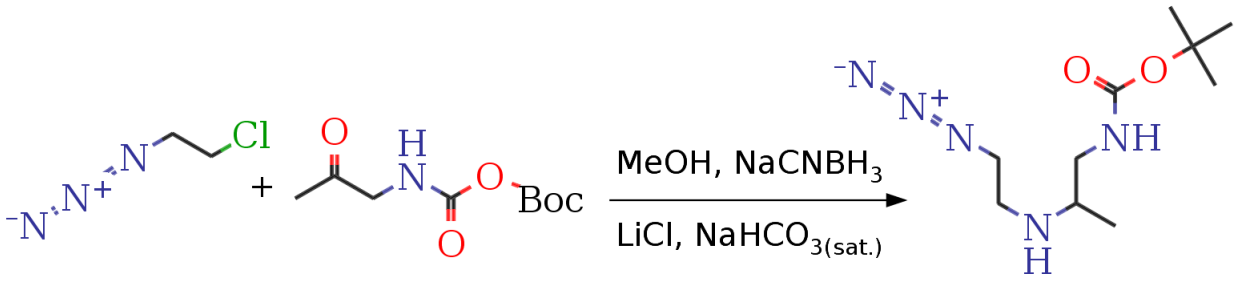

Figure 2.23: Synthesis of organic amine linker building block B

We will now discuss the synthesis of building block B from 1-aminopropan-2-ol conducted by Suvadeep Nath. The starting material was boc-protected in methanol to produce 1-[(tert-butoxy)amino]propan-2-ol, and the resulting alcohol was oxidized through a Swern oxidation to yield the product 1-[(tert-butoxy)amino]propan-2-one (Figure 2.22). Block B was synthesized by Suvadeep Nath similarly to Block A through a reductive amination with 2-azidoethan-1-amine (Figure 2.23).

2.4.5 Fluorophore conjugation to Block A and B

Block A and Block B both possessed secondary amines and were amenable to fluorophore conjugation using standard methods. Suvadeep Nath prepared the coumarin and FITC block A derivatives (Figure 2.8A-E), whereas I synthesized dervatives from sulfonyl chloride precursors (Figure 2.24F-H). The coumarin dyes were attached to building blocks A with ease using DCC coupling techniques via the addition of DMAP-HCl in dichloromethane at room temp. The dye, 7-(diethylamino)-2H-chromen-2-one (DEAC) was the first of the series (Figure 2.24A). The second dye, 7-methoxy-2H-chromen-2-one contains a methoxy group on the coumarin fluorophore (Figure 2.24B). The third fluorophore, 2H-benzo[h]chromen-2-one is an analog of the previous dye where a benzene ring

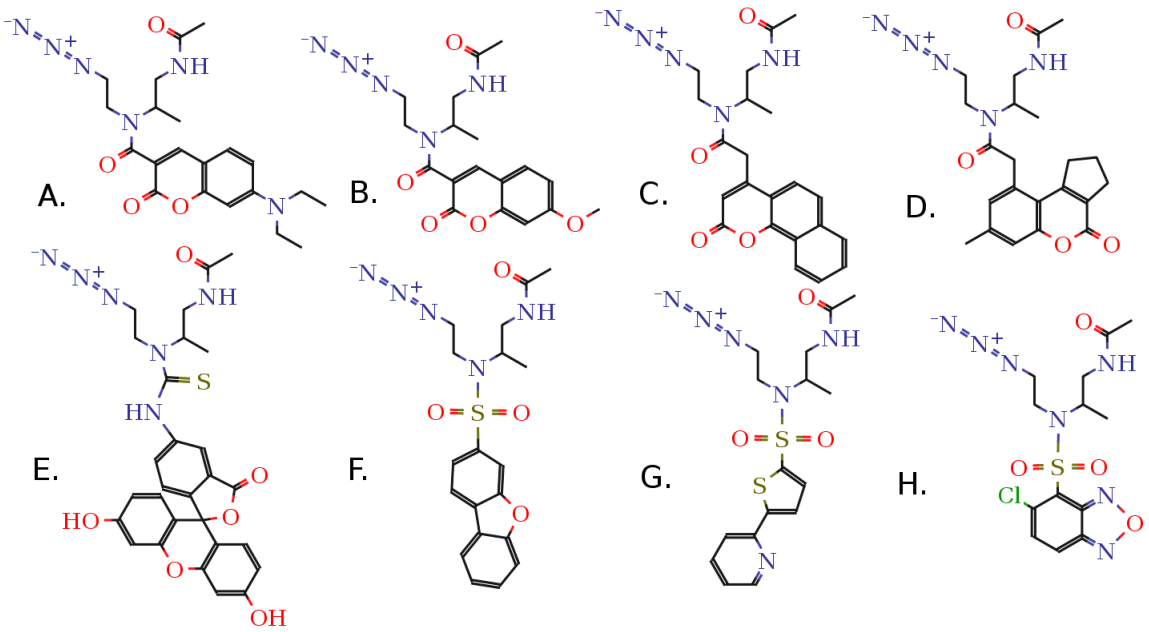

Figure 2.24: Panel of fluorophores attached to block A. A-D. Coumarin dyes. E. Fluorescien isothiocyanate. F-H. Dyes available as sulfonyl chloride precursors.

provides the red shift for the fluorophore (Figure 2.24C). The fourth, and last coumarin, 7-methyl-1H,2H,3H,4H-cyclopenta[c]chromen-4-one was the last coumarin synthesized by Suvadeep Nath (Figure 2.24D).

Fluorescein isothiocyanate (FITC) conjugate of Block A was prepared by dissolving a building block and FITC in tetrahydrofuran (THC) since FITC is a known secondary amine nucleophile (Figure 2.24E)(Suvadeep Nath). The final group of dyes were attached to block A from sulfonyl chloride precursors using DMAP, DIPEA in DCM (Figure 2.24F-H). The first tye of this group was dibenzofuran (Figure 2.24F). The second dye, 2-(thiophen-2-yl)pyridine from 5-(pyridin-2-yl)thiophene-2-sulfonyl chlorinde hereafter called pyridyl thiophene block A. The third dye, 5-chloro-2,1,3-benzoxadiazole-4-sulfonyl chloride, was also attached to block A. The structures of these three dyes was confirmed as click conjugates of clickable library acetylenes (Chapter 3: Chemical genomic screen).

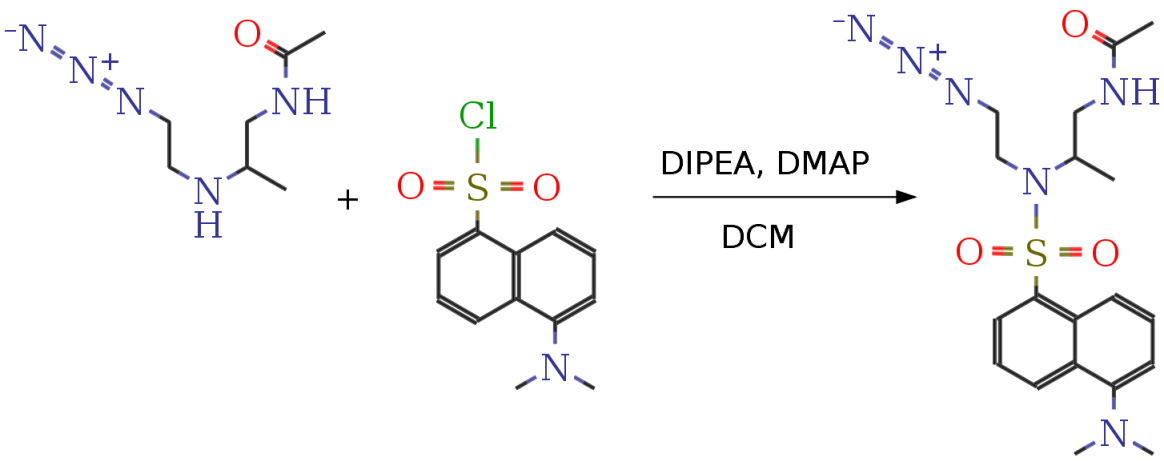

Figure 2.25: Dansylation of block A

The last dye attached to block A from a sulfonyl chloride precursor was N,N-dimethyl-5-sulfonylnaphthalene-1-amine (dansyl) chloride (Figure 2.25). This dye, when reacted with certain classes of the clickable library, lead to the most potent bioactive leads (Chapter 4: Phenotypes of bioactives). Therefore, the majority of this thesis pertains to the spectral characterization dansyl block A and block B dyes and select conjugates featured in later chapters (Chapter 4: Phenotypes of bioactives).

The addition of the dansyl fluorophore to block A and B (Figure 2.26) was accomplished through amine coupling methods in dry solvent and DMAP for a couple of hours. Contaminants were removed from the synthesis of dansyl block A using isocratic RP-HPLC conditions. 1HNMR confirmed dansyl azide amine block A had a sharp doublet at 0.96 ppm and several multiplets from 3-4 ppm for the CH2 groups on block A. Aromatic protons for the dansyl dye were observed as expected.

2.4.6 Synthesis of diazirine Block B

The synthesis of diazirine functionalized dansyl amine azide building block B (Figure 2.26) was accomplished by incubating the precursor in 10% trifluoracetic acid

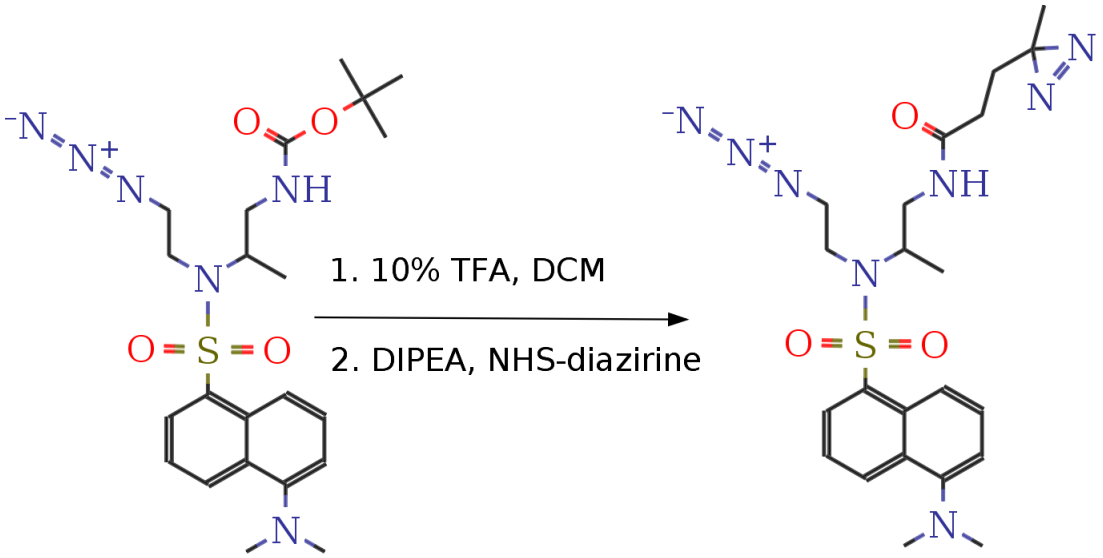

Figure 2.26: Synthesis of dansyl diazirine building block B

in DCM for thirty minutes, to de-protect the amine. Volatiles were removed *in vacuo* and the free amine was created *in situ* [51] through the addition of DIPEA to dissociate the amine-TFA salt. The resulting amine was reacted with commercially available N-hydroxysuccinimide ester alkyl diazirine or SDA-diazirine from Pierce.

The product, hereafter called dansyl daizrine amine azide (DDA) was purified from the reaction by using reverse phase chromatographic methods. The addition of the alkyl diazirine and linker to the building block caused an increase in retention time on the C8 column compared to the precursor (data not shown). HPLC purification was time intensive and low-yielding. To remedy this I chromatographed DDA using a step inject procedure to minimize overall run time and maximize isolation of pure product (data not shown). Three closely spaced automated injections had sufficient resolution under our elution conditions, and was highly reproducible between runs. The starting material co-elutes prior to the desired product in reverse and normal phase methods, and may be an avenue for recovery if necessary.

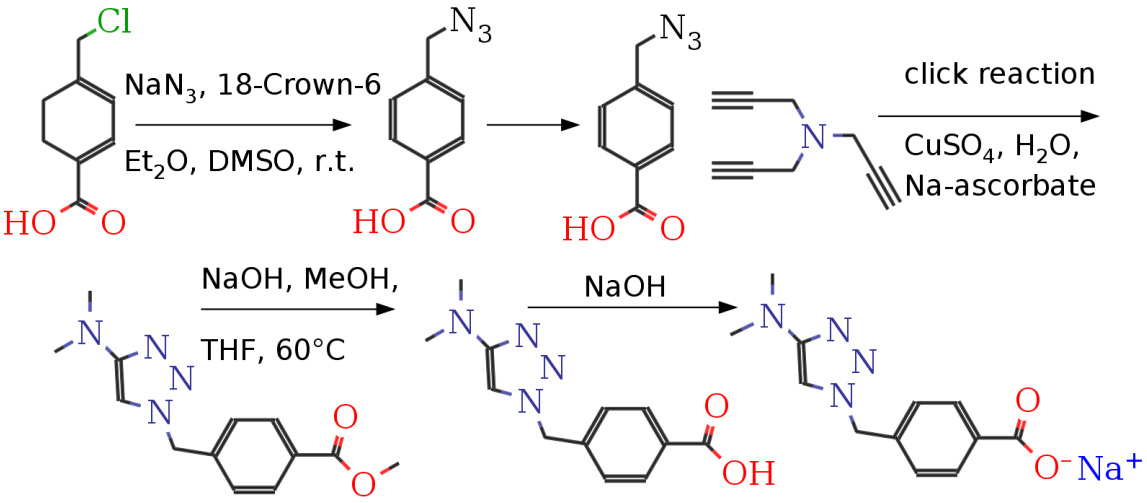

Figure 2.27: Synthesis of a cell impermiable TBTA-(CO2Na)3 catalyst for the click reaction.

The 1HNMR spectrum obtained was as expected, and the CH3 group on the diazirine ring is observed at the expected chemical shift with adjacent CH2 groups showing as a multiplet region upfield of 2.0 ppm. The synthesis of the alkyl and halogenated versions of the diazirine are discussed in the literature [52, 53][54]. The biochemical appication of the diazirine ring has been discussed extensively [55, 56] and will be discussed (Chatper 6: Target identification via Covalent capture). The stability of the compound is remarkable as the azide precursor and after the click reaction. The product remained intact and bioactive after several rounds of chromatography (Chapter 4: Phenotypes of bioactives). Nonetheless, precautions were taken to prevent exposure to sources of intense light and UV radiation.

2.4.7 Synthesis of cell impermeable TBTA ligand

A suitable TBTA ligand [39] was available from Sigam Aldrich to accelerate the copper catalyzed click reaction, but we chose to develop a novel cell impermiable version as a carboxylate salt. The TBTA variant enabled us to perform the biological screens of

the click reactions directly. We included 10 micromolar EDTA into the growth media to chelate remaining copper. TBTA-(CO2Na)3 was synthesized by Suvadeep Nath using a novel route (Figure 2.11A-E) in five steps from 4-(chloromethyl)benzoic acid combined with 18-crown-6 and sodium azide (Figure 2.27A). The product was extracted into diethyl ether with DMSO at room temperature, and reacted with acetyl chloride in methanol to give an esterified product (Figure 2.27B). The product was treated with tri(propargyl)amine in standard click reactions. The ester was converted to a salt via base hydrolysis (Figure 2.27C) using soidum hydroxide with heat in THF/methanol followed by gentle acid (Figure 2.27D), and then an equimolar amount of NaOH was added to create the final charged carboxylate salt (Figure 2.27E).

Biological testing confirmed TBTA-(CO2Na)3 was less bioactive than the ester and the carboxylic acid version. TBTA-(CO2Na)3 was tolerated up to 50 micromolar and higher (Chapter 3: Phenotypes of bioactives). The TBTA-(CO2Na)3 enabled us to perform bioassays directly on the click reactions with no purification (Chapter 3: Chemical genetic screen).

2.4.8 *in sillico* selection of the clickable library

The *in sillico* selection and automated assembly of the unique 4,002 compound library of drug-like terminal acetylenes was conducted using the R library(http://cran.r-project.org/) and ChemmineR [57] to select Rule Of Five (ROF) [58] compliant compounds with terminal acetylene functional groups from a collection of commercially available sources (emolecules.com) (Figure 2.28). Since, e-molecules requires a subscription to download the entire compound database I will elucidate a similar method to search the freely available ZINC database.

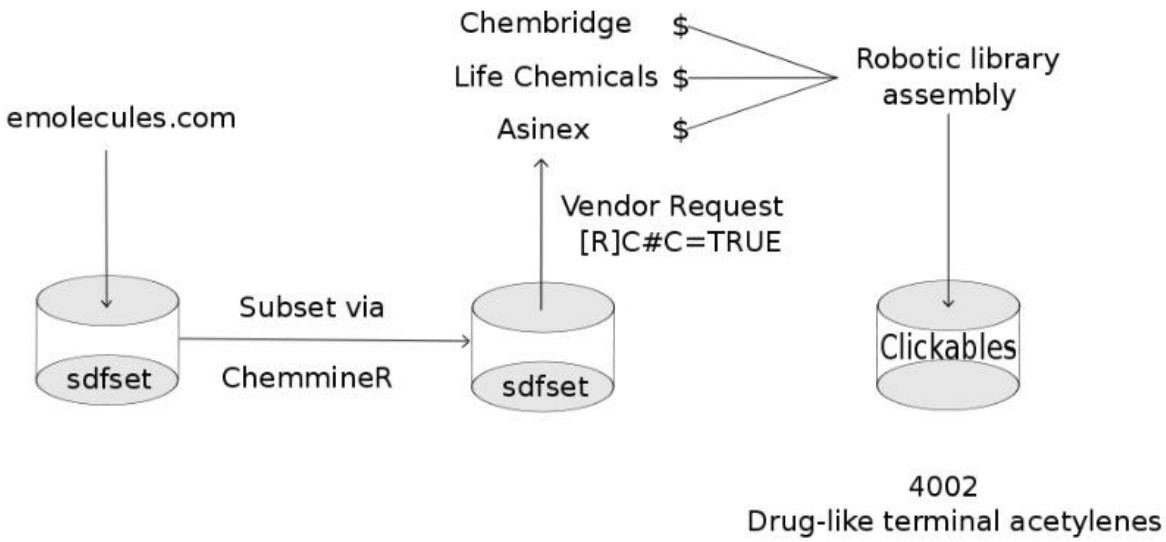

Figure 2.28: *in sillico* selection of the clickable library

ZINC is a host of approximately 2 million "cleaned" compounds in flexible chemical file formats (SDF, smiles, ihChI) along with availability and vendor info [47]. The library was downloaded as 49 gb of sdf.gz via wget and sent in chunks to the local 64-bit debian linux cluster (biocluster.ucr.edu) for parallel processing using R and the Bioconductor package ChemmineR [57].

2.4.9 Diversity of the clickable library

The clickable library was assembled in two phases. In the first phase 2,769 compounds were arrayed in plate format, shown as red dots (Figure 2.13A) and the remaining 1,233 compounds were added to the collection later, shown as orange dots (Figure 2.29B). Our acetylene search of ZINC determined the total purchasable chemical space with terminal acetylenes was greater than 80,000. We chose a subset of the whole for optimal "drug-likeness". The overall difference between the clickable collection, ZINC terminal acetylenes, and the resulting combinatorial libraries shown as merged,

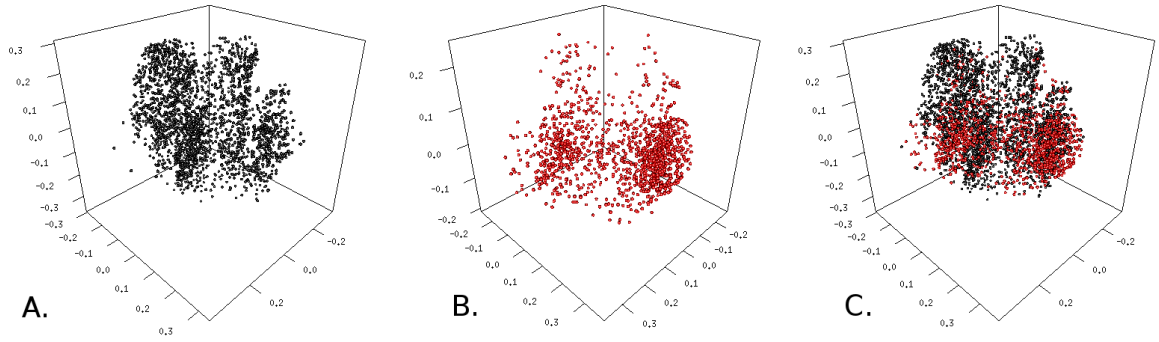

Figure 2.29: Three dimensional multidimensional scaling (3D-MDS) plot showing the chemical space of the clickable library. A. Black dots show the inital clickable collection of 2769 compounds. B. Red dots show the later addition to the library. C. The merge of the two collections in one MDS space.

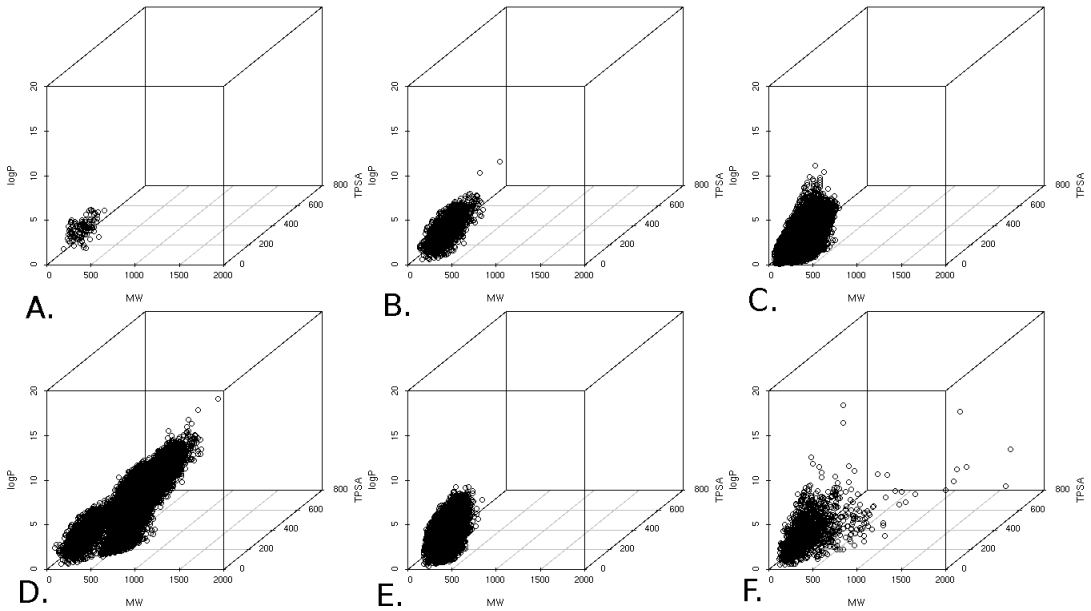

Figure 2.30: Comparison of three physicochemical properties for molecular weight (MW), total polar surface area (TPSA), and solubility (logP) of several small molecule libraries available for screening at University of California Center for Plant Cell Biology (CEPCEB). A. 114 *Arabidopsis* bioactives in clickable collection (Chapter 3: Chemical genetic screen) occupy a small domain. B. The clickable collection of 4,002 terminal acetylenes. C. ZINC80K, a collection of 80,000 purchasable terminal acetylenes. D. Properties of expected products from a 38,000 member combinatorial library produced by performing high throughput reactions with the clickable library and a panel of fluorophores and one glucose azide (not shown). E. The collection of 72,000 diverse small molecules at the Center for Plant Cell Biology (CEPCEB) assembled from Chembridge Diverset, Chembridge Novacore, Sigma TimTec, and Life Chemicals. F. The spectrum library a 2,000 compound natural product derived from Microsource.

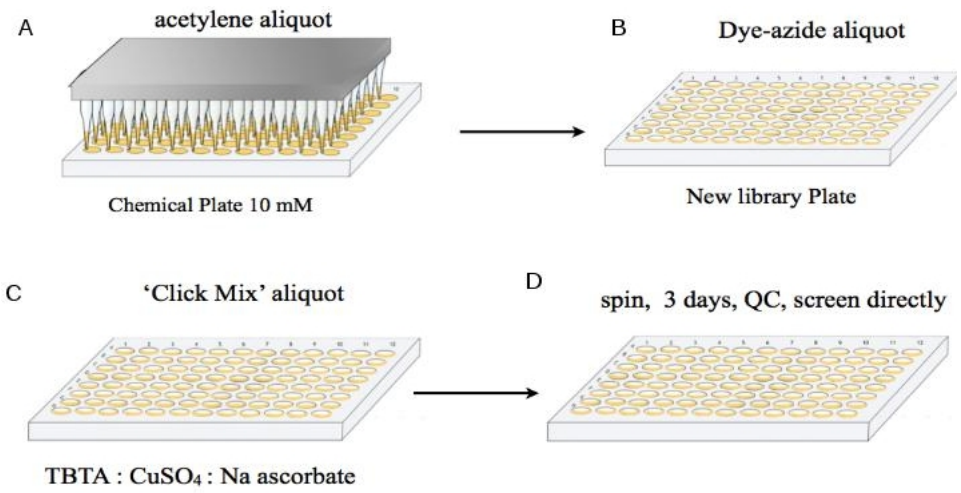

Figure 2.31: High throughput synthesis of fluorescent small molecule libraries form the clickable collection and azide building blocks.

is apparent by comparison of the molecular weight, logP and total polar surface area physicochemical features (Figure 2.30).

2.4.10 Physical assembly of the clickable library and combinatorial libraries

The clickable library of 4,002 terminal acetylenes was purchased from commercial vendors as a neat powder, and diluted in DMSO to a 10 mM master stock using the Biomek FX (methods). Aliquots of the master plates were made for future chemical genetic screens and 1 microliter was transferred to polypropylene plates in duplicate with a fluorescently labeled dye-azide building block (Figure 2.31A and B). One library was produced from each dye-azide linker, including one library from 1-beta-D-azido-glucopyranose. The click reaction was catalyzed through the addition of a 1:1:1 mixture of TBTA-(CO2Na)3 (10 mM) and CuSO4 (10 mM) and Na-ascorbate (100 mM) dissolved in water, hereafter called the click mix (Figure 2.31C). Reactor software from ChemAxon (http://www.chemaxon.com) was used to generate a chemical library file

with entries for each expected regiospecfific 1,2,3-substituted triazole and used for hit scoring, product prediction and record keeping.

Reactions were conducted for 3 days in the dark after centrifugation of the plates to collect all material in the center (Figure 2.31D). The library was screened thereafter. It is well known compounds in diverse collections are not always soluble in DMSO [58] and may be less soluble in aqueous. Sonication helps but does not solve this problem. Quality control was conducted on random wells from different plates each day a new combinatorial library was made (Chapter 3: Chemical Genetic Screen). HPLC monitoring revealed the reaction took, on average 12-24 hours to reach full completion in our 96-well format.

Precipitates appeared during the preparation combinatorial libaries from the clickables. This problem can be addressed systematically by removing the aqueous from the reaction *in vacuo* followed by re-suspsion in a minimum of DMSO.

2.5 Conclusion

We were successful to acquire 4,002 diverse clickable terminal acetylenes from vendors to build a large combinatorial library from click reactions with a fluorescent amine azide building block. The amine building block was synthesized from available precursors and we also synthesized a free amine building block as a precursor to a diazirine functionalized probe which will be highlighted later (Chapter 6: Target identification efforts by covalent modification). Furthermore, our combinatorial library was synthesized using TBTA-(CO2Na)3 to accelerate the screening process by testing the click reactions directly using *Arabidopsis thaliana* as a reporter (Chapter 3: Chemical Genetic Screen).

Chapter 3

Chemical Genetic Screen

Part I

3.1 Abstract

A library of 4,002 terminal acetylenes, the clickable library, and a collection of combinatorial reactions from this library were tested on *Arabidopsis* in high throughput. Our objective was to identify the subset of probes with biological effects on *Arabidopsis thaliana* by using the small plant growing in the dark as a reporter [59, 60]. In total 2.9 % of terminal acetylenes in the clickable library were bioactive and caused a wide spectrum of growth phenotypes such as aberrations in the growth of the hypocotyl, root, or cotyledon. The phenotype state was encoded into a vector of 26 features, hereafter called the phenotype fingerprint (PTFP). The PTFP enabled us to group bioactives with similar phenotypes. Bioactive compound structures were clustered based on atom pair similarity and core multiple common substructures (MCS) were used to partition the bioactives into clades A through M. Evidence is presented to compare the bioactivity trend of select clades of terminal acetylenes prior to derivitization with fluorescent amine azide building blocks. None of the terminal acetylenes clades examined in this section retained biological activity after derivitization into fluorescent probes which will be described in part II of this chapter.

3.2 Introduction

This chapter elucidates how these bioactive compounds were identified. A primary, secondary, tiered chemical genetic screen was conducted with the clickable library and fluorescent libraries made from the clickable library for cell expansion and growth inhibitors. The format of the primary screen was a high throughput 96-well plate format in duplicate (Figure 3.1 Step 1). Compounds identified as bioactive growth inhibitors were re-synthesized in bulk and re-tested in the original format (Figure 3.1

Step 2.). Candidates were then re-synthesized for analytical HPLC and re-testing of the observed phenotype (Figure 3.1 Step 3). Reactions were prioritized based on product formation as evidenced by analytical HPLC and phenotype strength. Lead candidates were subjected to preparative fractionation using RP-HPLC, and bioactive fractions were sent for validation by ESI-MS (Figure 3.1 Step 4 and 5). Proton NMR was performed on bioactive classes that were the subject to full phenotype investigation (Figure 3.1 Step 5) and (Chapter 3: Phenotypes of Bioactives). In the chemical genetic screen we examined the etiolated growth process in detail for the first seven days of growth and identified 119 bioactive compounds. A majority of which could not be categorized to phenotype classes, and several were fluorescent small molecules from the click reaction with dansyl building block A.

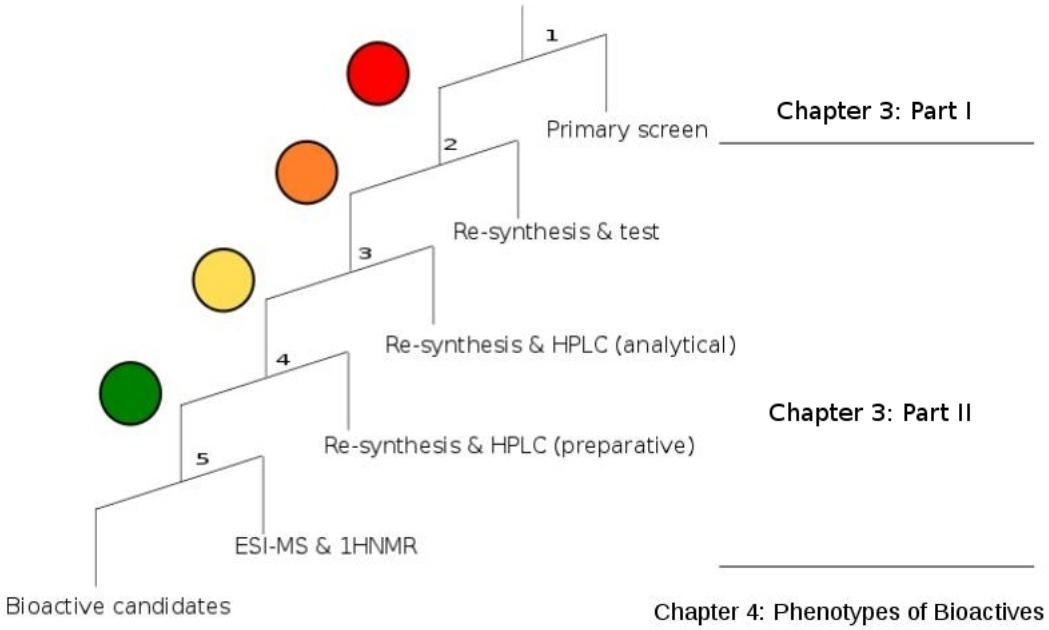

Figure 3.1: A tiered chemical genetic screen was conducted with the clickable library and combinatorial libraries produced via click reactions with our amine azide building blocks using *Arabidopsis thaliana* as a bioactivity reporter. Numbered 1-5 above. 1. The primary chemical genetic screen involved testing *en masse* available probe sets in 96-well plates. 2. Bioactives identified in the primary screen were re-synthesized in the original microscale and biologically tested again in a 96-well format. 3. Candidates that produced novel or confusing phenotypes were re-synthesized again in a larger scale and validated for product formation with analytical reverse phase HPLC. 4. A select pool of reactions were re-synthesized and fractionated using preparative HPLC. 5. The fractionate was tested biologically with *Arabidopsis* and bioactive fractions were sent for structural validation using ESI-MS. Expected products were re-synthesized and purified for elucidation by 1HNMR. Bioactive candidates are discussed at length in Chapter 4: Phenotypes of Bioactives.

3.2.1 Early growth and development of dark grown *Arabidopsis*

The growth of the dicotyledonous plant *Arabidopsis* follows a simple and distinct set of developmental milestones (Figure 3.2). When the seed absorbs water and is properly chilled at 4° C for four days the seed will germinate (Figure 3.2 stage 01). Hydrostatic pressure causes the radicle, or primordial root, to emerge from the seed coat (Figure 3.2 stage 3-5). Hypocotyl elongation is exaggerated in the dark and will

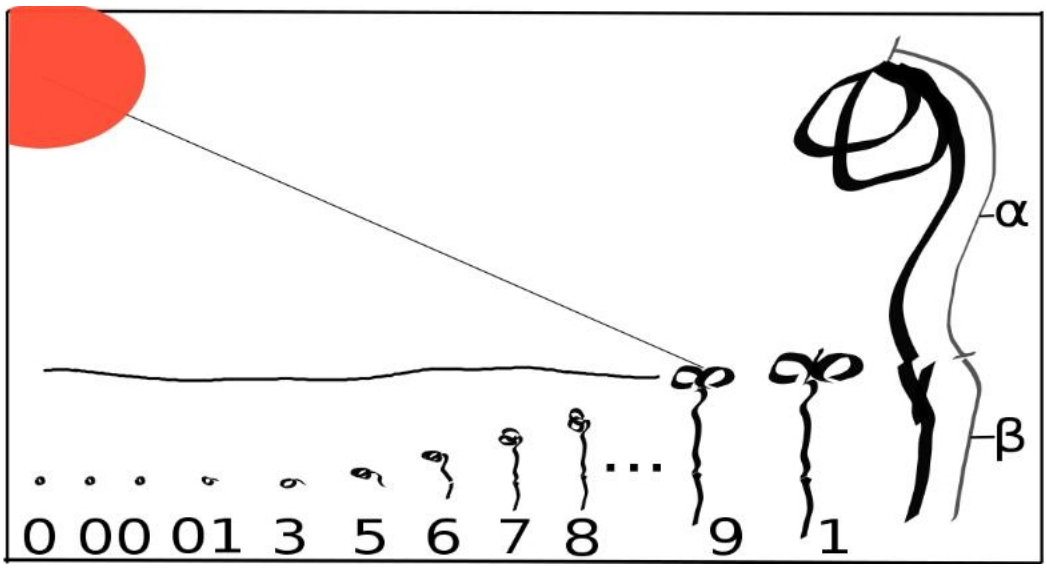

Figure 3.2: The life and development of etiolated *Arabidopsis* using the decimal code from 0 to 1 in the Biologische Bundesanstalt Bundessortenamt und CHemische Inustrie (BBCH) growth scale. The growth scale has the following descriptions. 0, time of setting down seed. 00, dry seed. 01, beginning of seed imbibition. 03, seed imbibition complete. 05, radicle emerged from seed. 07, hypocotyl with cotyledons breaking through seed coat. 09, emergence and cotyledon breaks through soil surface. 1. Cotyledons completely unfolded and growing point or true leaf initial visible.

continue until exposed to light (Figure 3.2 stage 6-9)[61]. The hypocotyl elongation aids the escape from the seed husk while the cotyledon and apical hook retain a protective angle until permissive conditions arise, as will be detailed.

3.2.2 The apical meristem and the cotyledon angle

The apical meristem, the future above ground body of the plant is a subpopulation of totipotential cells [62]. This important population of cell is physically protected from damage during seedling emergence by a coordinated process between the cell adjacent to the cotyledon on the stem (Figure 3.3A-E). These cells grow at an uneven rate on the apical and basal side of the stem to preserve the apical hook. Exposure to light accompanied with soil emergence causes a loss of skotomorphogenesis or differential cellular growth and the apical hook [63]. During this process the cotyledon is changing

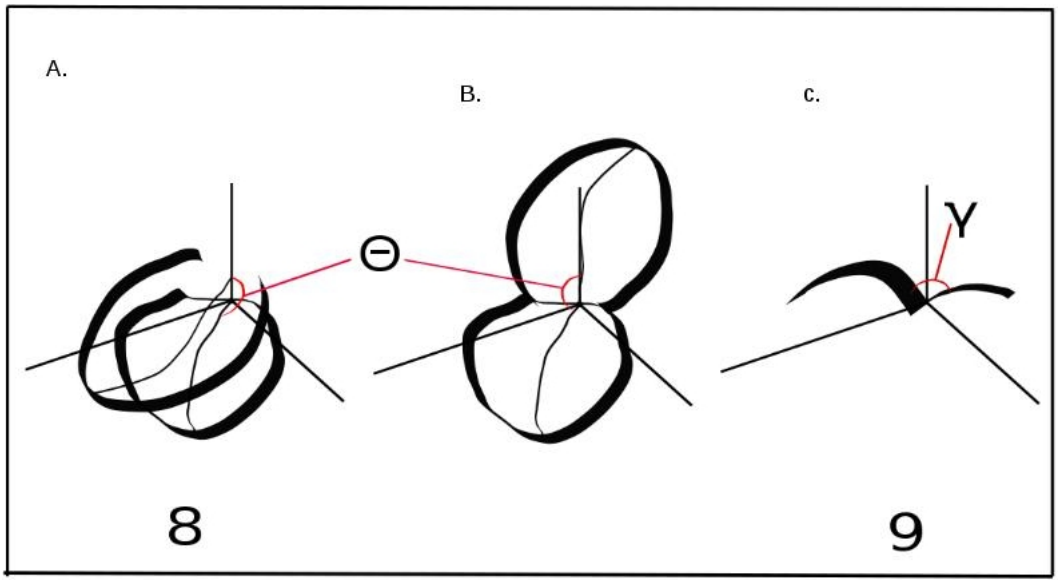

Figure 3.3: A schematic cotyledon of *Arabidopsis*. A. The cotyledon is pressed against itself, a consequence of embryonic development. B. The cotyledon opens and the angle phi shown increases. C. The cotyledon is fully open and above ground growth will continue from the apical meristem in this region. The numbers shown are BBCH developmental stage designations

internal metabolism, enlarges, and emerges from the soil gravitropically and then peels back (Figure 3.3B-C) to begin the first true leaf development from stem cells in the apical meristem (Figure 3.3C) [62].

3.2.3 Anisotropic growth of hypocotyl cells in the dark

Growth of dicots in the dark is characterized by cell elongation in the direction opposite to gravity, and this response terminates in response to light. Etiolated anisotropic hypocotyl expansion is a phenotypic process limited by hydrostatic processes in elongation and expansion regulated by cell wall synthesis and developmental programs [64, 61]. Etiolated growth facilitates emergence from the ground, and leads to an ideal spatial separation of the root and shoot with air, water and light gradients that dictate future organ differentiation in response to photosynthesis. It is our goal to discover compounds that perturb any or all of the processes of etiolated growth and

intend to study these bioactives to determine the overall mechanism of action.

3.3 Materials and Methods

3.3.1 Automated preparation of high throughput assay plates for chemical genetic screening

A biomek FXp Laboratory Automation Workstation was adapted to accurately and consistently dispense chemical, hot agar, and seed. All plates were inspected and visually and corrected by hand if necessary. Our automation procedure produced an aliquot error lower than expected from utilizing a mult-channel pipette.

A high throughput chemical genetic screen was performed in 96-well format in duplicate in all cases. The clickable library was aliquotted from 10 mM master plates and stored at -20° C when not in use. To test the clickable library 1 μL of chemical or DMSO was spotted into the 96-well plates and diluted with 99 μL of plate agar mix to a final concentration of 100 μM. The fluorescently taggaed library was screened similarly with 2 μl of unpurified reaction mixture, and to chelate copper from the click reaction EDTA was added to the growth media to a final concentration of 10 μM. *Arabidopsis* growth media was comprised of 0.67% agar, 0.13% Murashage and Skoog (Sigma-Aldrich), and ddH20. Assuming the click reactions proceeded to completion we estimated product concentration to be a maximum of 6.25 mM, thus the screening concentration was estimated up to 125 μM. Candidate hits from click reactions were re-synthesized and tested at a scale similar to the primary screen.

3.3.2 Preparation of *Arabidopsis* seeds and bioassay plates

Columbia-0 seeds, obtained from the National Arabidopsis Stock Center (NASC), were sterilized with 10% bleach and 0.005 Triton X-100 (Fisher) for 15 minutes and washed with sterile ddH20 four times. Seed aliquots of 12 μL were suspended in 0.015% agar and transferred to the top of growth media in the 96-well polystyrene plate (Greiner). On average 8-12 seeds were deposited per well, and was accomplished by suspending 17.73 mg of seed per 1 ml of suspension agar. Assay plates were subjected to vernalization at 4° C for four days to promote uniform germination, and moved to a humid and dark growth chamber for four days. This was the standard growth protocol for all experiments involving analysis of etiolated growth. On the fourth day plates were scored manually and with a Leica MZ12.5 microscope for inhibition of growth due to chemical treatment compared to a 1% DMSO control. Phenotypes pertaining to the hypocotyl, cotyledon, and root were scored, and hits that appeared in duplicate were considered candidates.

3.3.3 Clustering analysis using ChemmineR

3.3.3.1 Clustering of bioactive terminal acetylenes using hclust

R code:

```
##################
library(ChemmineR)
##################
#Clickable_ArabidopsisActives.sdf provided as a record

#load, generate apset
sdfset <- read.SDFset("Clickable_ArabidopsisActives.sdf")
apset <- sdf2ap(sdfset)

#create distance matrix to cluster apset
dummy <- cmp.cluster(db=apset, cutoff=0, save.distances="distmat.rda")
load("distmat.rda")

#plot apset distance matrix as hierarchical cladogram
hc <- hclust(as.dist(distmat), method="single")
hc[["labels"]] <- cid(apset)
plot(as.dendrogram(hc), edgePar=list(col=4, lwd=2), horiz=F)
```

3.3.3.2 Three dimensional multidimensional scaling (3D-MDS) plots

R code:

```
##################
library(ChemmineR)
##################
sdfset_actives<-read.SDFset("Clickable_Arabidopsis_Actives.sdf")

cid(sdfset_actives)<-datablocktag(sdfset_actives,tag="Click_ID")

Actives<-cid(sdfset_actives)

write.table(Actives, header=TRUE, quote=TRUE)
##################
load("Clickable_clusters.rda")

load("Clickable_coord.rda")

load("Clickable_sdfset.rda")
##################
cid(sdfset)<-datablocktag(sdfset,tag="Click_ID")

rownames(coord) <-clusters$ids<-cid(sdfset)
##################
Actives<-read.table("Actives")
```

```
rownames(Actives)<-Actives[,1]

coord[rownames(coord)%in%rownames(Actives),]

ActivesCoord<-coord[rownames(coord)%in%rownames(Actives),]

InActivesCoord<-coord[!rownames(coord)%in%rownames(Actives),]
#############
library(rgl)
#############
coord_beta <-InActivesCoord
rgl.open(); offset <- 50; par3d(windowRect=c(offset, offset, ...
640+offset, 640+offset))
rm(offset); rgl.clear(); rgl.viewpoint(theta=45, phi=30, fov=60, zoom=1)
spheres3d(coord_beta[,1], coord_beta[,2], coord_beta[,3], radius=0.005, ...
color="black", alpha=1, shininess=20); aspect3d(1, 1, 1)
axes3d(col='black'); title3d("", "", "", "", "", col='black'); bg3d("white")
#############
rgl.snapshot("coord_beta.png")
rgl.close()
#############
coord_alpha <-ActivesCoord
rgl.open(); offset <- 50; par3d(windowRect=c(offset, offset, ...
640+offset, 640+offset))
rm(offset); rgl.clear(); rgl.viewpoint(theta=45, phi=30, fov=60, zoom=1)
spheres3d(coord_alpha[,1], coord_alpha[,2], coord_alpha[,3], radius=0.005, ...
color="red", alpha=1, shininess=20); aspect3d(1, 1, 1)
axes3d(col='black'); title3d("", "", "", "", "", col='black'); bg3d("white")
#############
rgl.snapshot("coord_alpha.png")
rgl.close()
#############
rgl.open(); offset <- 50; par3d(windowRect=c(offset, offset, ...
640+offset, 640+offset))
rm(offset); rgl.clear(); rgl.viewpoint(theta=45, phi=30, fov=60, zoom=1)
spheres3d(coord_alpha[,1], coord_alpha[,2], coord_alpha[,3], radius=0.005, ...
color="red", alpha=1, shininess=20); aspect3d(1, 1, 1)
spheres3d(coord_beta[,1], coord_beta[,2], coord_beta[,3], radius=0.005, ...
color="black", alpha=1, shininess=20); aspect3d(1, 1, 1)
axes3d(col='black'); title3d("", "", "", "", "", col='black'); bg3d("white")
#############
rgl.snapshot("coord_alphaVsbeta.png")
rgl.close()
#############
```

3.3.4 Multiple common substructure clustering using atom pair similarity via Tanimoto and hclust

```
##################
library(ChemmineR)
library(fmcsR)
##################
files <- list.files(recursive=FALSE, pattern=".sdf")
##################
A <-c(2,3,2,2,2,2,3,1,1,3)
B <-c(3,3,3,3,3,3,3,4,4,3)
##################
#CladeSelection <-c("B","C","D","E","Gi","Gii", "I", "Lii", "Li", "M", "N")
CladeSelection <-c("CladeB.sdf","CladeC.sdf", "CladeD.sdf", "CladeE.sdf", ...
"CladeGi.sdf", "CladeGii.sdf", "CladeI.sdf", "CladeLii.sdf", "CladeLi.sdf", ...
"CladeM.sdf", "CladeN.sdf", "CladeO.sdf", "CladeP.sdf")
##################
DoThis <- function(a){
##################
name <- gsub(".sdf", "", CladeSelection[a])
png_name <- paste(name, "_hclust.png", sep="")
png_name2 <- paste(name, "_SAR.png", sep="")
table_name <- paste(name, "_SAR.txt", sep="")
##################
png(file=png_name, width=800, height=600, units="px")
##################
sdfset <- read.SDFset(CladeSelection[a])
##################
blockmatrix <-datablock2ma(datablocklist=datablock(sdfset))
blockmatrix[,4]<-gsub("\\s__","",blockmatrix[,4])
IDs<-blockmatrix[,4]
blockmatrix[,4]<-gsub("CLK","",blockmatrix[,4])
blockmatrix[,4]<-gsub("_","",blockmatrix[,4])
datablock(sdfset)<-blockmatrix
blockmatrix <-datablock2ma(datablocklist=datablock(sdfset))
##################
cid(sdfset)<-datablocktag(sdfset,tag="Click_ID")
##################
d <- sapply(cid(sdfset), function(x)
fmcsBatch(sdfset[x], sdfset, au=0, bu=0,
matching.mode="aromatic")[,"Overlap_Coefficient"])
##################

##################
write.table(round(d*100), file=table_name, sep=" & ", quote=FALSE)
##################
hc <- hclust(as.dist(1-d), method="complete")
hc[["labels"]] <- cid(sdfset)
```

```
plot(as.dendrogram(hc), edgePar=list(col=4, lwd=2), horiz=TRUE, main=name)
dev.off()
##################
printZ<-LETTERS[1:length(A)]
printZ <- paste(printZ, ". CLK", cid(sdfset), sep="")
##################
png(file=png_name2, width=1024, height=768, units = "px")
plot(sdfset, griddim=c(A[a],B[a]), print_cid=printZ)
dev.off()
}
##################
a <- 1:length(CladeSelection)
##################
lapply(a, DoThis)
##################
```

3.3.5 Phenotype fingerprint (PTFP) annotation guidelines

Previous work has been done to formalize the collection and annotation of diverse phenotypes into discrete binary bins [65]. These phenotype bins are shown below (Table 3.1). Most of the procedures require only the naked eye and have been designated "eyeball". The matrix resulting from the scoring of phenotype bins in our scheme were used to prioritize candidate leads, and interpret structure activity relationships. Therefore, I will detail how the determination is made to score the presence or absence of a phenotype during the scoring procedure.

3.3.5.1 Apical hook

An apical hook in this sense is more like an apical corkscrew. The control apical hook shown here (Figure 3.8) is not considered an "apical hook".

3.3.5.2 Apical collar

If a radial constriction appears at the interface between the hypocotyl and the cotyledon it is considered an apical collar.

3.3.5.3 Apical meristem protrusion

A protrusion from the center of the open cotyledon or shoot apical meristem.

3.3.5.4 Hookless

This term was inhereted from Ecker, and has some subtlety. If the cotyledon was 90 degrees or more with respect to the hypocotyl it is considered "hookless", as observed in ethylene mutants.

3.3.5.5 Open-T cotyledon

If the cotyledon is of a V-shape and is open beyond 45 degrees it is considered a T-cotyledon.

3.3.5.6 Open-V cotyledon

A cotyledon can be considered an "open-v" under circumstances when the cotyledon does not point up. If the shape of the open cotyledon is shaped like the letter "v" it qualifies for inclusion into this phenotype bin.

3.3.5.7 Open scissor cotyledon

A cotyledon is considered to by an "open scissor" if the cotyledon is open and if the two leaves are scissored against each other before opening. Contrast the scissor cotyledon with the V-cotyledon, where the leaves are in the same plane. The cotyledon is considered a scissor even if the angle between the center of each leaves is small. If the leaves of the cotyledon are pressed together this is not considered "open scissor". Cotyledons pointing down that satisfy these characteristics qualify for a hit in this phenotype bin.

3.3.5.8 Bleached cotyledon

Etiolated or dark-grown seedlings have inactive chlorophyll and are subsequently yellow prior to exposure to light. "Bleached cotyledons" do not green along the typical time course, and are delayed in respect to photosynthetic greening.

Stained cotyledon

A cotyledon may exhibit spots or characteristic stains in small domains on the cotyledon.

3.3.5.9 Left handed twisted hypocotyl

The orientation of the elongated cells of the hypocotyl recapitulate a left handed twist, as if one held up their left hand with thumb pointing up to the sky and fingers curling inward to the palm to reprsent the twisting axis.

3.3.5.10 Right handed twisted hypocotyl

The orentation of the elongated cells of the hypocotyl recapitulate a right handed twist.

3.3.5.11 Swollen hypocotyl

A hypocotyl is considered swollen if the diameter relative to the seedling size is larger than expected. Even slightly swollen hypocotyls were considered for this class.

3.3.5.12 Curly hypocotyl

A hypocotyl is considered curly if the hypocotyl curls onto itself of if the whole seedling makes a shape characteristic of the letter "C" or "S". This annotation is applied to seedlings of large and small size equally. This may be problematic as a long seedling

may be more accurately considered curly whereas a short seedling may be described as curved, but would be expected to curl if of larger size. Lastly, if a hypocotyl has a considerable spiral this can be considered curly.

3.3.5.13 Bulging hypocotyl cells

If the hypocotyl cells are discernible bulging or diverging from the typical columnar shape the cell is considered bulging.

3.3.5.14 Distorted cell alignment

Cells in any tissue may exhibit distorted alignment. Often hypocotyl cells are long, and have a distinct and parallel, end-to-end alignment. Deviations from this pattern are considered "disordered cell alignment".

3.3.5.15 Stained hypocotyl

A hypocotyl may exhibit a reproducible stain in any part of the stem.

3.3.5.16 Root promotion

If the root is larger than would be expected for the relative seedling size this phenotype is indicated. Especially if the hypocotyl is small but the root is half or more the size of the hypocotyl.

3.3.5.17 Swollen root

The root must be a minimum size to be considered for the "swollen root" phenotype bin. For example, the compounds auxin and isoxaben create very short roots that are undeveloped, but swollen. These are not considered.

3.3.5.18 Stained root

Any reproducible root stains are considered.

3.3.5.19 Germination inhibition

Seeds that do not germinated or are arrested in the germination process are scored for inclusion in this phenotype bin.

3.3.5.20 Radical

A seedling is considered a radical if it is barely germinating and the radical is protruding from the seed coat.

3.3.5.21 20-80% hypocotyl inhibition

Assignment of phenotype based on hypocoty inhibition value compared to control seedlings with the DMSO vehicle at 1%.

3.3.5.22 Root inhibition

Root size is judged relative to seedling size, rather than hypocotyl length which is considered with respect to the control seedling.

Phenotype Bin	Short Details	Procedure	Value
apical hook		eyeball	1
apical collar		eyeball	1
apical meristem protrusion	apical meristem early growth	eyeball	1
hookless	no hook, straight up and down	eyeball	1
open T-cotyledon	T	eyeball	0.5
open V-cotyledon	V	eyeball	0.5
open scissor cotyledon	$>$ or $<$	eyeball	0.5
bleached cotyledon		eyeball	0.2
stained cotyledon		eyeball	0.2
left handed twisted hypocotyl	LH twist	eyeball	0.5
right handed twisted hypocotyl	RH twist	eyeball	0.5
swollen hypocotyl	swollen hypocotyl	eyeball	0.8
curly hypocotyl	spiral	eyeball	1
bulging hypocotyl cell	swollen hypocotyl cells	eyeball	1
distorted cell alignment	alignment of cells is disordered	eyeball	0.8
stained hypocotyl	dye localized in domains	eyeball	0.2
root promotion	root longer than control	eyeball	0.5
swollen root	root is swollen	eyeball	0.8
stained root	dye localized in domains	eyeball	0.8
germination inhibition	germination inhibtion	eyeball	0.5
radical	radical protrusion	eyeball	0.5
20-50% hypocotyl inhibition	80-50% control	imagej	0.8
51-80% hypocotyl inhibition	49-20% control	imagej	0.2
>80% hypocotyl inhibition	<20% control	imagej	1
>80% root inhibition	<20% control	imagej	1

Table 3.1: Phenotype annotation guidelines used to assign scores of one or zero for each phenotype annotation bin. The weight shown was used as a parameter for the clustering of the phenotype fingerprint (PTFP) matrix of all bioactive terminal acetylenes in the clickable library.

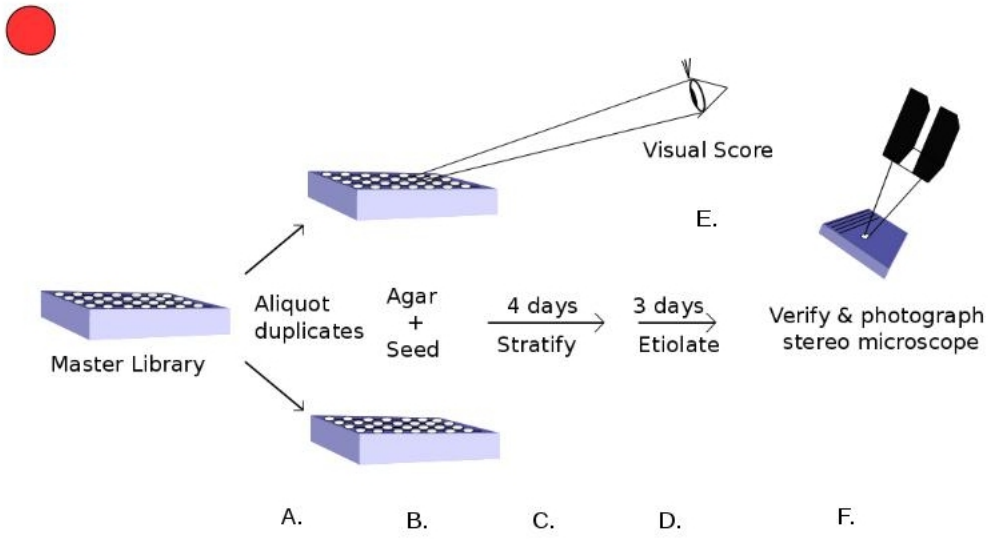

Figure 3.4: *Arabidopsis thaliana* primary chemical genetic screen performed in a 96-well assay. A-B. Chemical is transferred to empty plates diluted with growth media in agar and seed are added to the top of the hardened surface. C-F. Seeds are chilled in the refrigerator on assay plates for four days and placed in a dark growth chamber for three days prior to inspection by eye or microscope.

3.4 Results

3.4.1 Primary Chemical Genetic Screen

The *Arabidopsis* chemical genetic screen was conducted by transferring a small volume of compound library in duplicate to assay plates with growth media and seed to a concentration up to 100 μM (Figure 3.4 A and B). This procedure can be performed by hand using a multichannel pipette, but was automated by using a custom Biomek FXp workstation with an AP-96 multichannel head. The assay plates were stratified, chilled, for four days to promote even germination (Figure 3.4C) and grown in a humid and dark growth chamber for three days to observe etiolated growth (Figure 3.4D). Visual inspection of assay plates at multiple angles was performed to isolate seedlings did not grow to approximately 80% of the well height (Figure 3.4E)[66]. The visual scoring procedure accelerated prioritization of wells for microscope investigation (Figure 3.4E

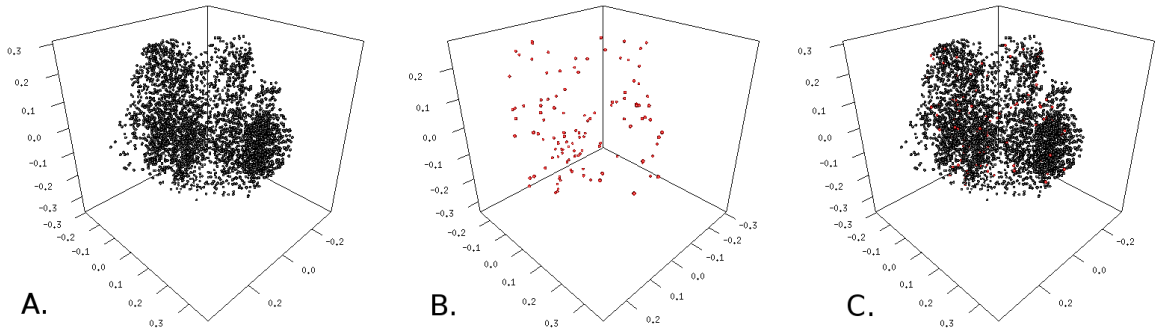

Figure 3.5: Three dimensional multidimensional scaling (3D-MDS) plots showing the chemical diversity of bioactives in the clickable collection. A. Plot showing the chemical space of the inactive compounds in the clickable collection as black dots. B. Plot showing active compounds in the clickable collection as red dots. C. Both active and inactive compounds shown as an overlay in the same chemical space.

and F). The process of photomorphogenesis or light response is known to occur in less than thirty minutes, and can reverse dark grown phenotypes. For this reason assay plates were not exposed to light prior to examination. The process of dark growth or skotomorphogenesis will be discussed in the following chapter (Chapter 4: Phenotypes of bioactives).

3.4.2 Diversity analysis of bioactive terminal acetylenes

The SDFset of bioactive terminal acetylenes was used to generate an atom pair descriptor set (APset). The Tanimoto metric was chosen to evaluate atom pair similarity and embed compounds into a 3-dimensional distance matrix which is the product of a multidimensional scaling operation hereafter 3D-MDS [57]. Tanimoto values were computed by matching the connectivity matrix of two compounds. A Tanimoto value of 1.0 indicated a one-to-one match for all bonds if the cutoff is set to 1.0 or 100 % match.

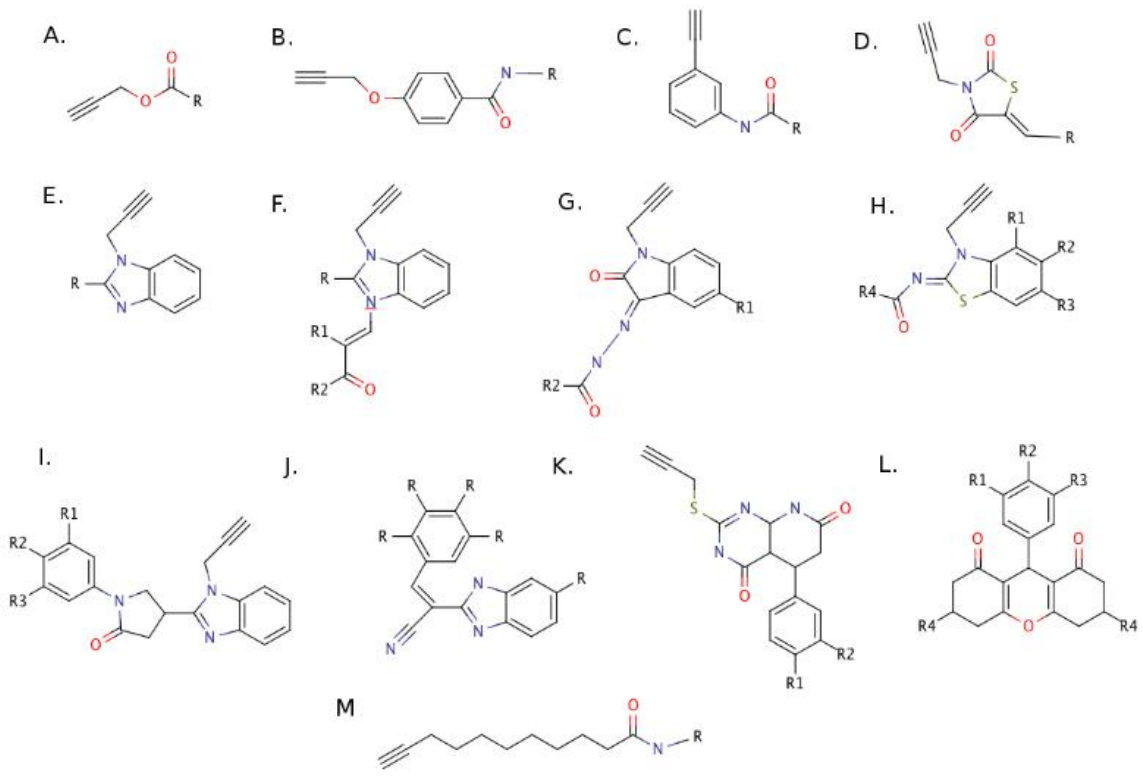

Figure 3.6: A panel of the core multiple common substructures (MCS) identified in bioactive acetylenes in the clickable library. A-M. MCS of the clades discussed in the text. Clades A-M.

The primary chemical screen determined the clickable collection contained a total of one hundred nineteen bioactive compounds (Figure 3.5). A quick glance shows that bioactives were loosely spread across the chemical library (Figure 3.5A and B). To compare the results of bioactives that clustered closely in chemical space we subset the bioactives based on shared multiple common substructures (MCS) (Figure 3.6). We explored bioactivity trends among these clades of acetylene compounds by re-testing at a 25 μM concentration in a 24-well format. Seedlings were transferred to a hard agar surface and photographed using a dissecting microscope mounted with a digital camera. Typical control seedlings exhibited elongated hypocotyls and long roots with a preserved apical hook and closed cotyledons (Figure 3.8). These are the features of an etiolated seedling, and will be discussed in more detail in the next chapter (Phenotypes of Bioactives). Use the phenotype key to look up the phenotype classification for bioactive compounds shown in the next figures (Table 3.2).

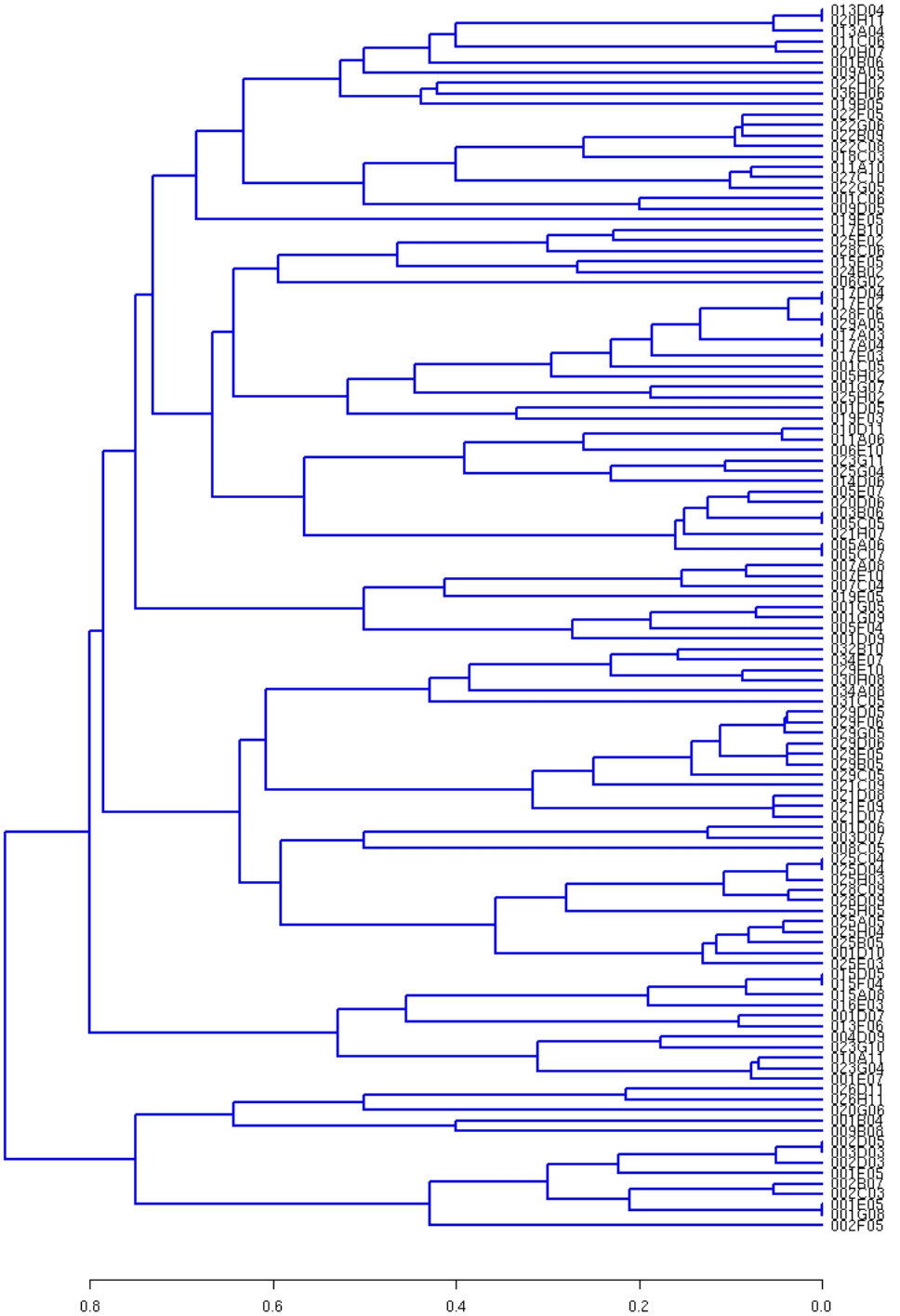

Figure 3.7: Cladogram showing the distance between all *Arabidopsis* active compounds in the clickable collection, distances were derived through the Tanimoto atom pair similarity metric in the R package fmcsR. Compound IDs are visible by zooming into the cladogram.

Figure 3.8: Representative control *Arabidopsis thaliana* Columbia-0 seedling grown in the dark and photographed on the seventh day. Seedling is shown at 1.0 magnification grown on 1% dimethylsulfoxide (DMSO) as solvent vehicle control. Scale bar indicates 2 mm.

	Category	Ref
K	curly	
C	cytokinin-like	[67]
A	auxin-like	[4]
I	isoxaben-like	[68]
H	hookless	[63]
G	gibberellin	
R	short root	
D	radical protrusion	
M	germination disturbance	
T	twisted hypocotyl	
Y	hypocotyl inhibition	
S	stained-root	
O	oryzalin-like	[69]

Table 3.2: Phenotype key for assigned phenotype categories of bioactives from the clickable collection. The short letter is used in following figures to designate assignment to a phenotype group. K, denotes curly hypocotyl or roots. C, denotes phenotypes recapitulated by cytokinin. A, denotes short stubby roots, apical collars, and swollen short hypocotyls. I, denotes phenotype appearance similar to cellulose synthase inhibitor isoxaben. H, denotes a missing apical hook that is typical for etiolated seedlings. G, denotes phenotypes similar to treatment with gibberellin. R, denotes roots shorter than control. D, denotes the emergence of the radical or primordial root from the seed. M, denotes any disturbance that prevented the seed from germinating. T, denotes hypocotyls that have a twisted cell file. Y, denotes shorter than control hypocotyls. S, denotes roots exhibit a conspicuous collection of compound creating stained puncta. O, denotes hypocotyl and root features associated with the microtubule disruptor oryzalin.

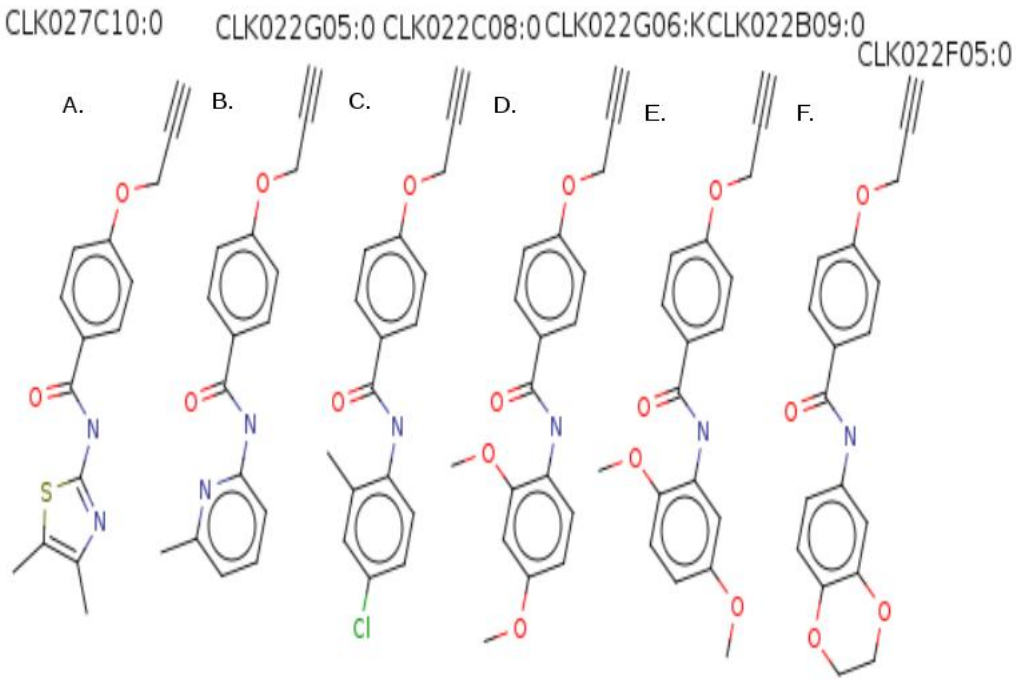

Figure 3.9: Clade B compounds with bioactivity annotation.

3.4.3 Clade overview of bioactive terminal acetylenes

3.4.3.1 Clade B compounds

Clade B compounds (Figure 3.9), contain a 4-(prop-2-yn-1-yloxy) benzamide core (Table 3.3). The table (Table 3.3) and following table have been sorted to match the heatmap, and the order of cladograms produced by executing hclust on the distances of compounds in the respective clade (described in methods). The heatmap indicates the score of percentage simlarity when executing a search with a cutoff of 70%(Figure 3.10).

The compound CLK022G06 can be observed with a curly hypocotyl phenotype (Table 3.2) -(Figure 3.11) when observed with nearest neighbors in the clade with scores

	ID
2.	CLK022G05
1.	CLK027C10
6.	CLK022F05
4.	CLK022G06
3.	CLK022C08
5.	CLK022B09

Table 3.3: Table showing the index and ID of clade B compounds shown in the next heatmap.

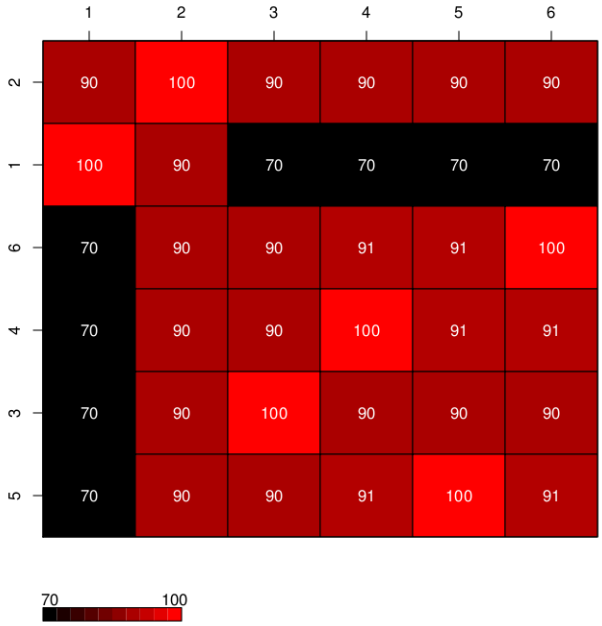

Figure 3.10: Heat map showing the distance between clade B compounds in the clickable collection, distances were derived from the Tanimoto atom pair similarity metric in the R package fmcsR. Values shown are the Tanimoto value *100, with a score of 100 indicating 70 % similarity.

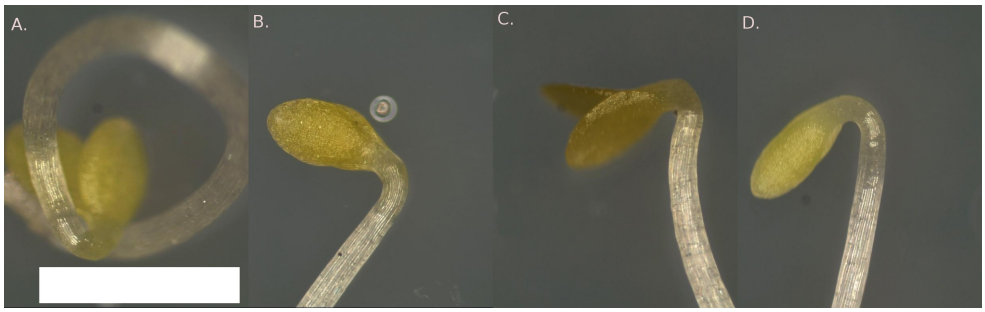

Figure 3.11: Phenotypes of clade B compounds in the clickable collection. A. CLK022G06. B. CLK022F05. C. CLK022B09. D. CLK022C08. All images at the same scale with a scale bar shown indicating 1μm.

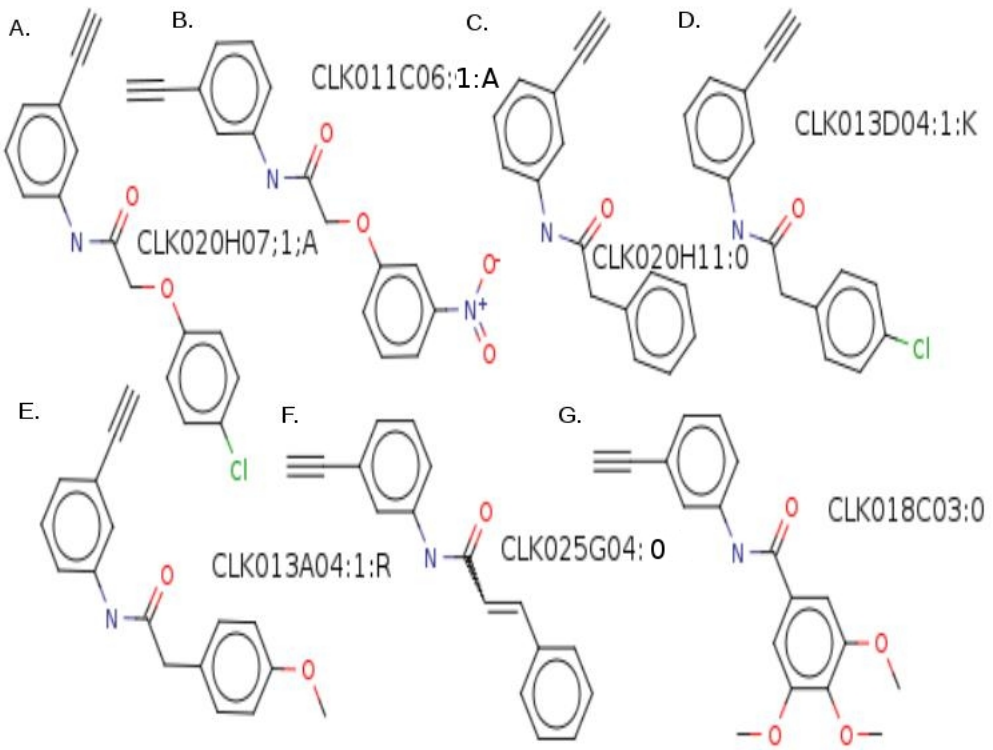

Figure 3.12: Clade C with bioactivity information. A-G.

of 90-91%(Figure 3.10).

3.4.3.2 Clade C compounds

Clade C compounds (Figure 3.12), 3-ethynylphenyl acetamide class (Table 3.4), causes the formation of apical collars reminiscent of the auxin phenotype (Figure 3.14A). The compounds CLK020H07 (not shown) and CLK011C06 (Figure 3.14A) both contain electronegative substituents which may be responsible to cause auxin-like phenotypes. The compound CLK13A04 is 60% similar to CLK020H07 and CLK011C06, and had root problems not shown. The root inhibtion caused by CLK013A04, not shown, may or may not be related to phenotypes seen in this clade. It was not determined.

	ID
2.	CLK011C06
1.	CLK020H07
5.	CLK013A04
4.	CLK013D04
6.	CLK025G04
3.	CLK020H11
7.	CLK018C03

Table 3.4: Table showing the index and ID of clade C compounds shown in the next heatmap.

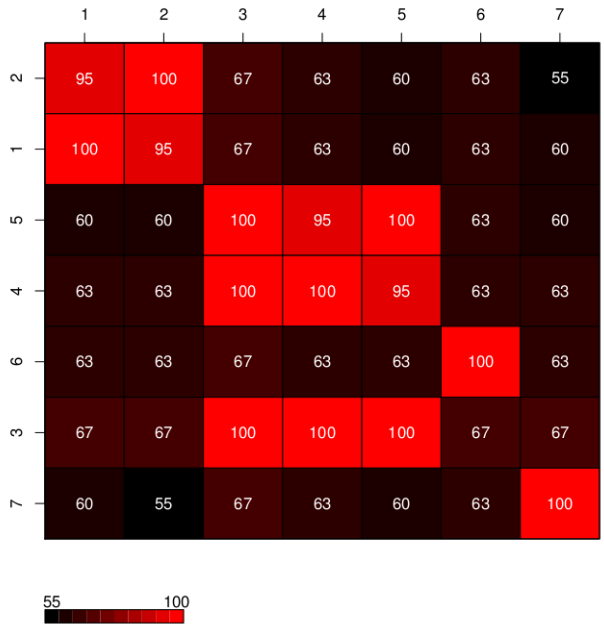

Figure 3.13: Heat map showing the distance between clade C compounds in the clickable collection, distances were derived from the Tanimoto atom pair similarity metric in the R package fmcsR. Values shown are the Tanimoto value *100, with a score of 100 indicating 70 % similarity.

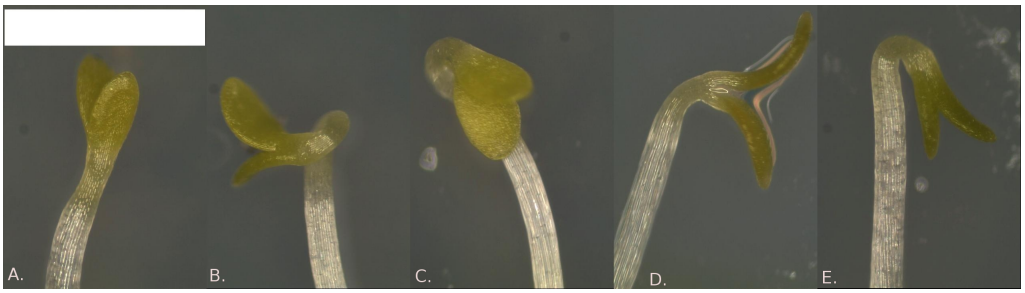

Figure 3.14: Phenotypes of clade C compounds in the clickable collection. A. CLK011C06. B. CLK013D04. C. CLK013A04. D. CLK018C03. E. CLK020H11. All images at the same scale with a scale bar shown indicating 1μm.

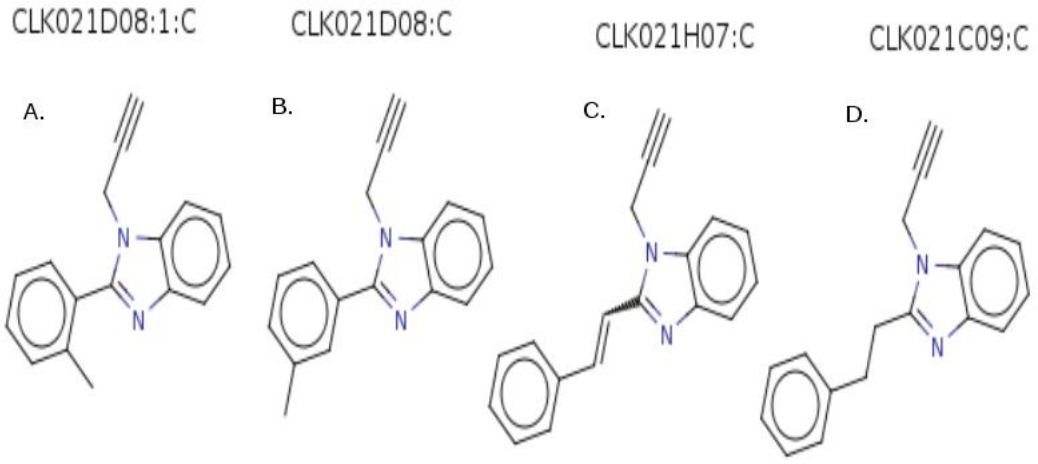

Figure 3.15: Clade E compounds with bioactivity annotation.

3.4.3.3 Clade E compounds

Clade E compounds (Figure 3.15 and (Table 3.4) are the most potent to cause severe apical hook problems (Figure 3.17A-D). These compounds were not considered further.

	ID
2.	CLK021D08
1.	CLK021F09
4.	CLK021C09
5.	CLK021D07
3.	CLK021H07

Table 3.5: Table showing the index and ID of clade E compounds shown in the next heatmap.

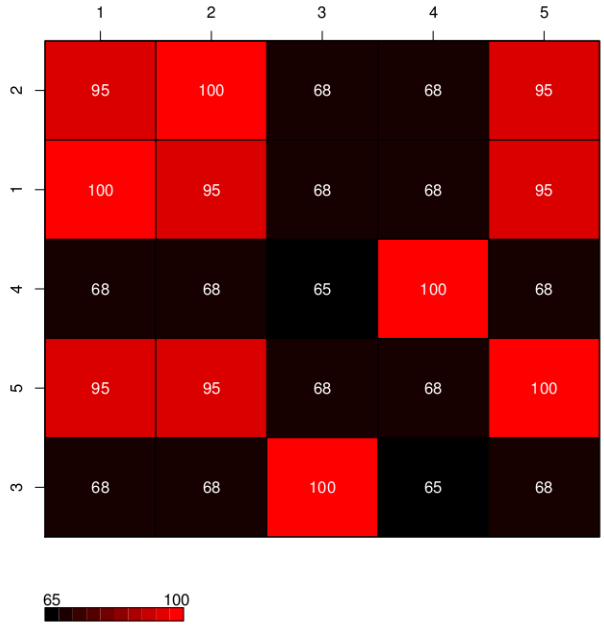

Figure 3.16: Heat map showing the distance between clade E compounds in the clickable collection, distances were derived from the Tanimoto atom pair similarity metric in the R package fmcsR. Values shown are the Tanimoto value *100, with a score of 100 indicating 70 % similarity.

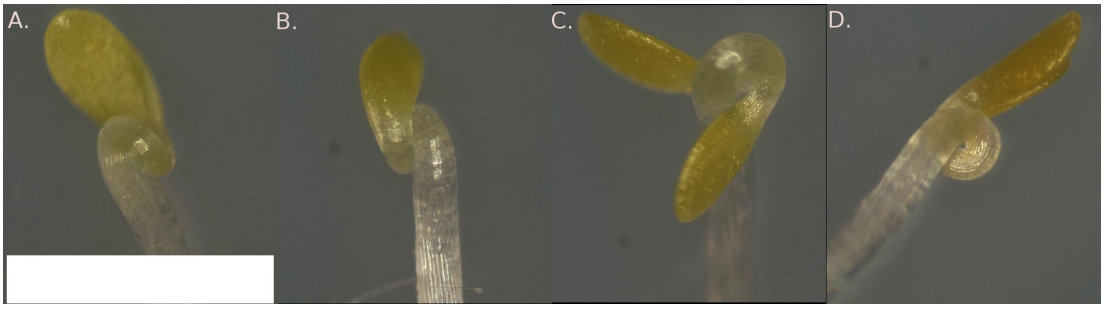

Figure 3.17: Phenotypes of clade E compounds in the clickable collection. A. CLK021D08. B. CLK021F09. C. CLK021H07. D. CLK021C09. All images at the same scale with a scale bar shown indicating 1μm.

	ID
2.	CLK029D05
1.	CLK029F06
3.	CLK029G05
5.	CLK029D06
4.	CLK029E05
7.	CLK029B05
6.	CLK029C05

Table 3.6: Table showing the Tanimoto similarity coefficients for clade I compounds. The similarity between compound pairs is shown with compound identifier (CLK ID) and value as percent. The matrix has been sorted to match the compound ID order of the matching cladogram on this page.

3.4.3.4 Clade I compounds

Clade I compounds (Figure 3.19) stain the hypocotyl and root in distinct domains and cause swelling reminiscent of the isoxaben phenotype (Figure 3.19). Close neighbors (Figure 3.19A and B) with a score of 92% similarity cause severe phenotypes. The neighbors (Figure 3.19A, C, and D) all score 96% similar (Table 3.6) and have a phenotype where portions of the hypocotyl and or root have localized stains.

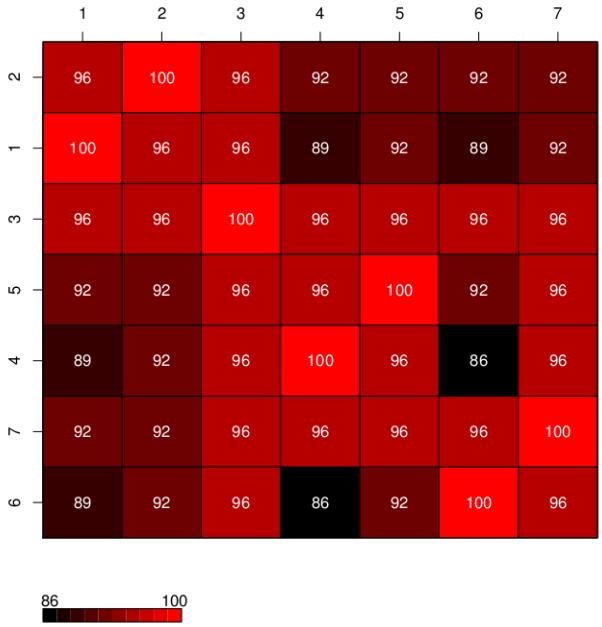

Figure 3.18: Heat map showing the distance between clade I compounds in the clickable collection, distances were derived from the Tanimoto atom pair similarity metric in the R package fmcsR. Values shown are the Tanimoto value *100, with a score of 100 indicating 70 % similarity.

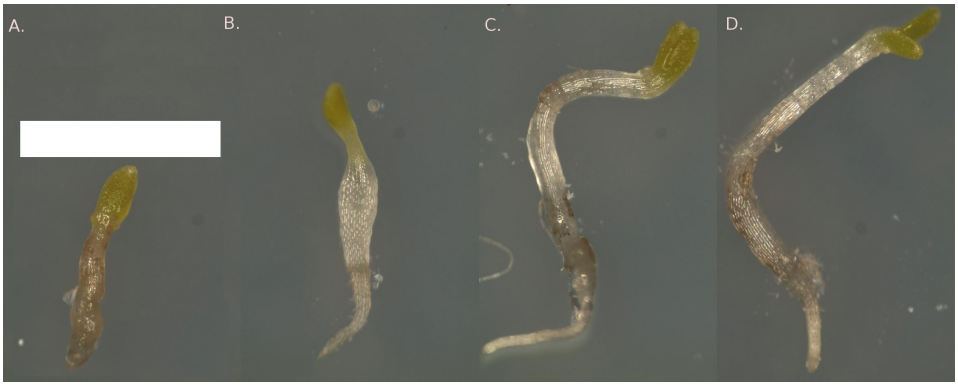

Figure 3.19: Phenotypes of clade I compounds in the clickable collection. A. CLK029D06. B. CLK029C05. C.CLK029G05. D. CLK029B05. All images at the same scale with a scale bar shown indicating 2μm.

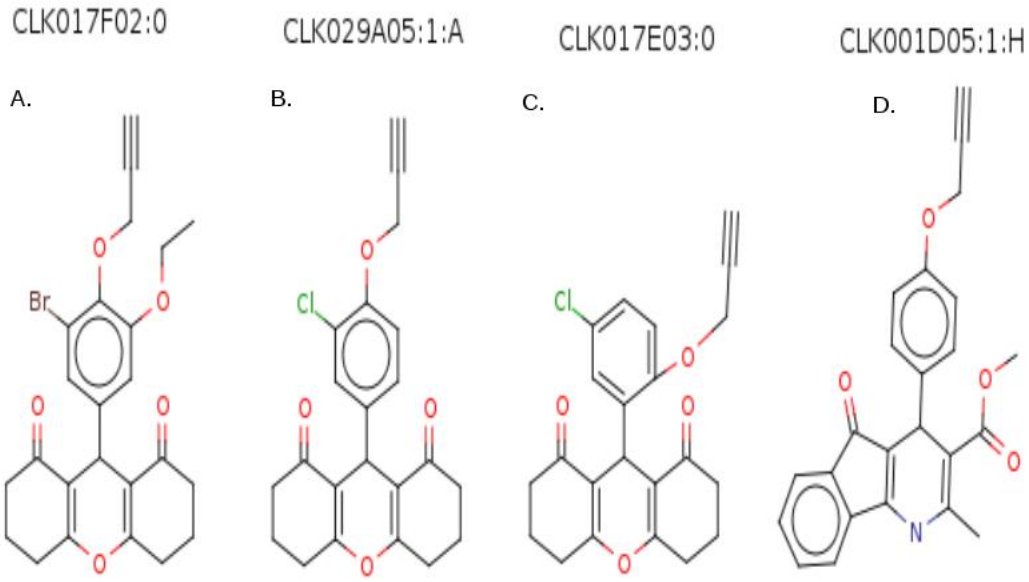

Figure 3.20: Clade Lii compounds with bioactivity annotation.

3.4.3.5 Clade L compounds

Clade L compounds (Figure 3.20)-(Figure 3.21) and (Table 3.6) also cause exaggerated apical hooks (Figure 3.23) despite having different core structures. The compound CLK017F02 is inactive and the compound CLK006G02 is 47% similar and causes an exaggerated apical hook (Figure 3.23A and B). Nearest neighbors to CLK006G02 with lower scores of 44 and 40% caused similar apical hook defects (Figure 3.23C and D)(Table 3.6). This group was not considered further.

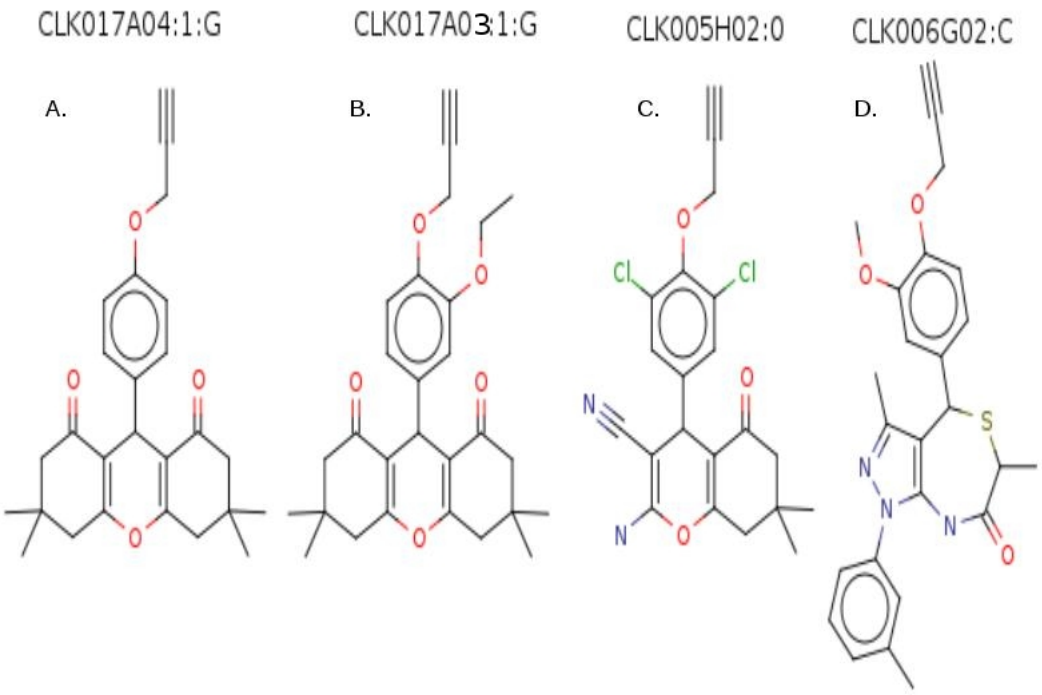

Figure 3.21: Clade Li compounds with bioactivity annotation.

	ID
4.	CLK001D05
3.	CLK017E03
7.	CLK005H02
2.	CLK029A05
6.	CLK017A03
5.	CLK017A04
8.	CLK006G02
1.	CLK017F02

Table 3.7: Table showing the Tanimoto similarity coefficients for clade L compounds. The similarity between compound pairs is shown with compound identifier (CLK ID) and value as percent. The matrix has been sorted to match the compound ID order of the matching cladogram on the next page.

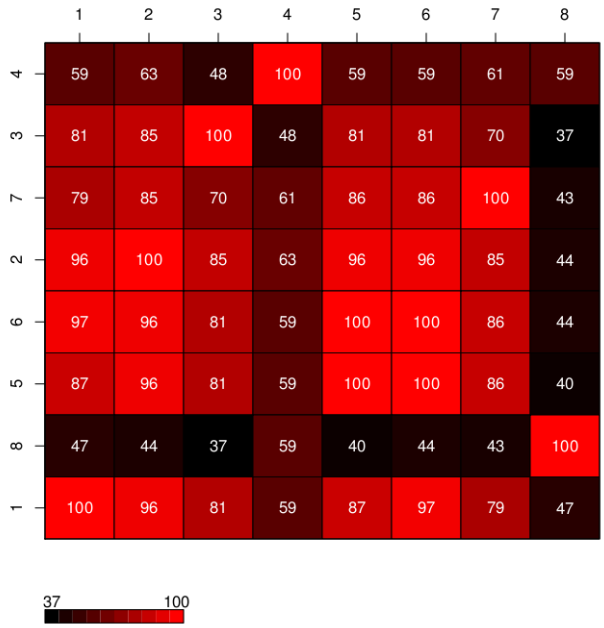

Figure 3.22: Heat map showing the distance between clade L compounds in the clickable collection, distances were derived from the Tanimoto atom pair similarity metric in the R package fmcsR. Values shown are the Tanimoto value *100, with a score of 100 indicating 70 % similarity.

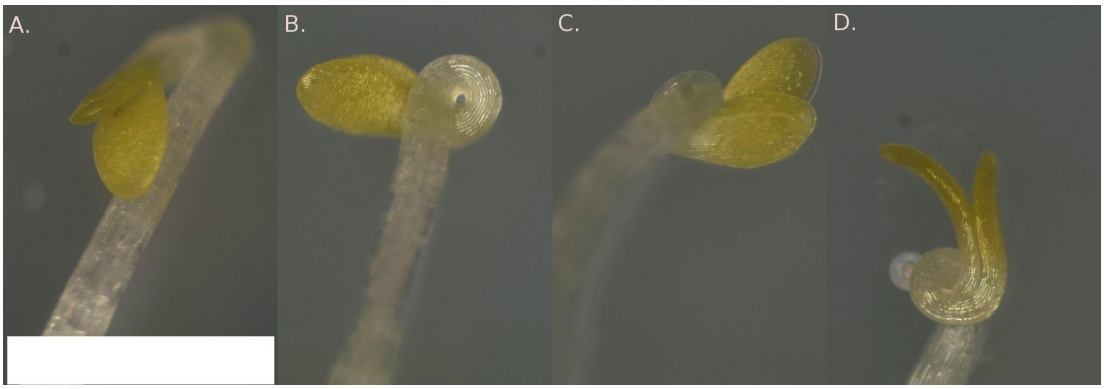

Figure 3.23: Phenotypes of clade L compounds in the clickable collection. A. CLK017F02. B. CLK006G02. C. CLK017A03. D. CLK017A04. All images at the same scale with a scale bar shown indicating 1μm.

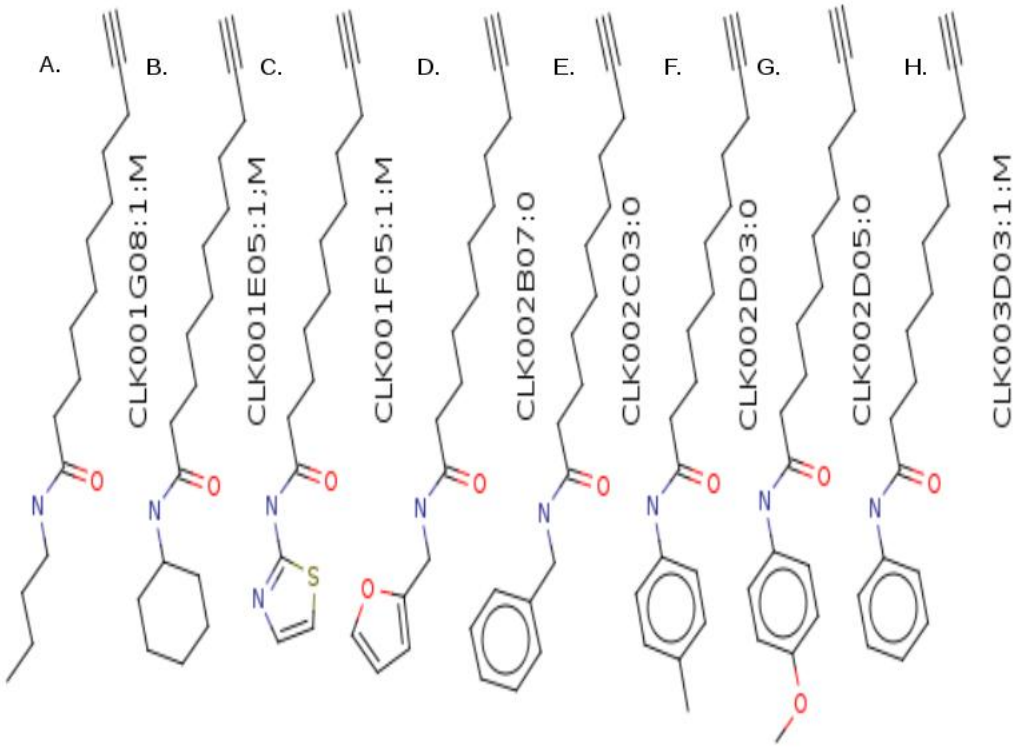

Figure 3.24: Clade M compounds with bioactivity annotation.

	ID
2.	CLK002D05
1.	CLK003D03
3.	CLK002D03
6.	CLK001G08
5.	CLK002B07
8.	CLK001E05
4.	CLK002C03
7.	CLK001F05
9.	CLK002F05

Table 3.8: Table showing the Tanimoto similarity coefficients for clade M compounds. The similarity between compound pairs is shown with compound identifier (CLK ID) and value as percent. The matrix has been sorted to match the compound ID order of the matching cladogram on the next page.

3.4.3.6 Clade M compounds

The last clade in this analysis clade M (Figure 3.24)(Figure 3.25) causes severe germination inhibition (Figure 3.26). The core MCS of this clade (Table 3.8) has been identified in the literature in relation to *Arabdiospsis* growth and germination and is discussed in Chapter 4.

3.4.4 Phenotype Fingerprint approach for phenotype annotation of the clickable library

3.4.4.1 The phenotype fingerprint approach to library annotation

The library was annotated using a phenotype bin approach inspired by the Masters thesis of Freeman Chow [65] since a number of clustering techniques can be performed on the resultant binary matrix. In addition, our data may be appended to a larger matrix downstream as information becomes available. Images were analyzed manually to annotate each compound with a binary score for 26 phenotype bins (Figure 3.27) (Table 3.1). The resulting 26 character phenotype score, hereafter called the

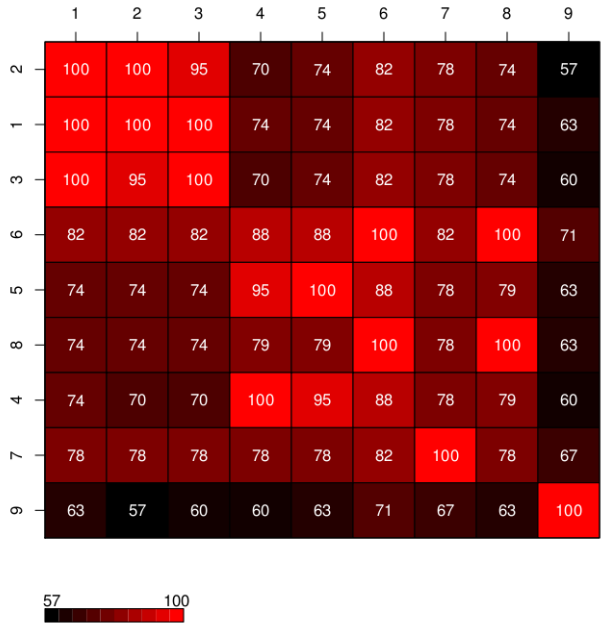

Figure 3.25: Heat map showing the distance between clade M compounds in the clickable collection, distances were derived from the Tanimoto atom pair similarity metric in the R package fmcsR. Values shown are the Tanimoto value *100, with a score of 100 indicating 70 % similarity.

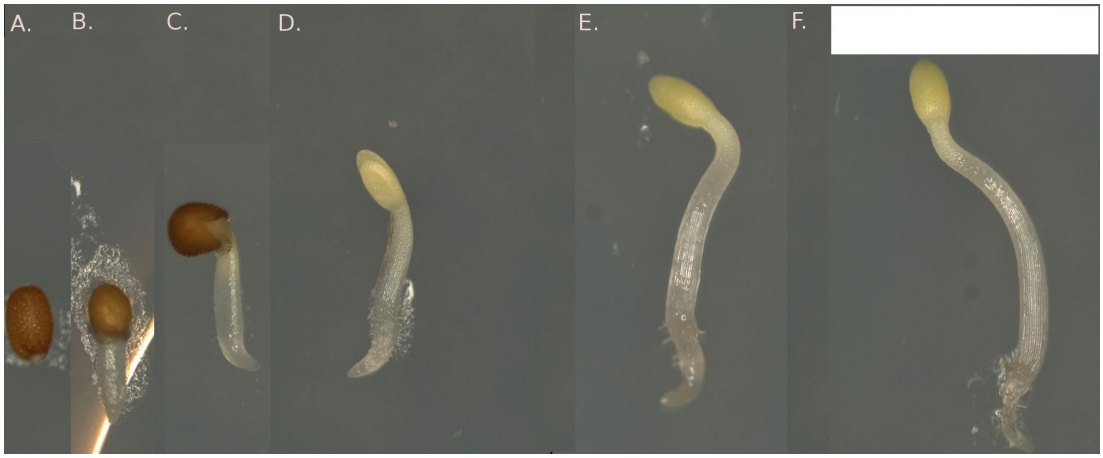

Figure 3.26: Germination effects of Clade M germination inhibitors. A. CLK003D03 B. CLK002B07 C. CLK001G08. D. CLK001F05 E. CLK002D03 F. CLK002C03 . All images at the same scale with a scale bar shown indicating 2μm.

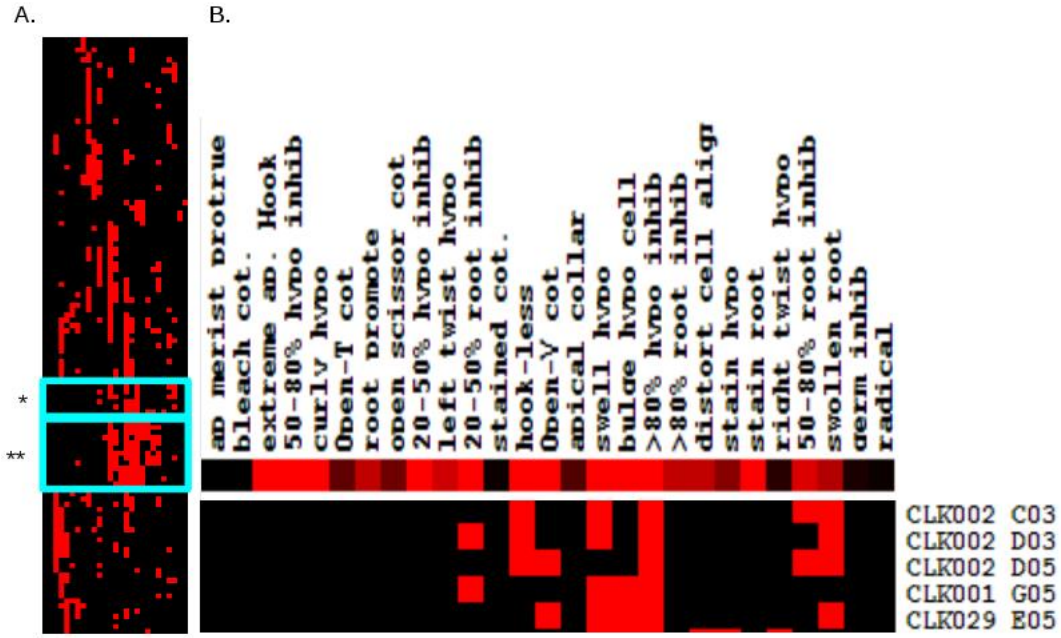

Figure 3.27: A complete cladogram of phenotype fingerprint (PTFP) vectors collected for bioactive acetylenes. B. The inset cyan box with a single astrisk indicates three of the clade M germination inhibitors (CLK002C03, CLK002D03, CLK002D05), shown here all displaying root problems and greater than 80% hypocotyl inhibition shown with red color. The double astrisks indicate a block of compounds that cause swelling and disordered cells, shown in the next figure.

PhenoType FingerPrint (PTFP), was used to perform hierarchical clustering based on a weighted category scheme (Figure 3.27). As expected, the PTFP approach successfully grouped phenotypically similar compounds, such as members of clade M, shown as a zoom of the total cladogram (Figure 3.27B).

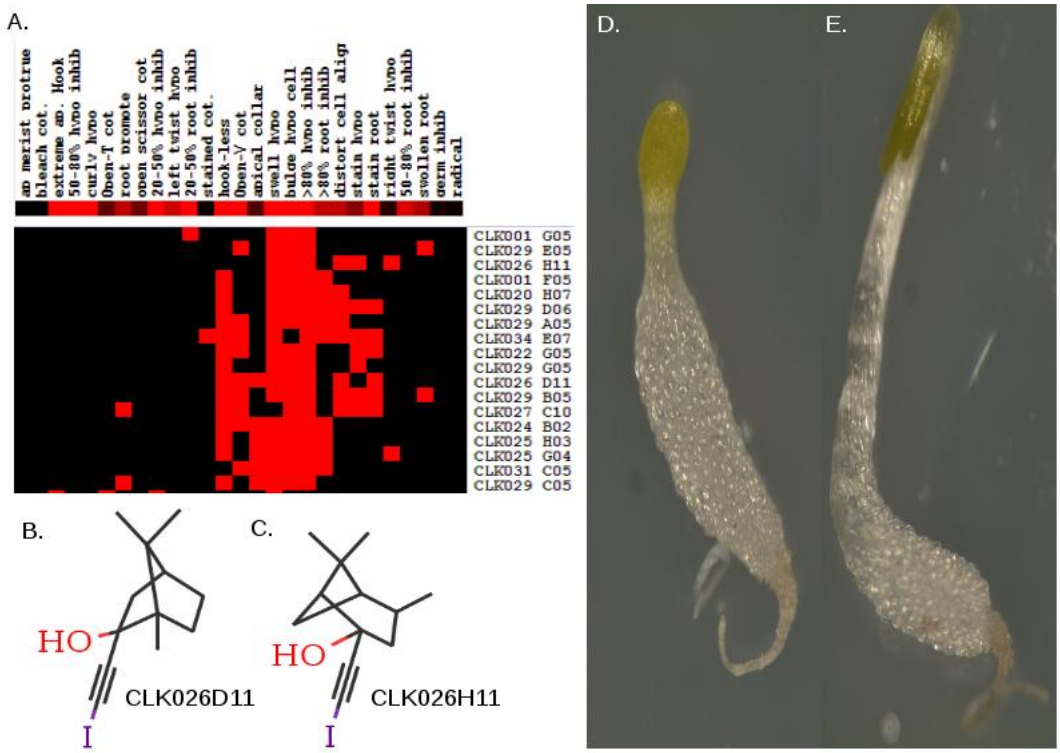

Figure 3.28: A block of compounds with similar PTFPs causing swelling and disordered cells. B-C. CLK026D11 and CLK026H11 cause extreme swelling phenotype at low doses. D-E. Swelling phenotype for CLK026D11 and CLK026H11, both Clade O members, respectively shown at identical magnification to previous clade M photos.

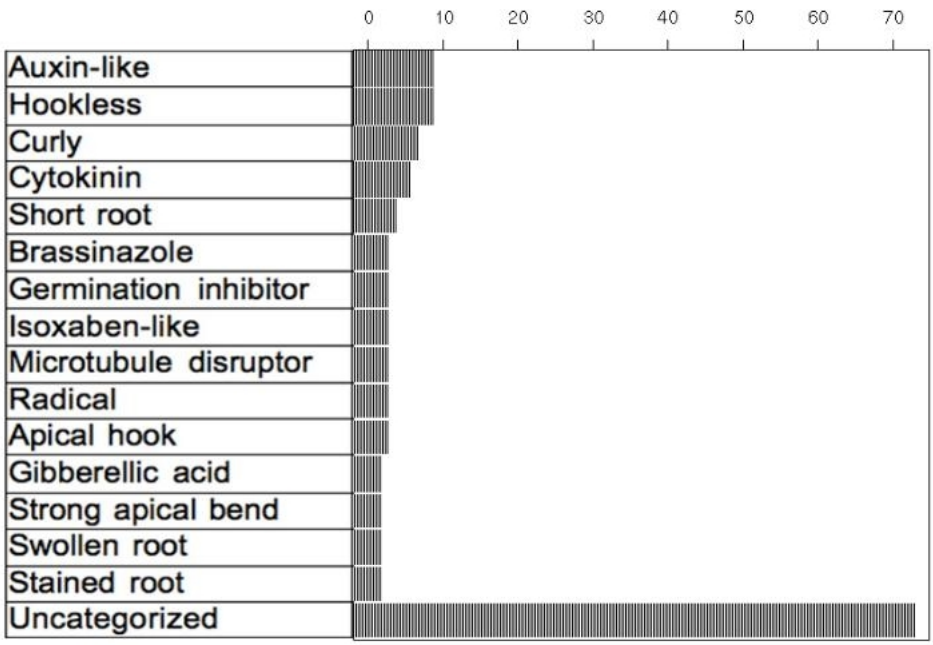

Figure 3.29: A grouping of classes of bioactive terminal acetylenes found in the clickable library. Data is based on the PTFP, comments on primary chemical screening data, and comparison of phenotypes and compounds to the literature. Values shown are percent per category for total number of bioactives. The largest category is uncategorized hits.

3.4.4.2 PTFP identifies a block of compounds potent to cause bulging and disordered cells and lead to severe inhibition of hypocotyl and root development

PTFP identifies a block of compounds potent to cause bulging and disordered cells and lead to severe inhibition of hypocotyl and root development (Figure 3.28A compared to D and E). Two interesting compounds in this block, CLK026D11 and CLK026H11 were camphor-like compounds possessing iodoacetylenes (Figure 3.28B-C). These compounds were serendipitously included into our library because the connection matrix of the iodoacetylene was similar enough to a terminal acetylene for inclusion.

3.4.4.3 Phenotype spectrum of bioactives in the clickable library

A large number of compounds produced auxin or auxin-like phenotypes such as the apical collar with swollen and short hypocotyl and a very short root (Figure 3.29), (shown as a frequency distribution). The prevalence of a spectrum of auxin, cytokinin, and isoxaben-like phenotypes is evidence that our library of terminal acetylene small molecules was suffciently diverse and bio-orthogonal in the sense that the terminal acetylene alone did not appear to cause a phenotype bias.

Part II

3.5 Abstract

We proposed to discover small molecule compounds that were both fluorescent and tagged for biochemical follow up. Therefore, we were uncertain which leads would serve as best candidates for fluorescent probes and would retain bioactivity after derivitization with an alkyl diazirine for covalent labeling. A systematic approach was taken to test the amine azide building block and a panel of eight fluorophores with the clickable library, thus producing 32,016 reactions. After screening, the whole set or reactions positive candidates were prioritized for specific investigation paradigms (Chapter 4: Phenotypes of Bioactives).

3.6 Materials and Methods

3.6.1 Validation of click products using analytical RP-HPLC

RP-HPLC using an Agilent Technologies 1200 Series HPLC with a Zorbax RX-C8 5μm column using a gradient elution from 5% to 95% acetonitrile containing 0.05% formic acid. Prior to injection analyte was dissolved in either 99μL of 40% acetonitrile or methanol. When precipitation was observed the sample was sonicated. The retention of the parent acetylene, residual dye-azide, and new peaks was used to determine product formation and if possible reaction efficiency.

3.6.2 Validation of click products using preparative RP-HPLC

Preparative HPLC fractionation was performed in duplicate on re-synthesized compounds using aqueous solvents of 5% methanol and 95% methanol. For injection 15μL of reaction was placed into 85μL of methanol and injected. Fractionate was collected into 48 wells of a 96-well plate, split into two sets. One set was stored directly

stored at -20° C for mass spectrometry validation. The other set was dried *in vacuo* using a Labcono centrifugal evaporator at 35-40° C for several hours and dissolved in 2μL of DMSO and growth media was then added to the well along with seed, as previous.

3.6.3 Validation of bioactive fractions by ESI-MS

ESI-MS was provided by University of California Department of Chemistry open access mass spectrometry services using an Agilent ESI-TOF-MS.

3.6.4 Dose curve of TBTA variants

A dose curve was performed with 1,5, 15, 25, 50, 75, 100 μM doses using TBTA-(CO2Me)3, TBTA-(CO2H)3, and TBTA-(CO2Na)3 on *Arabidopsis* grown in single well plates (NUNC) grown vertically. Measurements and photographic procedures are detailed in Materials and Methods of Chapter 4: Phenotypes of bioactives.

3.7 Results

3.7.1 Bivariate graphs of hypocotyl and root growth values using the bagplot

Previously we highlighted the features of individual plants displaying phenotypes by photography. It is evident the specifics of each phenotype has an influence on hypocotyl and root length at the very least. Observe these values shown in a simple scatter plot of DMSO treated seedlings from three replicate plates (Figure 3.30A). Even if we have no *a priori* knowledge of the cause of a phenotype we can observe the overt characteristics through bi-variate analysis. A multi-variable analysis would offer a finer description of each phenotype but was too costly. Therefore, we chose to evaluate

quantitative measurements of hypocotyl and root growth to determine the bioactivity of several compounds on whole plates of *Arabidopsis* plants. Hereafter, the bi-variate graphs are visualized using a plotting technique called the bagplot [70].

Visualization of values in the bagplot are shown for three replicate plates treated with DMSO (Figure 3.30B). Three transparent layers have been used to partition values to the replicate data set. The darker blue circle called a baghull indicates values in accordance with a 95% confidence interval. The red asterisk in the center of the baghull indicates the mean value for both variables of hypocotyl and root. Outliers from the dataset are shown as red dots and red lines called whiskers connect data points to the mean for the merge of the previous data set (Figure 3.30C). For instance, Comparison of control treatments with treatments with either TBTA-(CO2H)3 (data points shown in violet) or TBTA-(CO2Na)3 (data points shown in green) (Figure 3.30D) demonstrates that TBTA-(CO2Na)3 is less bioactive than TBTA-(CO2Me)3 even though the seedlings appear (by eye) morphologically similar (not shown). In this work we will refine our description of bioactivity for select compounds using these style of bagplot visualization.

3.7.2 Testing TBTA catalyst variants for bioactivity

A dose curve analysis revealed the TBTA-(CO2Na)3 catalyst was the least potent to cause bioactive effects and inhibit hypocotyl or root growth (Figure 3.31A-F, where F is from previous panel for comparison). The control values for hypocotyl and root (125,100) are achieved for lower doses of 1, 5,and 15μM for the TBTA variants, as seen with the position of the red astrisk (Figure 3.31A-C). The higher doses of 25, 50, 75 and 100 μM are more potent to cause a decrease in values to (100,100) for TBTA-(CO2H)3 (Figure 3.31D) and (100,75) for TBTA-(CO2Me)3 (Figure 3.31E) compared

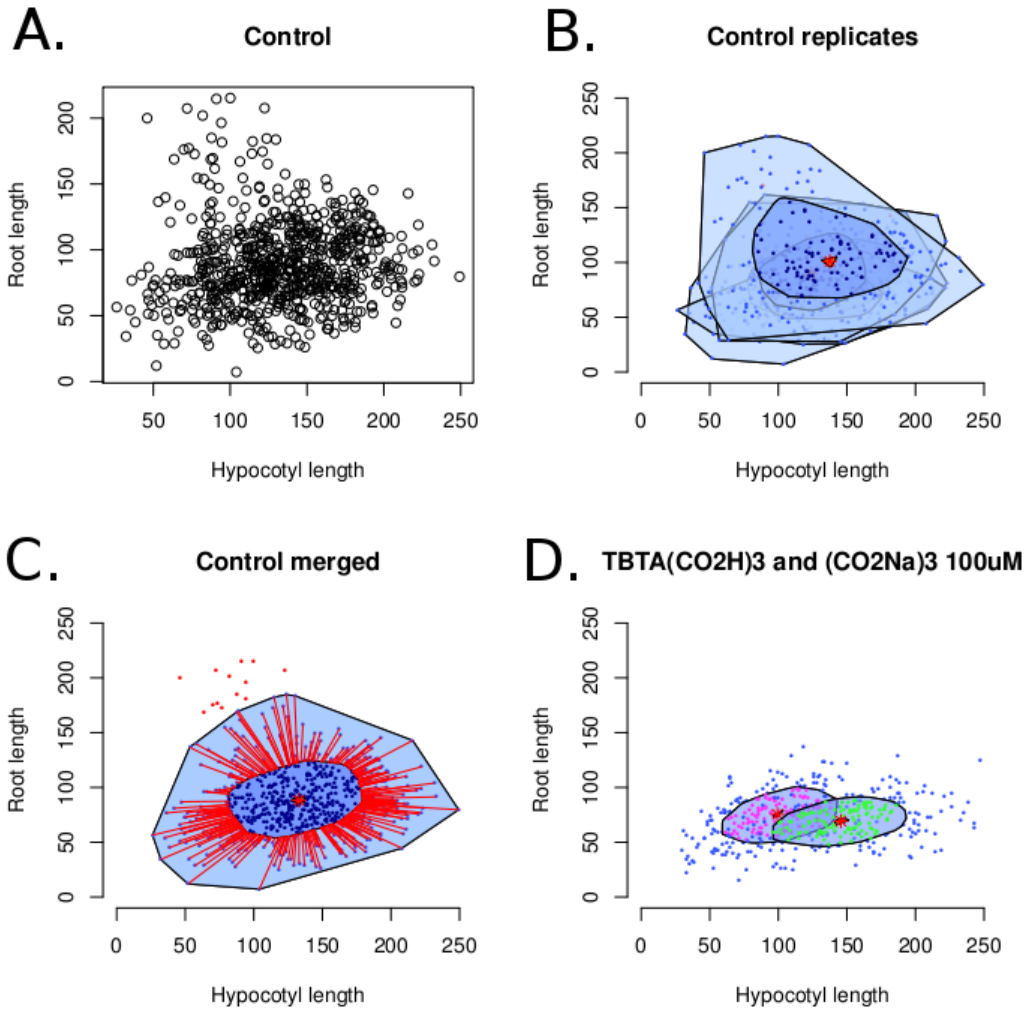

Figure 3.30: Bivariate analysis of hypocotyl and root values of *Arabidopsis*. A. Scatterplot of hypocotyl and root values for DMSO treated seedlings (data from three replicates). B. The same three replicates partitioned to transparent layers of a bagplot with darker color indicating the baghull or region of 95% confidence. C. Replicate values merged into the same bagplot showing outliers in red and a modified baghull as a result of the averaging. D. Baghull demonstrating bioactivity differences between TBTA-(CO2H)3 (shown in violet) versus TBTA-(CO2Na)3.

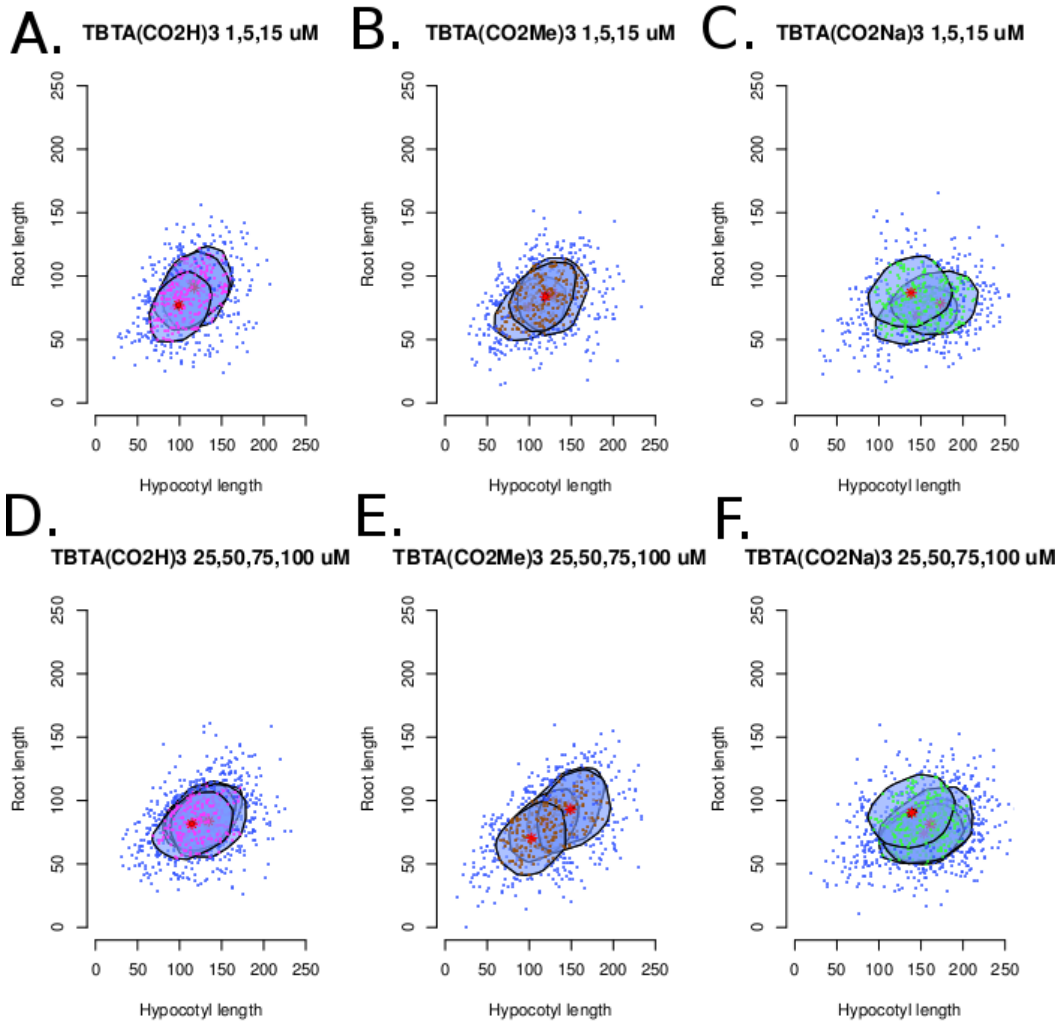

Figure 3.31: Bivariate bagplot analysis of TBTA-(CO2H)3, TBTA-(CO2Me)3, and TBTA-(CO2Na)3. A-F. Respective TBTA treatment with the dose shown as transparent layers, with higher doses shown on the foreground layer for all bagplots.

to (125,100) for TBTA-(CO2Na)3.

3.7.3 Validation of fluorescent click reactions by bioactivity guided fractionation

The analysis of bioactivity of the terminal acetylene library and reagents such as TBTA-(CO2Na)3 was a prerequisite to our intended investigation which was to identify bioactive fluorescently tagged molecules. We previously described several libraries

derived from high throughput reactions with acetylenes and azides using ligand catalyzed click chemistry (Figure 2.31) with a panel of fluorescent dye azides (Figure 2.24A-D) and one 1-azido-beta-D-glucopyranose. The unpurified click reactions were screened in duplicate 96-well format to identify bioactive click products in *Arabidopsis thaliana*.

In the biological screen we were able to buffer the copper salts necessary for the click reaction with the addition of μM EDTA as a chelator and utilized the TBTA-(CO2Na)3 catalyst. A number of bioactive reactions were discovered in this fashion, but it was not known if the bioactive result was from the original acetylene, a new product, or undesired side products. Analytical HPLC and bioactivity guided fractionation were used to identify reactions with true bioactive 1,2,3-triazole products, as evidenced by mass spectrometry (Table 3.9).

Reaction ID	PP	PC	FM	CP	CC
CLK003G05_Dansyl	1	0	FG	26*	26*,3
CLK003H03_Dansyl	1	0	F3		34-37*,41
CLK004F10_C2	1	0	FG	2,24	2
CLK015C07_Glc	0	1	F5	n/a	26-29,33-35
CLK018G02_FITC	0	1	F6	n/a	25,28,43
CLK018C11_Glc	0	1	F4	14-15,25-28	19*,20,23-26
CLK015H06_C2	1	1	F3	6,23*,27*	13,26-28,27
CLK016D02_FITC	1	1	F3	2,24-26,31*	1,6
CLK017F11_C3	1	0	F3	2,3,24	2,39
CLK018G02_FITC	1	0	F3	1,2,44-45	15*,44
CLK019E08_FITC	1	1	F3	17*,19,32	2,21
CLK019E08_C3	1	1	F3	18	1-3,31,*40
CLK020A07_C2	1	0	F	1-3,36,37	2,21,35*-37
CLK021C05_Dansyl	0	1	F3	2	2,21,35*,37
CLK022F02_Dansyl	1	1	F3	2	25,35*,37,42
CLK023C04_Dansyl	0	1	F3	0	0
CLK024F02_Dansyl	0	1	F3	18	1-3,31*,40
CLK026G08_Dansyl	0	1	F3	1-3,36,37	2,21,35*-37
CLK026G08_Glc	0	1	F3	2	25,35*,37,42
CLK029B07_Dansyl	0	1	F3	2,41	36*
CLK031C06_Glc	1	1	F	3	2,9*,34*,40
CLK034F10_Dansyl	1	1	F3	2	2,40

Table 3.9: Table showing the results for the fractionation screen of candidate bioactive click reactions. The column headers are product parent (PP), product conjugate (PC), fractionation method (FM), comment on phenotype of parent (CP) with fraction indicated or phenotype of the conjugate (CC). Where * denotes a fluorescent fraction.

3.7.4 Analytical HPLC and fractionation of bioactive click reactions

Candidate reactions were studied by analytical HPLC for product formation and fractions corresponding to the bioactive product(s) were isolated by fractionation.

3.7.5 A demo click reaction with a bioactive product and inactive analog

The bioactive compound CLK042A09 dansyl block A, was identified from a bioactive click reaction with an inactive parent acetylene (shown at similar magnifica-

tion) (Figure 3.32). Here, CLK042A09 dansyl block A, was fractionated and produced bioactive fractions unique to the parent acetylene CLK042A09 (Figure 3.32). These compounds were not identified in the primary screen of terminal acetylenes. A typical click reaction with the terminal acetylene CLK042A09 and the nearest neighbor CLK039G03 with dansyl block A recapitulated bioactivity observed during our screen Figure 3.32A-E). When we used DMSO alone or CLK042A09 there was no significant change in phenotype Figure 3.32A and B). When we used dansyl block A with CLK042A09 and no TBTA-(CO2Na)3 catalyst we observed no product and no phenotype (Figure 3.32C). Addition of the TBTA-(CO2Na)3 catalyst to the click reaction caused a novel phenotype corresponding to the expected product (Figure 3.32D). The nearest neighbor CLK039G03 failed to produce a phenotype from a click reaction with dansyl block A although the reaction produced the expected product (Figure 3.32E). In the next chapter (Chapter 4: Phenotypes of Bioactives) we will explore the product CLK042A09 dansyl block A.

3.8 Summary

3.8.1 Bioactives from failed click reactions

A majority of false positive hits produced interesting phenotypes through a synergy between the mild bioactivity of TBTA-(CO2Na)3, residual dye-azide, and/or terminal acetylene. Hits were scored and prioritized after analytical and preparative HPLC especially when the mass ion for the predicted click product was observed (Figure 3.33). When bioactivity was due to residual terminal acetylene the fractionation experiment yielded identical bioactive fractions for both the acetylene and the reaction. In some cases synergy alone was not sufficient to explain the reason for a false positive

Figure 3.32: Representative click reaction to demonstrate the bioactivity of a click reaction from the terminal acetylene CLK042A09 and the amine azide block A. A. DMSO. B. Block A. C. Reaction with no TBTA-(CO2Na)3 added. D. A successful bioactive click reaction with TBTA-(CO2Na)3, CLK042A09 and block A. E. The nearest neighbor CLK039G03 was not bioactive. All seedlings are shown at the same magnification.

phenotype. A number of cases can be explained where no major product was generated in the reaction, but new fragments with less retention on the reverse phase appeared. These were most likely break down products from the click reaction conditions.

Further investigation is necessary to elucidate the structure of a number bioactive click reactions that could not be structurally validated by mass spectrometry, but appeared by analytical HPLC and the fractionation experiment to be the expected fluorescent click product. This is true of CLK029B07 Dansyl which did not generate a mass ion in accordance with the expected molecular formula.

3.9 Discussion

At the onset of our investigation we chose to evaluate the bioactivity of 4,002 terminal acetylenes (clickables) by high throughput biological screening with *Arabidopsis*

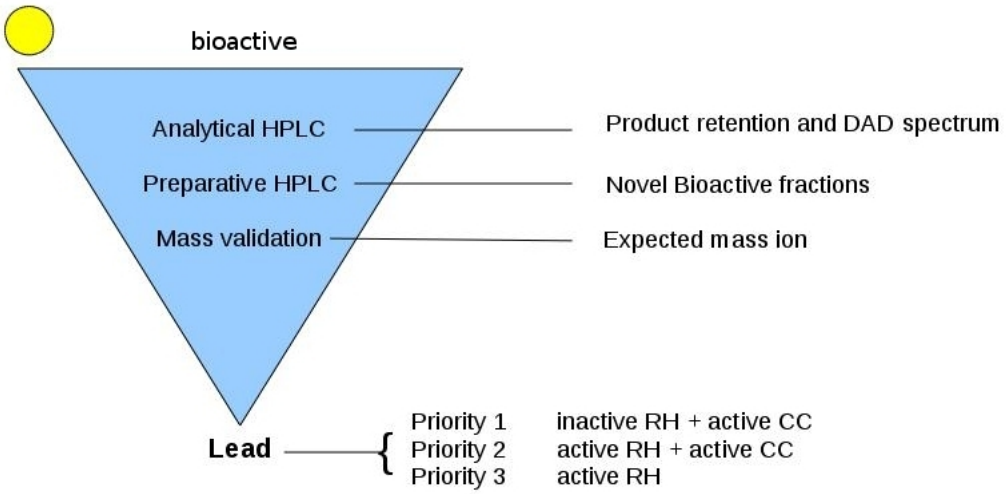

Figure 3.33: Lead prioritization scheme. Prioritization of hits for follow up was determined continually during the chemical genetic screen by collecting data for activity, HPLC product retention, and DAD spectrum. The highest priority was reserved for click reactions that produced bioactive fractions not attributable to the parent acetylene and had the expected mass ion in ESI/MS.

(Figure 3.4). A number of bioactive multiple common substructures (MCS) were identified (Figure 3.11)-(Figure 3.26), and the effects on *Arabidopsis* was presented briefly. We sought a means to use our clickable library for chemical genetics and for the combinatorial generation of our libraries using click chemistry (Figure 2.31). We hypothesized the TBTA-(CO2Na)3 ligand would be more charged than the original ligand, thus affecting permiability and this could be advantageous for screening reaction products directly. The effects of the TBTA-(CO2Na)3 were characterized as less than the acid counterpart as seen in the bagplot graphs (Figure 3.30). It remained to be determined if the permiability was decreased for TBTA-(CO2Na)3, but this was inconsequential for our purposes after the biological effects of this ligand were determined to be lessened (Figure 3.31F).

The fractionation of hits from the combinatorial libraries was a useful procedure to isolate pure fractions of compounds for bioactivity and mass spectrum analysis where necessary. One compound, CLK042A09 Dansyl, was determined to be the expected structure (Figure 2.16). We see clearly that the unpurified reaction between the

acetylene CLK042A09 and dansyl block A produces a phenotype in *Arabidopsis* when there are suitable conditions for the click reaction (such as inclusion of TBTA-(CO2Na)3) (Figure 3.32D).

3.10 Conclusion

Two terminal acetylene clades are interesting chemical genetic probes in there own right these will be explored in the begining of the next chapter (Chapter 4: Phenotypes of Bioactives). The clade M germination inhibitors (Figure 3.26) and the Clade O inhibitors of anisotropic growth (Figure 3.28B and C). Three probes were identified as bonifide bioactive click reaction products: CLK021C05 Dansyl, CLK024F02 Dansyl and CLK042A09 Dansyl. These were all priority 2 leads (Figure 3.33) as shown in our scheme since the acetylene fragments themselves had some bioactivity. These rare probes will also be examined further in the next chapter using bagplots. All other candidates were not considered further. We did not find any priority 1 leads. These priority 1 leads may potentially be found through the exploration of the chemical space available in the remaining ZINC80K or through the generation of novel probes. One method could be through the use of Ruthenium to catalyze the formation of the other regioisomer of the click reaction the 1,4,5-triazole. Another method could put the whole scheme on a molecular weight reduction to capitalize on earlier studies of fluorogenic click reactions with azido coumarin variants [71].

Chapter 4

Phenotypes of Bioactives

4.1 Abstract

We discovered a handful of bioactive terminal acetylenes from the chemical space available in the clickable library. These probes were potent to inhibit the hypocotyl and root growth of *Arabidopsis*. Bioactivity of Clade M and Clade O compounds was confirmed using assays appropriate to the specific phenotype. The clade M compounds were tested in a 24-well plate to confirm CLK003D03 was the most potent to inhibit germination of Col-0 and Ler *Arabidopsis* ecotypes. Bagplot analysis of the clade O compound CLK026D11 demonstrates it is potent as a volatile to inhibit the growth of hypocotyls and caused a radial swelling phenotype of the hypocotyl cell file, as visualized by scanning electron microscopy (SEM).

4.2 Introduction

Our rationale was to use *Arabidopsis thaliana* and classical genetics to determine the mechanism of action of novel probes discovered [5, 3]. The analysis included

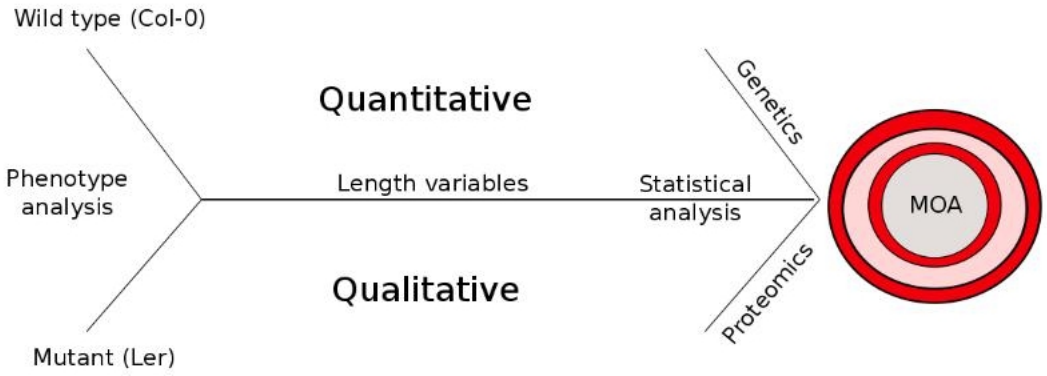

Figure 4.1: The end game of our forward genetic investigation is to determine the mechanism of action (MOA) for newly discovered probes. Two ecotypes Col-0 and Ler were used together in a phenotype analysis platform that integrates quantitative and qualitative length variables. The goal to produce robust phenotype descriptions that enable the identification of target allele(s) for novel bioactives, an investigation which will be described in detail (Chapter 5: Target identification efforts via EMS).

examination of several active compound classes using the two ecotypes Columbia-0 (Col) and Landsberg erecta (Ler)[66]. In this process qualitative and quantitative variables were collected and used in a combination of approaches to determine the Mechanism Of Action (MOA) [72]. Discussion and review of contemporary MOA approaches will be elucidated in the coming chapter (Chapter 4: Target Identification efforts with EMS) and are reviewed briefly here [73]. The objective of this chapter was to establish which probes were amenable to downstream examination.

4.3 Materials and Methods

4.3.1 Twenty-four well assay plates for the Clade M germination inhibitors

Standard 24-well assay plates were used to grow *Arabidopsis* at different doses of the clade M germination inibitors. Plates were prepared as indicated previously (Materials and Methods: Chapter 3).

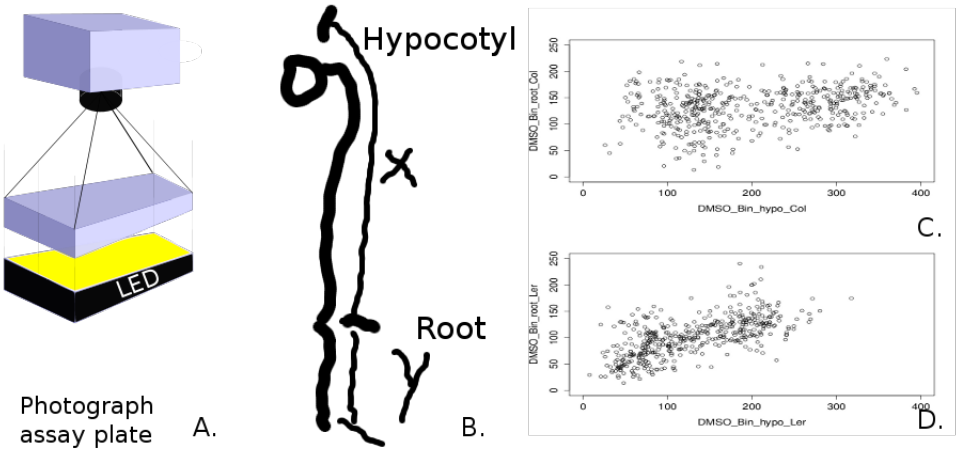

Figure 4.2: Two dimensional capture of vertically grown *Arabidopsis thaliana* by digital imaging. A. Photographs were taken of the assay plate from a fixed distance above plates transilluminated from below. B. Measurements of the hypocotyl (shown as x variable for x-axis) and root (shown as y variable for y-axis are conducted using ImageJ. C. Demonstration of typical control (DMSO 1% data plotted as a bi-variate x-y graph for hypocotyl and root length of Col and Ler seedlings.)

4.3.2 Preparation of assay plates for the testing of the bioactive volatile CLK026D11

Single well assay plates (NUNC) were grown vertically and taped together with different doses of 1, 5, 10, 15, 25 μM CLK026D11 and (1%) DMSO was used as control. Growth media and seed preparation was indicated previously (Materials and Methods: Chapter 3), except seedlings were placed onto the media on the fourth day or the first day of growth in dark. Plates were photographed on the seventh day using a digital camera and LED transillumination.

4.3.3 Quantitative bagplot: image acquisition, photography setup, and digital processing for length variables

Images for quantitative bagplots were acquired using a hand held 2.0 Mega Pixel digital camera. Assay plates were photographed using standard position and

distances for consistancy (Figure 4.2). Illumination was provided by an LED transilluminator placed directly underneath the assay plates. Photographs were measured by hand using the ImageJ polyline tool. Hypocotyl and root variables were exported from ImageJ and processed in R using the bivariate bagplot() call in the package aplapack.

4.3.4 Preparation of *Arabidopsis* for scanning electron microscopy

Arabidopsis seedlings prepared identically to the chemical genetic screen are imbibed for 4 days at 4° C and grown for 3 days in the dark. On the final day seedlings were transferred directly onto a Hitachi scanning electron microscope (SEM) specimen stage with adhesive and flash frozen immediately using isopentane chilled via liquid nitrogen. The sample was placed into the vacuum chamber of the SEM for electron scan and imaging.

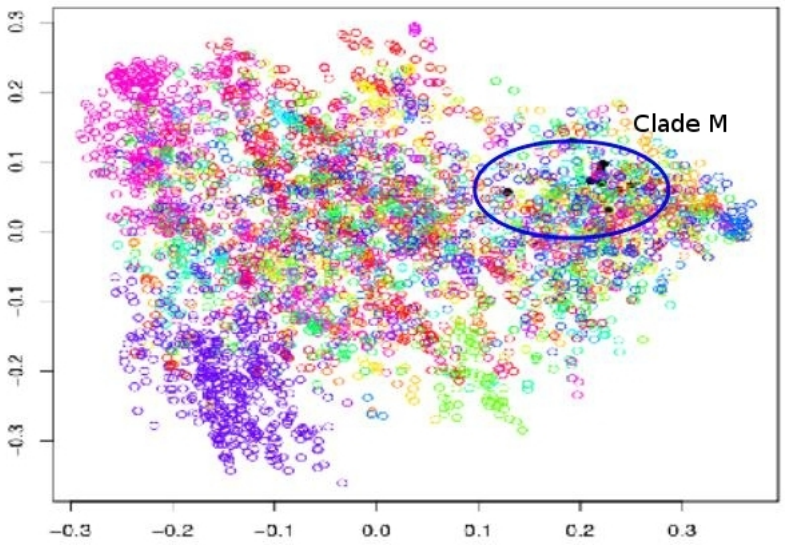

Figure 4.3: Two dimensional multi-dimensional scaling (2D-MDS) plot showing clade M germination inhibitors as black circles in the clickable library. Colors indicate compounds that group together and are 70% similar. The number of colors for plotting is limited as such the colors have been re-used on other clades. Blue circle inset shows clade M compounds in the plot.

4.4 Results

4.4.1 Revisiting the clade M germination inhibitors

The clade M germination inhibitors (Figure 3.24) are members of a clade of fatty acid n-acylethanolamine (Figure 4.3) in the clickable library [74, 75]. A Multiple Common Substructure (MCS) consisting of a unsaturated hydrocarbon chain of length (C8) terminating with ethanolamine linked heterocycles (Figure 3.24) was used as a query to identify related compounds in the clickable collection (Figure 2.30).

The clade M and analogs in the clickable collection and cited literature have been collected in this MDS space for comparison (Figure 4.4). Compounds are shown inset the MDS space with a line connecting to the coordinate corresponding to actual scaled distance.

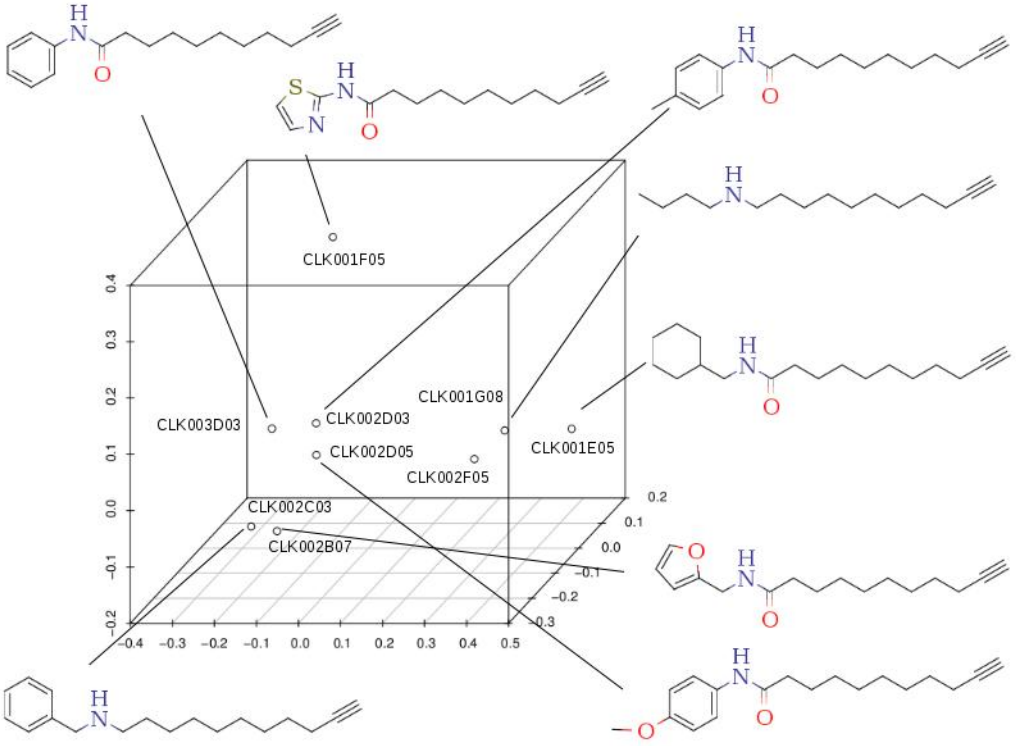

Figure 4.4: Clade M and analogs in the clickable collection.

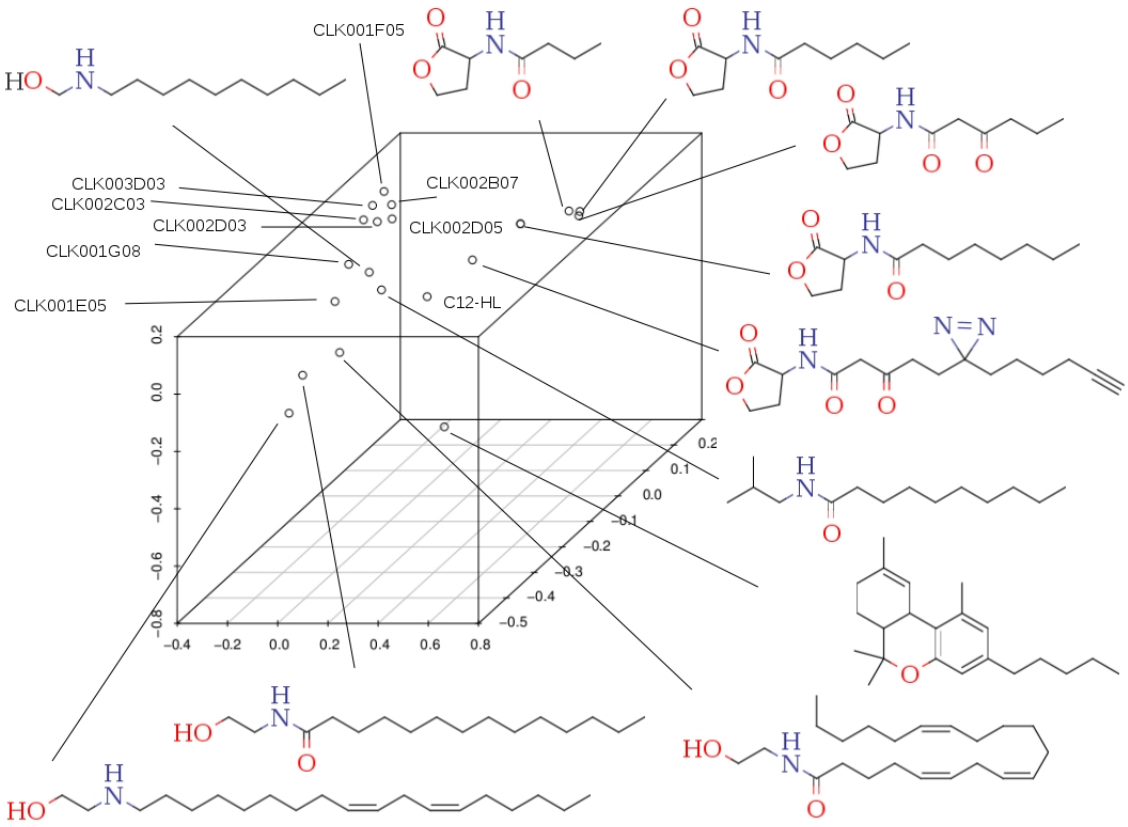

Figure 4.5: Clade M and analogs in the clickable collection and and relevant literature.

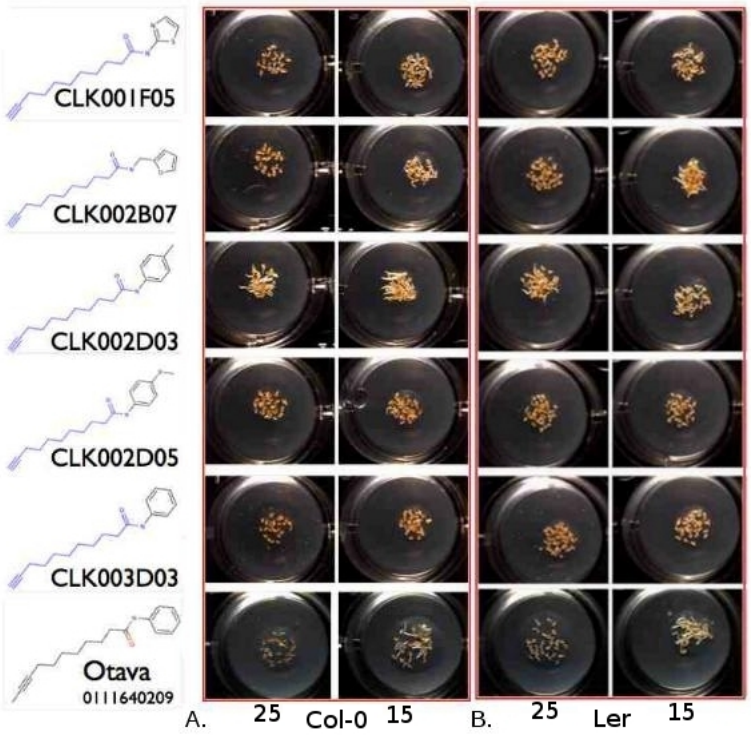

Figure 4.6: Dose curve of select clade M compounds differentiates CLK003D03 as a strong candidate at 25 μM and 15 μM. A. CLK003D03 is the most potent clade M compound to inhibit germination of Col-0. B. CLK003D03 is equally potent to inhibit the germination of the Ler ecotype

A number of related compounds have been mentioned previously in the literature in relation to *Arabidopsis* growth regulation. These compounds have been collected into an MDS space for inspection (Figure 4.5)[74, 76][75].

A detailed structure activity relationship was performed to determine the potency of clade M germination inhibitors (Figure 4.6 A-B). The clade M compounds were not active as fluorescent conjugates with our amine azide linker (data not shown), and were used directly as is. The compound CLK003D03 was used to generate resistant mutants in *Arabidopsis* and identify a causative allele (not described here).

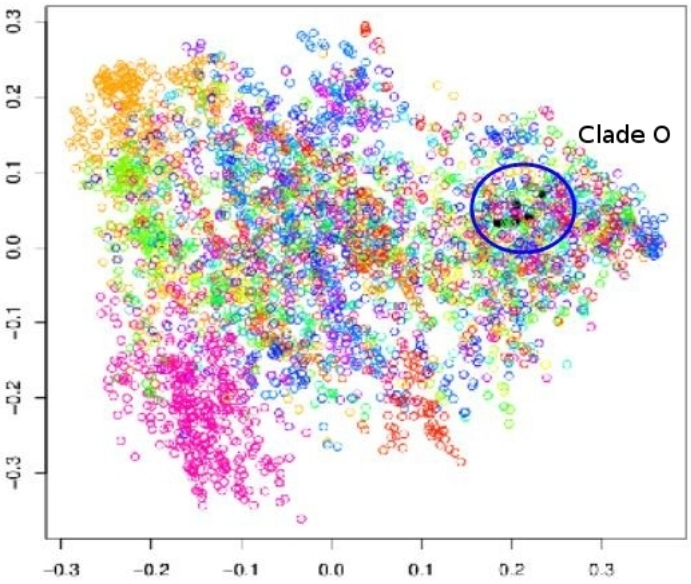

Figure 4.7: Clade of compounds related to CLK026D11 possessing a camphor-like carbon skeleton shown as black dots. Colors indicate compounds that group together and are 70% similar. The number of colors for plotting is limited as such the colors have been re-used on other clades.

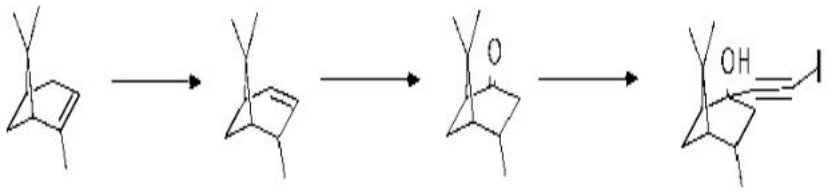

Figure 4.8: Reported synthesis of CLK026D11 from verbanone [77].

4.4.2 Clade N inhibitors of anisotropic hypocotyl growth

Clade N, with the core iodoacetylene member CLK026D11 was compared to compounds with similar pinane-like and adamantane carbon skeletons in the clickable library (Figure 4.7A, inset blue circle, see supplementary table). Despite possessing similar carbon skeletons or iodoacetylenes the compounds CLK026D11 and CLK026H11 distinctly caused the same phenotype (Figure 3.28). CLK026D11 is similar to compounds in natural product datasets, and this may be a consequence of the synthesis of CLK026D11 and CLK026H11 from verbenone and verbanone respectively (Figure 4.8A-D)[77].

4.4.2.1 CLK026D11 is a potent volatile inhibitor of anisotropic growth

Volatiles, such as ethylene, can function as growth regulators in plants [61]. Preparation of the iodoacetylenes CLK026D11 and CLK026H11 describes the compounds as unstable, although 1HNMR analysis confirmed the structure as reported [77]. Volatility of CLK026D11 was suspected due to a low molecular weight and the reason for observed variation on the edges of treatment and control plates. An experiment was conducted to demonstrate CLK026D11 was potent at a distance. *Arabidopsis* was grown in media vertically and another plate containing compound was placed on top to exchange head space.

Growth of plates containing DMSO alone showed typical bagplot variance whereas CLK026D11 was potent to inhibit hypocotyl and root growth down to 1 μM (Figure 4.9A and B). Iodine was used as a control as it was suspected that the iodoacetylene CLK026D11 may degrade to release iodine. Nonetheless, iodine at all concentrations was inactive, and did not cause a phenotype by exchanging head space with the con-

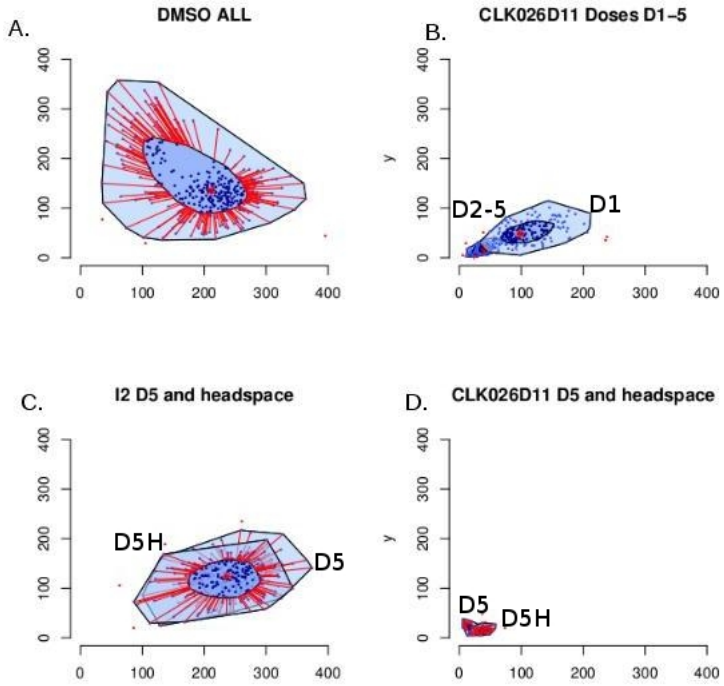

Figure 4.9: CL026D11 is bioactive as a volatile. A. Merge of DMSO. B. Increasing doses of CLK026D11 causes a significant growth defect as low as 1μM. C. Iodine and the headspace of this plate does not cause growth inhibition. D. The head space of CLK026D11 (D5H) causes similar growth defects as plants grown directly on the same dose. Doses are denoted D1-5 for 1, 5, 10,, 15, 25 μM.

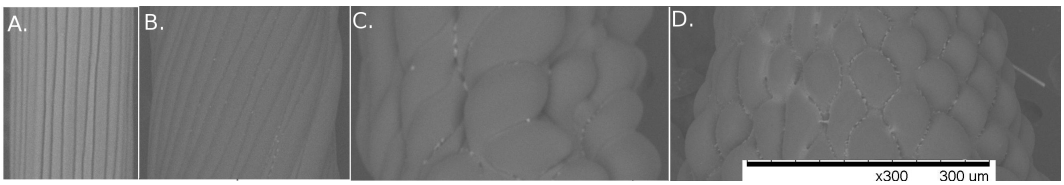

Figure 4.10: CLK026D11 causes anisotropic hypocotyl growth phenotypes with a handedness similar to chuboxypyr. A. Columbia-0 seedling control treatment (1% DMSO). B. 25μM of chuboxypyr causes right handed twisting of cell files in the hypocotyl. C. 1 μM dose of CLK026D11 causes a similar right handed twisting defect with enhanced severity. D. 5 μM of CLK026D11 causes loss of cell polarity and leads to swollen and bulging cells. Scale bar shows 300 μ meters.

trol plate (Figure 4.9C). CLK026D11 was potent to produce identical phenotypes in seedlings on the control plate when place a few centimeters away at 25 μM doses (Figure 4.9D).

4.4.2.2 Cellular features of the perturbation caused by CLK026D11

Low doses of CLK026D11 and structurally dissimilar chuboxypyr [66] induced right handed helical hypocotyl twisting as visualized by scanning electrom microscopy (SEM) (Figure 4.10B and C), whereas higher doses disrupted etiolated ansiotropoic hypocotyl expansion to result in radially expanding cells rather than typical expansion lengthwise against gravity (Figure 4.10D). These growth features were known to be caused by a number of structurally dissimilar compounds such as oryzalin [78, 79], taxol [80, 81], and morlin [68] with modes of action related to microtubule homeostasis [185] and cellulose synthase function (CeSA) [68].

4.5 Summary and Discussion

4.5.0.3 CLK026D11 is unique and potent bioactive but could not be used to reveal mechanism of action (MOA)

Exploration of the swollen cell phenotype caused by CLK026D11 and CLK026H11 (Clade O) is of significance to elucidate molecular and cellular events that lead to the adaptive evolution of skotomorphogenesis and etiolated growth in a number of angiosperms [82, 83]. The clade N compounds phenocopy mutants characterized as microtubule proteins [69] and can be examined further with mutant analysis.

Clade M is has probes that come with a literature history, thus when evidence is discovered there is already information available for comparison. We will see how one of two clade M probes can be used to find the responsible allele in Chapter 5. Target identification efforts with EMS.

The bioactivity of the CLK026D11 was interesting but was difficult to determine MOA in *Arabidopsis* since the compound was volatile in the growth media and caused a significant amount of phenotype variation in large plates. CLK026D11 did participate in the click reaction to give the expected product, but was not bioactive when reacted with the fluorophore azide building blocks tested. Therefore, CLK026D11 was not useful, in our scheme, as a tagged compound and was discarded.

4.6 Abstract

Chemical genetic screening of the combinatorial click reactions with different building blocks and dye-azides (Chapter 3) enabled us to find several 1,2,3-triazoles to study in detail. Four compound classes and selected nearest neighbors as dansyl block A conjugates were further evaluated. By the use of quantitative bagplots phenotypic

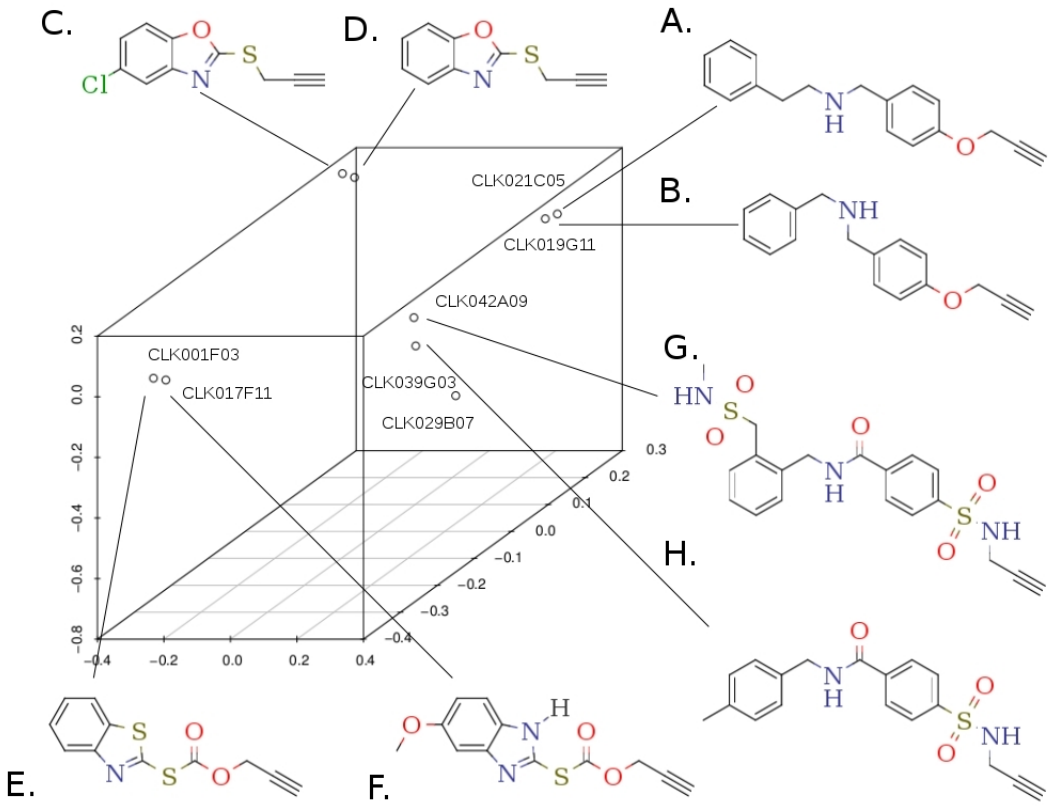

Figure 4.11: MDS of bioactive acetylenes explored quantitatively in this section, shown as parent acetylenes. A. CLK021C05. B. CLK019G11. C. CLK024F02 D. Proposed de-chlorinated CLK024F02 E. CLK001F03. F. CLK017F11. G. CLK042A09. H. CLK039G03.

trends were identified to evaluate the diversity of growth responses in two *Arabidopsis* accessions. The bioactives from click reactions with CLK021C05, CLK024F02, and CLK042A09 were grouped separately since they are distinct compound classes. In fact, through comparison of bagplots we can determine that each of the aforementioned compound classes inhibit hypocotyl and root growth in a distinct way as can be seen by the bagplot characteristics.

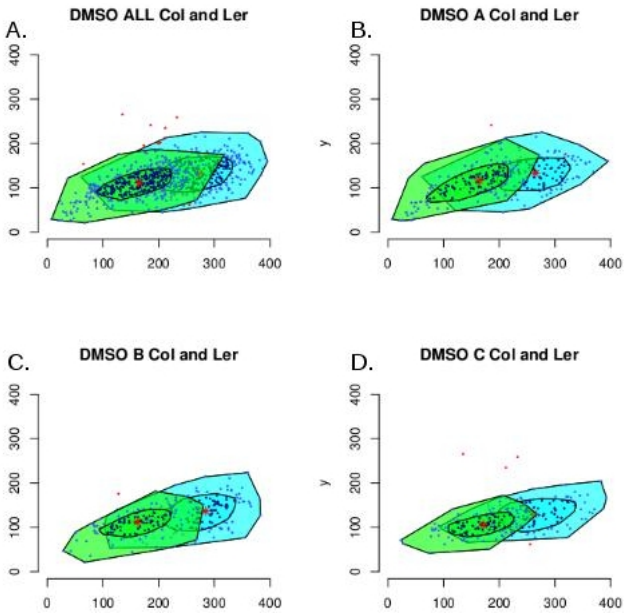

Figure 4.12: The bagplot visualization will be used to for hypocotyl (x-axis) and root (y-axis) values for Col (shown in blue) and Ler (shown in green). A. Average of control treatments with 1% DMSO. B-C. Separate replicates shown have subtle variation but the averages (the center hull of the bagplots) are similar.

4.7 Results

4.7.1 Bioactives explored quantitatively

Several bioactives were identified from combinatorial libraries (Figure 4.11). These bioactives will be explored quantitatively in this chapter.

4.7.2 Hypocotyl and root growth as bivariate bagplots

To further evaluate bioactives the hypocotyl and root length growth in the presence of compounds was quantitatively determined using Col-0 and Ler accessions. Data is displayed in the bi-variate bagplot graphs in which the x-axis represents the hypocotyl values and the y-axis represents the root values. To start out analysis observe the control values for both Col and Ler are reproducible and consider the average over replicates as a reference for expected values (Figure 4.12A-D).

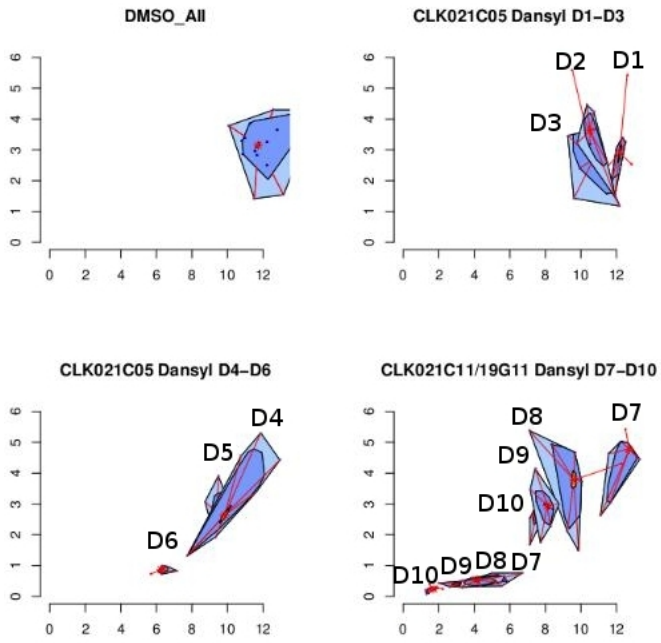

Figure 4.13: CLK021C05 dansyl block A is bioactive at high doses. A. Bagplot showing average of DMSO treatments. B. CLK021C05 dansyl block A at doses D1-D3 do not look much different from control. C. The D4 dose of CLK021C05 dansyl block A shows visible growth inhibition that continues to D6. D. High doses of CLK021C05 at D7-10 are strongly inhibited for hypcotyl and root growth compared to the nearest neighbor CLK019G11 dansyl block A at similar doses. D1-D10 are doses 1,5,15,25,50,75,100,150,175,200 μM.

4.7.3 CLK021C05 Dansyl and derivatives: poor potency and potential for development

The lead bioactive CLK021C05 dansyl block A, was considered not potent enough for further investigation (Figure 4.13). The bioactive effects of CLK021C05 dansyl block A were evident at 100 μM (Figure 4.13C). However an analog CLK019G11 dansyl was inactive found even at 200 μM (Figure 4.13D). Based on lack of response at a high dose CLK019G11 dansyl is a good negative control.

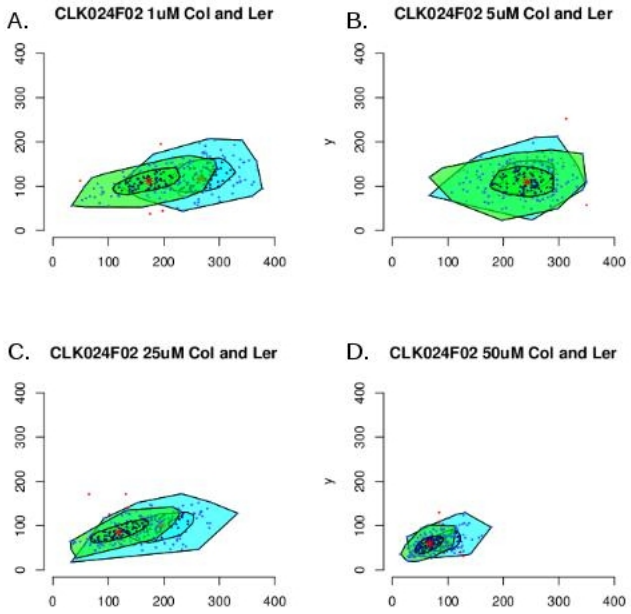

Figure 4.14: CLK024F02 acetylene dose curve with Col (blue) and Ler (green). A-D. Doses are 1, 5, 25, and 50 μM.

4.7.4 CLK024F02 an inhibitor of growth in *Arabidopsis*

Testing of CLK024F02 indicated the mean values for hyocotyl and root were halved around 50 μM for both Col and Ler (Figure 4.14). CLK024F02 caused bleached cotyledons in etiolated seedlings and delayed response upon exposure to light in *Arabidopsis thaliana* (data not shown).

4.7.5 CLK024F02 dansyl block A is an inhibitor of hypocotyl and root growth in *Arabidopsis thaliana*

The 1,2,3-triazole CLK024F02 dansyl block A was bioactive and reduced hypocotyl and root growth reproducibly (Figure 4.15). The phenotype of CLK024F02 dansyl block A appeared distinct from the acetylene CLK024F02 (Figure 4.16A-C) and did not cause cotyledon bleaching as observed for CLK024F02. This compound also resulted in swollen hypocotyls (Figure 4.16C), although this phenotype was not quantified.

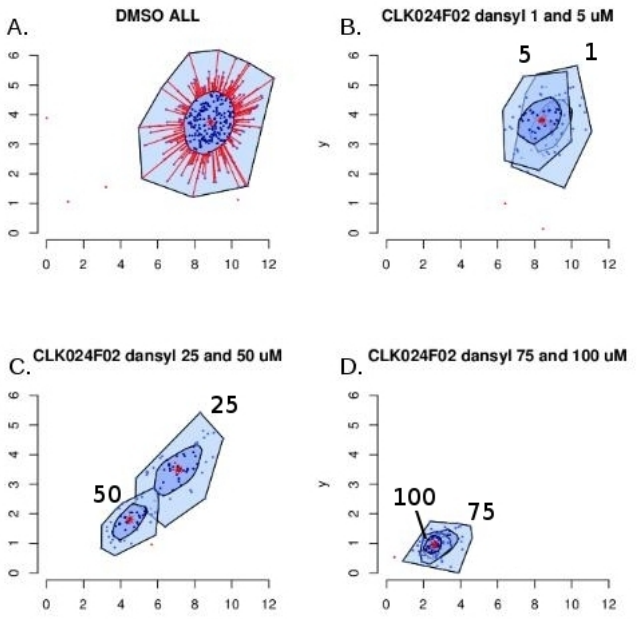

Figure 4.15: Dose curve for CLK024F02 dansyl block A. A. Merge of DMSO has expected values. B. 1 and 5 μM do not inhibit growth. C. 25 and 50 μM cause a reduction in growth. D. Higher doses 75 and 100 μM further inhibit growth.

Figure 4.16: Phenotypes of CLK024F02 dansyl on *Arabidopsis* Columbia-0 ecotype. A-C. Doses of 25, 50, and 75 μM of CLK024F02 dansyl block A. Scale bar is 2 mm.

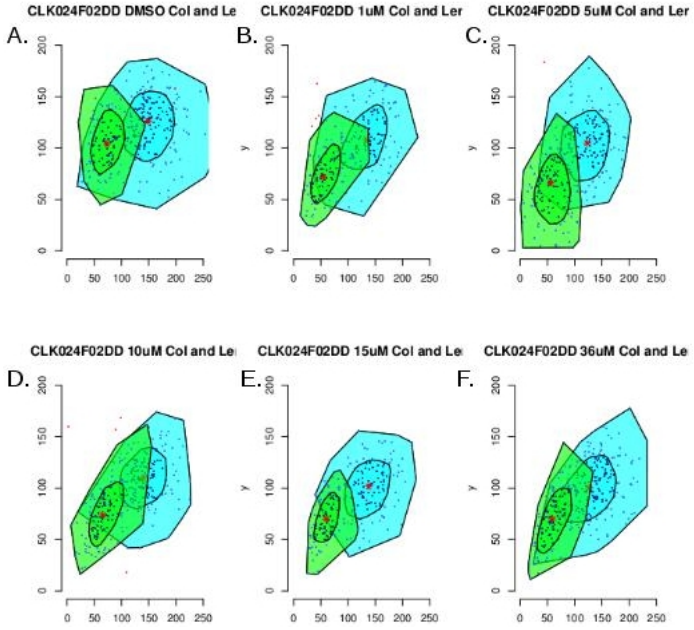

Figure 4.17: CLK024F02 dansyl diazirine quantitative dose curve on Col (blue) and Ler (green). A-F. Treatments with DMSO, 1, 5, 10, 15 and 36 μM CLK024F02 dansyl diazirine.

4.7.6 CLK024F02 dansyl diazirine is an inhibitor of hypocotyl and root growth in *Arabiodpsis thaliana*

The 1,2,3-triazole CLK024F02 dansyl diazirine is bioactive and produces a phenotype similar to the dansyl block A version (Figure 4.17). A search for inactive analogs for 1,2,3-triazole of CLK024F02 in the clickable library was unsuccessful despite a search using a number of candidates were generated from the clickable library. The closest candidate, in structure, CLK041B06 dansyl appeared to cause a different phenotype, and was abandoned (Figure 4.18).

Preliminary analysis of CLK024F02 analogs in the clickable library suggested substitutions on the heterocycle and on the propargyl group lead to a decrease in bioactivity of the 1,2,3-triazole block A conjugate. The phenotypic relationship between the analogs could not be established and the MCS of the similar compounds was very small,

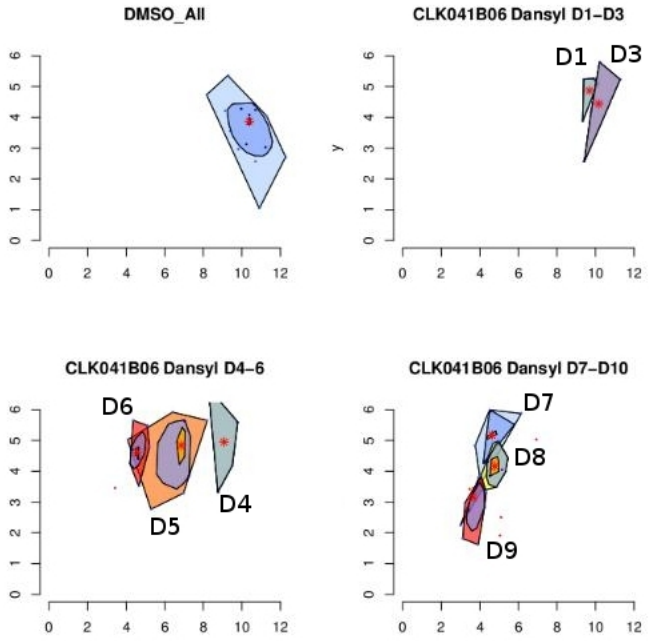

Figure 4.18: CLK041B06 dansyl block A is a bioactive nearest neighbor to CLK024F02 dansyl block A. A-D. Treatments with doses D1-D9, up to 150 μM cause growth inhibition at D4-D6 and at higher doses. Doses D1-D9 are 1,15,25,50,75, 100,125, 150, 200 μM.

thereby misrepresenting similarity in the small clickable set. Synthesis and testing of the de-chlorinated version of CLK024F02 (Figure 4.11D), 2-(prop-2-yn-1-ylsulfanyl)-1,3-benzoxazole from 2,3-dihydro-1,3-benzoxazole-2-thione, may be a useful step step in the search for an inactive analog.

4.7.7 CLK042A09 is a weak inhibitor of hypocotyl and root growth

The sulfonamide CLK042A09 or N-(2-[(methylsulfamoyl)methyl]phenylmethyl)-4-[(prop-2-yn-1-yl)sulfamoyl]benzamide was the most interesting dansyl block A lead. The terminal acetylene CLK042A09 was not a potent inhibitor of hypocotyl and root growth even at 50 μM (Figure 3.32) and (Figure 4.19)A-D.

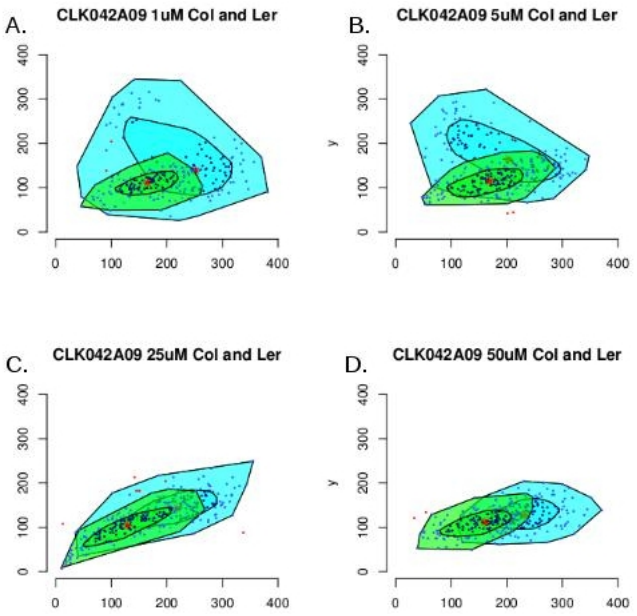

Figure 4.19: CLK042A09 acetylene quantitative dose curve with Col and Ler. A-D. 1, 5, 25 and 50 μM CLK042A09.

4.7.8 CLK042A09 dansyl block A is an inhibitor of hypocotyl and root growth

CLK042A09 dansyl block A was potent to inhibit root length by half at 15 μM and 5 μM (Figure 4.20) as seen in Figure 4.20A-C. The swollen hypocotyl and short root phenotype caused by CLK042A09 dansyl block A was reminiscent of ethylene and cellulose synthase mutants.

4.7.9 CLK042A09 analogs CLK039G03 and CLK039G03 dansyl block A are inactive

The analog CLK039G03 dansyl block A was not active at 100μM (Figure 4.21D) despite close similarity to CLK042A09.

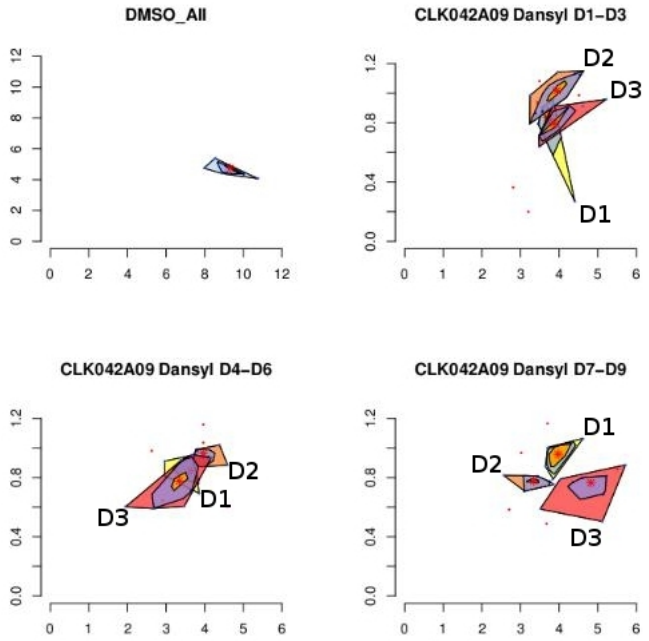

Figure 4.20: CLK042A09 dansyl block A quantitative dose curve. A. DMSO control. B. Doses D1 and D2 seem to elongate the root, whereas dose D3 causes some root inhibition. C. Doses D4-6 have more pronounced effects on growth. D. Treatment at higher doses may have confounding results due to precipitation. Doses D1-D5 are 5, 25,50, 75, 100 μM.

Figure 4.21: CLK042A09 dansyl block A is bioactive compared to the nearest neighbor CLK039G03 dansyl block A. A-C. 25, 50, and 75 μM of CLK042A09 dansyl block A. D. Seedlings treated with 100 μM CLK039G03 dansyl block A are not inhibited. Scale bar is 2 mm.

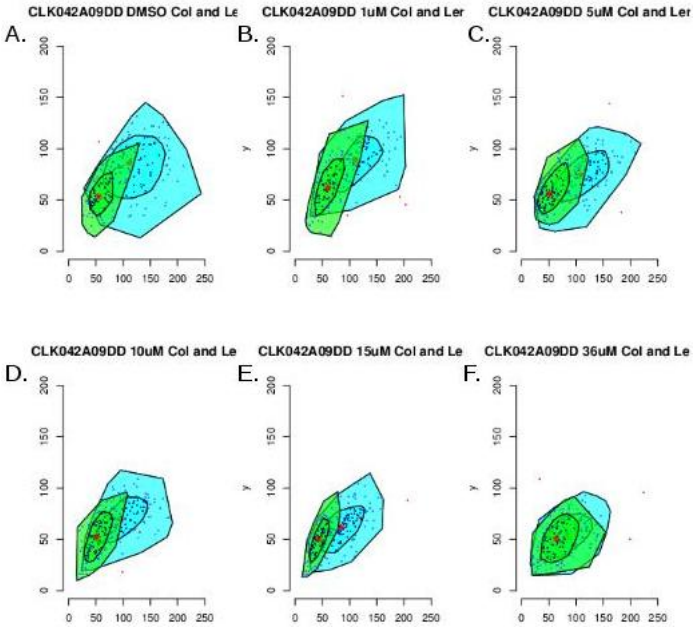

Figure 4.22: CLK042A09 dansyl diazirine quantitative dose curve of Col (blue) and Ler (green). A-F. Treatments with DMSO, 1, 5, 15, and 36 μM of CLK042A09 dansyl diazirine

4.7.10 CLK042A09 dansyl diazirine is an inhibitor of hypocotyl and root growth in *Arabidopsis thaliana*

Bioactivity of CLK042A09 dansyl diazirine was identical to the dansyl block A version (Figure 4.20) compared to (Figure 4.22).

4.8 Summary and Discussion

Several promising candidates for fluorescent probes (Figure 4.11) were evaluated for potency to disrupt early seedling growth. Bi-variate bagplot graphs were used to characterize the domains of hypocotyl and root values under varying doses. CLK021C05 Dansyl did not prove to be potent enough to develop further despite having an inactive analog CLK019G11 Dansyl (Figure 4.13).

The next probe CLK024F02 Dansyl was evaluated another day from the parent

acetylene (shown with a different pixel scale) (Figure 4.15). Nonetheless, it could be determined that CLK024F02 Dansyl (Figure 4.15) was moderately stronger as a fluorescent probe, and this trend was similar for the diazirine functionalized version CLK024F02 Dansyl Diazirine (Figure 4.17). An inactive analog was not found.

The last probe CLK042A09 Dansyl, although the biggest, had the most dramatic change in bioactivity as a fluorescent probe (Figure 4.20). The parent acetylene was mildly bioactive (Figure 4.19), but we can observe large changes in bioactivity in the final fluorescent probe.

4.9 Conclusion

The best candidate for further investigation was CLK042A09 dansyl. Therefore, the diazirine version may be used for covalent efforts to identify the target, and was selected as the focus of a later chapter (Chapter 6: Target Identification through Covalent Capture). A complement to the dose curve analysis would be a resistant individual or line. This is the goal of our analysis in the next chapter (Chapter 5: Target Identification Efforts through EMS screening).

Chapter 5

Target Identification efforts with a population Ler *Arabidopsis* M0

5.1 Abstract

Previous chapters elucidate the discovery of several novel terminal acetylene and 1,2,3-triazole probes with an unknown mechanism of action. A resistance screen was conducted with two select bioactives to identify resistant candidates from a population of *Arabidopsis* M0. A quantitative approach was chosen to interrogate the resistant candidates, but strong resistant candidates were not identified for the bioactive and fluorescent 1,2,3-triazoles CLK024F02 dansyl block A and CLK042A09 dansyl block A. A discussion of future research to identify the target of these compounds is presented.

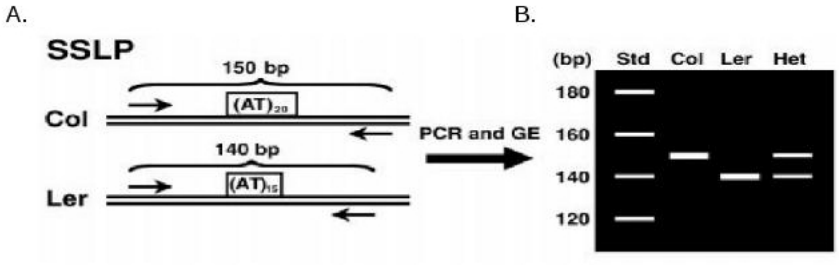

Figure 5.1: Figure adapted from Lukowitz et al. [72]. A. Simple sequence length polymorphisms (SSLPs) exist between *Arabidopsis* ecotypes Col and Ler. B. Polymerase chain reaction (PCR) amplification of the polymorphic regions visualizes the nucleotide length difference for this allele in the homozygous Col and Ler backgrounds as well as the heterozygote. Std is a ladder with a range of 120-180 bp, Het indicates heterozygote.

5.2 Introduction

5.2.1 EMS as a tool to identify small molecule targets in *Arabidopsis thaliana*

Ethanemethylsulfonate (EMS) is a mutagen that has been utilized to introduce heritable mutations in *Arabidopsis* and isolate resistance alleles to novel probes [84, 26]. EMS causes stable mutations by alkylating guanine of G-C pairs [62]. The mutated base pair is replaced during replication or mismatch repair as an A-T pair, and the nucleotide change can alter the triplet codon to disrupt protein synthesis or to produce a non-functional protein. During the process of meiosis the mutant alleles are introduced to the germplasm from mutant sectors of the apical meristem that develop into the sillique structure [62]. Screening the germplasm of EMS treated populations is an accepted method to find candidates with a genetic lesion involved in the mechanism of action of novel probes [66]. We will refer to this process in short as EMS screening.

5.2.2 RFLP mapping of candidate mutants with a 30 cM primer walk

Natural genetic variation between Columbia and Landsberg is a classical approach to determine the linkage of ecotype specific markers [72]. The rough chromosomal location of a genetic lesion is found through mapping techniques and can refine down to a couple of open reading frames [85]. Mapping of an F2 population is accomplished through DNA extraction and PCR amplification of Simple Sequence Polymorphisms (SSLPs) (Figure 5.1A), where amplified DNA is subjected to 12% agarose electrophoresis to determine the presence or absence of ecotype specific amplicons (Figure 5.1B) [72].

5.3 Materials and Methods

5.3.1 Resistance screens with EMS library and bioactive probe

EMS-mutagenized seed for the *Arabidopsis* Ler ecotype was soaked in mutagen and grown to maturity [84](Cutler lab). Seed was collected in segregated seed pools, to isolate siblings [86]. Chlorotic sectors and lethality was observed in less than 1% of seedlings screened, a typical observation of a successful mutagenesis procedure [86]. Therefore, we used the seed pools for the identification of a resistant individual.

Assay plates were made from standard petri dishes filled with ms agar growth media, as previous, with the addition of compound to the final concentration of 50 μM for CLK024F02 dansyl block A plates and 25μM for CLK042A09 dansyl block A plates. Assay plates were chilled and grown in the dark as previous. Resistant candidates were identified with either long hypocotyls or roots or both and rescued on the final day using tweezers. Candidates that grew to maturity and set seed were re-tested under conditions identical to dose curves (Chapter 4: Phenotypes of Bioactives), and displayed as bagplot

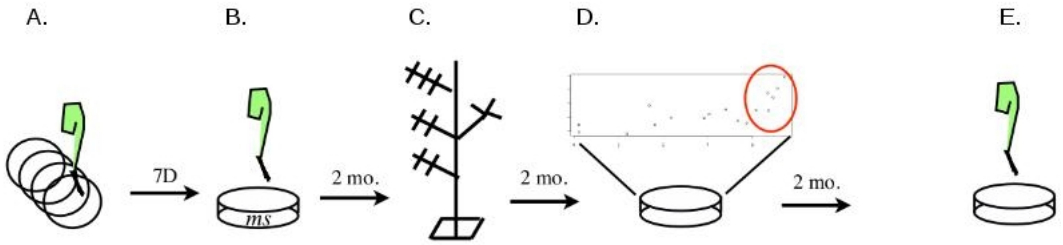

Figure 5.2: Schematic showing the steps involved to generate resistant mutant candidates from an EMS library of *Arabidospsis* seed. A. A resistant individual is identified from assay plates containing bioactive probe. B. The resistant candidate is rescued to normal growth media for a couple days and transferred to soil to produce more progeny for testing. C. *Arabidopsis* is allowed to grow to maturity and seed is collected when ready. D. Progeny from resistant individuals is tested in the original assay plate format with bioactive probe. E. The strongest candidates are selected, selfed or crossed to further molecular mapping strategies and assign dominance.

graphs. Discussion regarding bagplot graphs and significance values has been detailed in Chapter 3 at the opening of the Results section.

5.4 Results

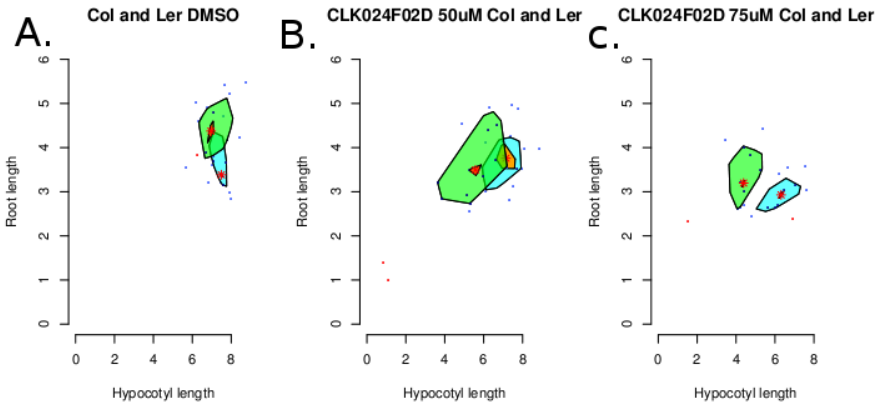

Figure 5.3: Baseline test of CLK024F02 dansyl on Col and Ler shown as blue and green bagplots. A. Both ecotypes shown on 1% DMSO treatment. B. Treatment with 50 μM probe causes exagerrated changes in Ler and a drop in root values. C. Increased dose of 75 μM leads to a reduction in mean hypocotyl and root values (center of red star) for both Col (shown in blue) and Ler (shown in green).

5.4.1 Baseline tests of CLK024F02 dansyl on Col and Ler

Initial tests with CLK024F02 dansyl on both ecotypes determined the expected range variation when dealing with two ecotypes as seen with the control values (Figure 5.3A.). Nonetheless, the probe's effects are clear as the dose is increased to 50 and 75 μM causing an decrease in hypocotyl and root values from (8, 3.5) to (6,3) for Col and (7,4.5) to (4,3) for Ler.

5.4.2 The search for resistant candidates for CLK024F02 dansyl block A

The resistant candidate CLK024F02D-SC-43-1 had a typical growth on DMSO with inhibitory effects observed at 50 and 75 μM (Figure 5.4A and D). Notice the characteristics of the bag and loop hull are retained at the lower dose of 50 μM (Figure 5.4D). CLK024F02D-SC-43-2 has a wide spread of high values at 75 μM, but the mean has dropped from (8,5) to (6,4) (Figure 5.4B). CLK024F02D-SC-43-3, the last line in this

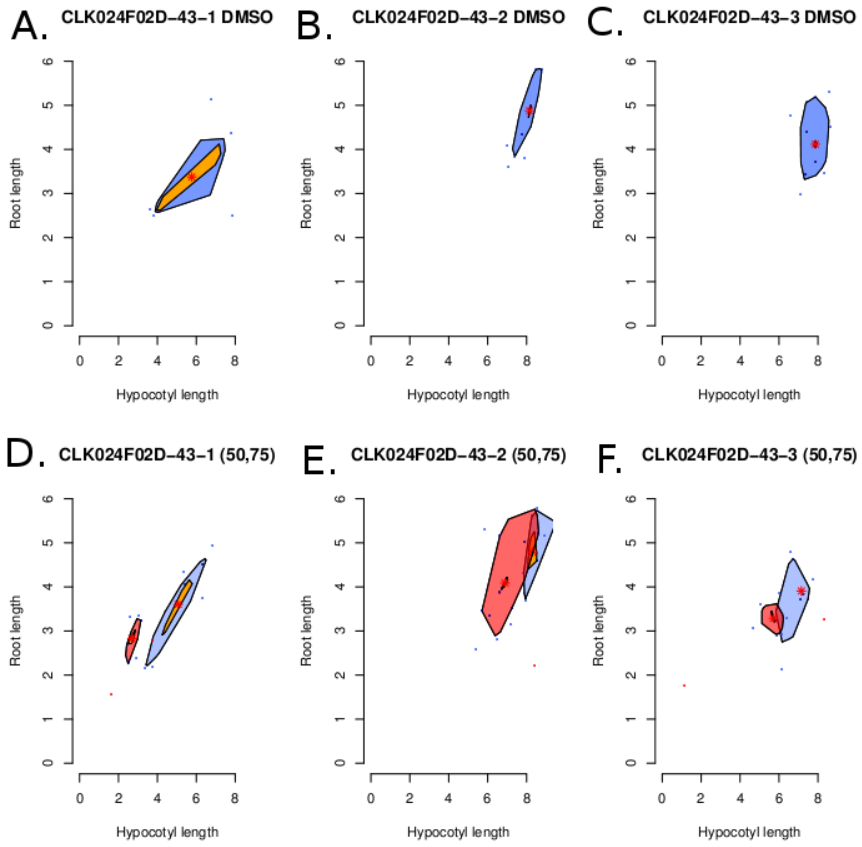

Figure 5.4: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 50 μM (light blue) and 75 μM (red) for the lines shown.

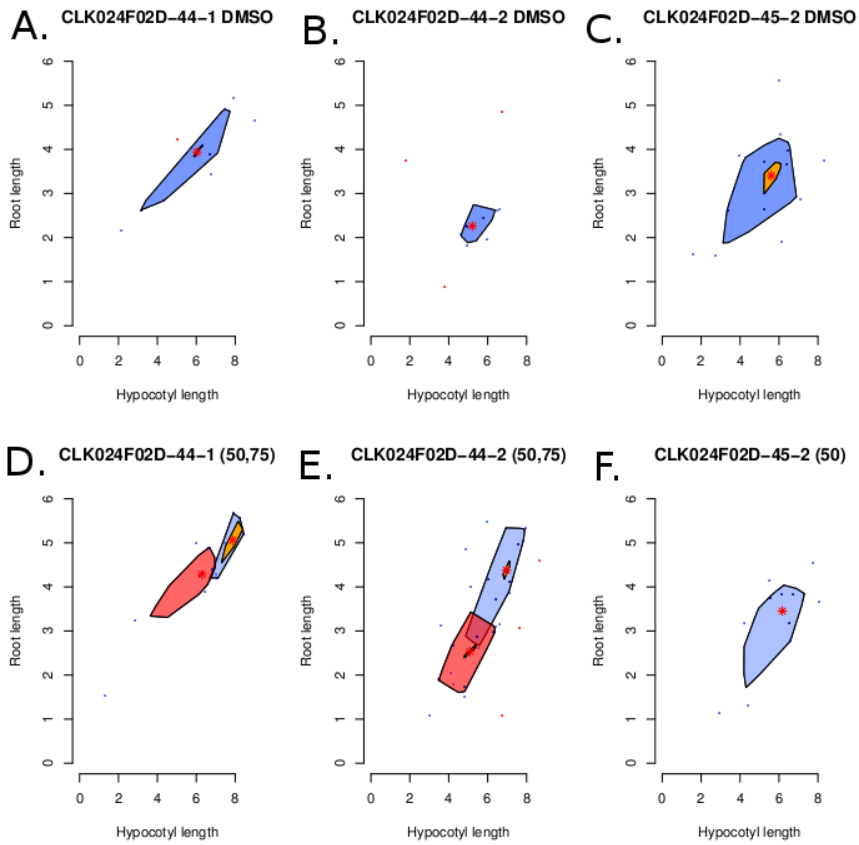

Figure 5.5: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 50 μM (light blue) and 75 μM (red) for the lines shown.

panel has a marked reduction in mean values at 50 and 75 μM (Figure 5.4F).

The mean values of CLK024F02D-SC-44-1 appear unchanged at 50 and 75 μM with some outliers showing values of (2,2) (Figure 5.5A and D). Similarly the mean of CLK024F02D-SC-44-2 is not changed at 75 μM, and out performs control values at 50 μM (Figure 5.5B and E). The last line in this panel, CLK024F02D-SC-45-2 was scarce and only data for 50 μM was derived to demonstrate top performers are slightly inhibited at this dose (Figure 5.5C and F). The performance of CLK024F02D-SC-45-2 can be compared to CLK024F02D-SC-46-1 and CLK024F02D-SC-46-2 at a similar dose. Both CLK024F02D-SC-46-1 and CLK024F02D-SC-46-2 have values larger than

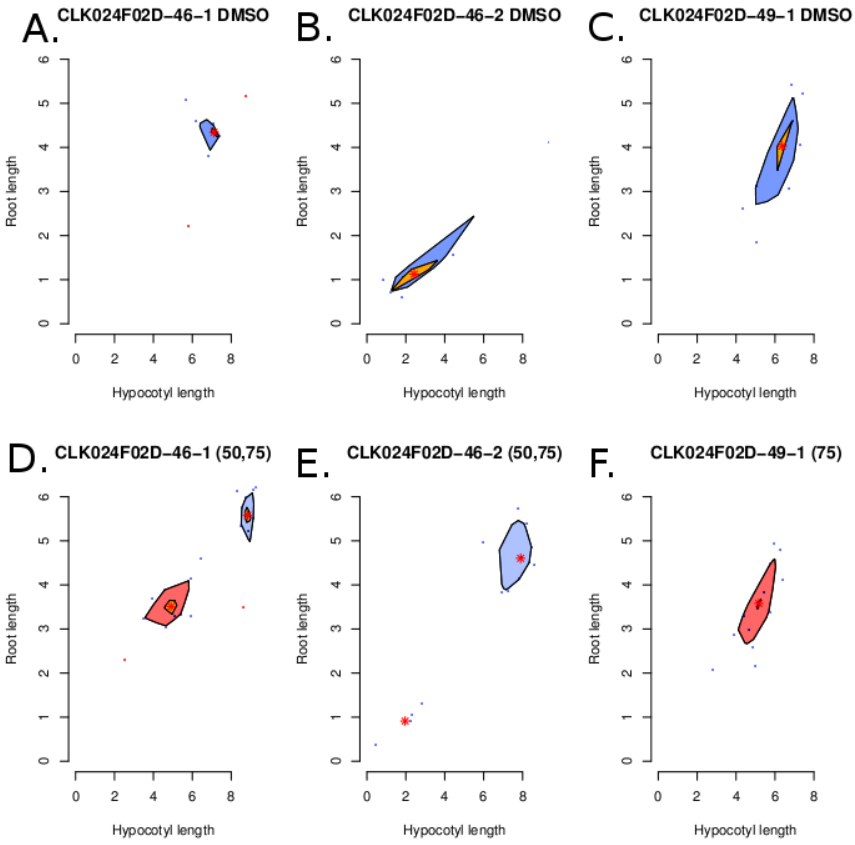

Figure 5.6: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 50 μM (light blue) and 75 μM (red) for the lines shown.

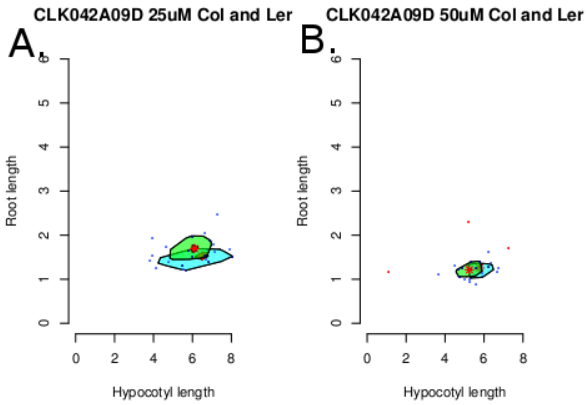

Figure 5.7: Baseline test of CLK042A09 dansyl on Col and Ler shown as blue and green bagplots. A. Both ecotypes shown on 1% DMSO treatment. B. Treatment with 25 μM probe causes exaggerated changes in Ler and a drop in root values. C. Increased dose of 50 μM leads to a reduction in mean hypocotyl and root values (center of red star) for both Col (blue) and Ler (green).

for respective DMSO control (Figure 5.6A-B, D-E). CLK024F02D-SC-49-1 on the other hand has a bivariate mean at (6,4) that drops slightly to (5,3.5) (Figure 5.6C and F).

5.4.3 The search for resistant candidates for CLK042A09 dansyl block A

CLK042A09 dansyl block A is potent to restrict the hypocotyl and root values significantly at 25 μM with subtle changes up to 50 μM (Figure 5.7A and B.). CLK042A09D-SC-53-2 was a promising candidate with a subtle change in bivariate mean from the control values (6,3) to (6,2.5) when dosed at 25 and 50 μM (Figure 5.8A and D). The line CLK042A09D-SC-53-2 had a number of candidates with long roots and average sized hypocotyls (Figure 5.8D). The next candidate CLK042A09D-SC-53-3 performed well at 25 and 50 μM with a majority of the sample having mean values above control for the line (Figure 5.8B and E). The last in this panel CLK042A09D-SC-53-4 has a reduction in root growth at 50 μM with a significant proportion with short roots

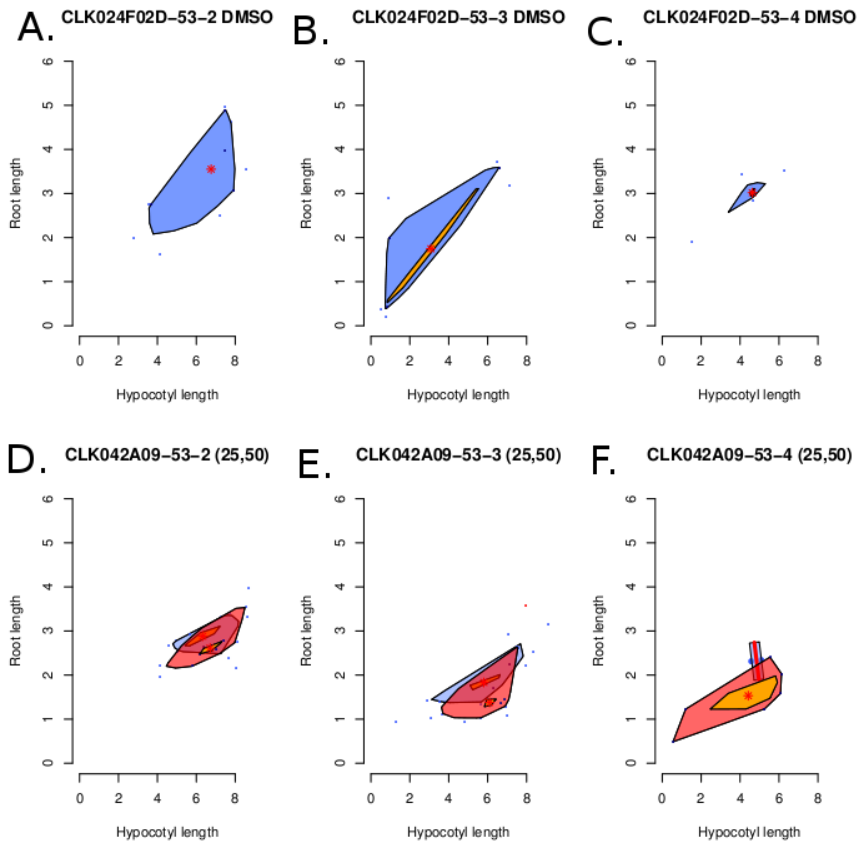

Figure 5.8: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 25 μM (blue) and 50μM (red) for the lines shown.

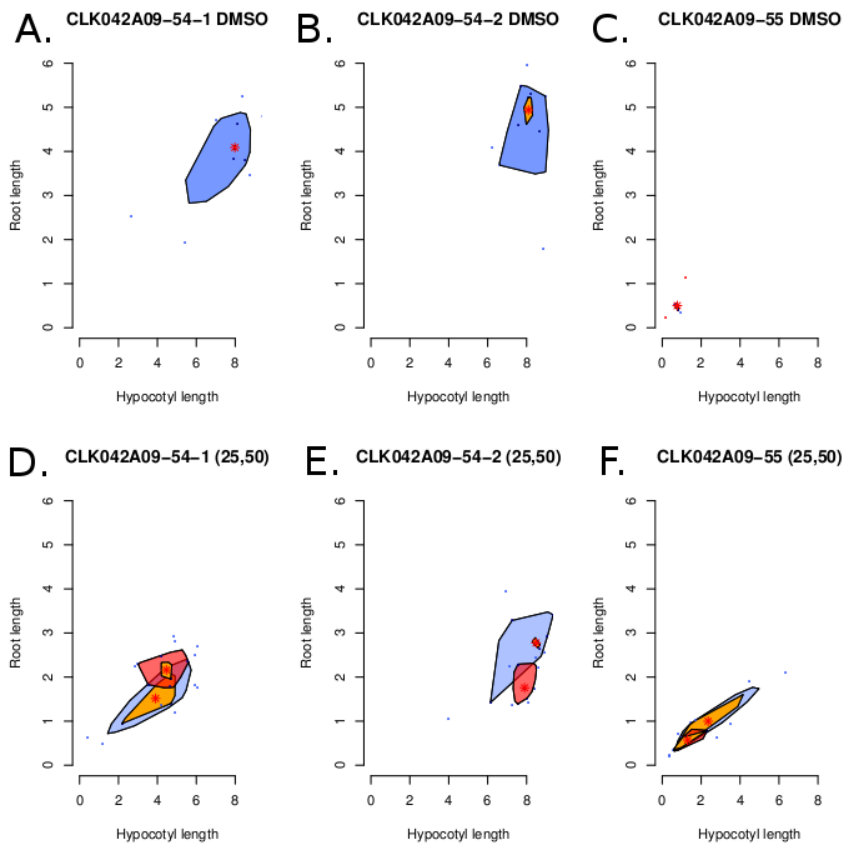

Figure 5.9: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 25 μM (blue) and 50 μM (red) for the lines shown.

and hypocotyls (Figure 5.8C and F).

CLK042A09D-SC-54-1 had a clear reduction in hypocotyl and root values at 25 and 50 μM despite having a good performance on DMSO (Figure 5.9A and D). The line CLK042A09D-SC-54-2 had unchanged mean values at 25 and 50 μM for the hypocotyl and a reduction in root size (Figure 5.9B and E). The last line CLK042A09D-SC-55 performed poorly on DMSO and on 50 μM (Figure 5.9C and F). The values for 25 μM were well below the means for Col and Ler (Figure 5.9F) compared to (Figure 5.8B). The results (Figure 5.4-Figure 5.9) were derived from the set of Ler-EMS mutant candidates

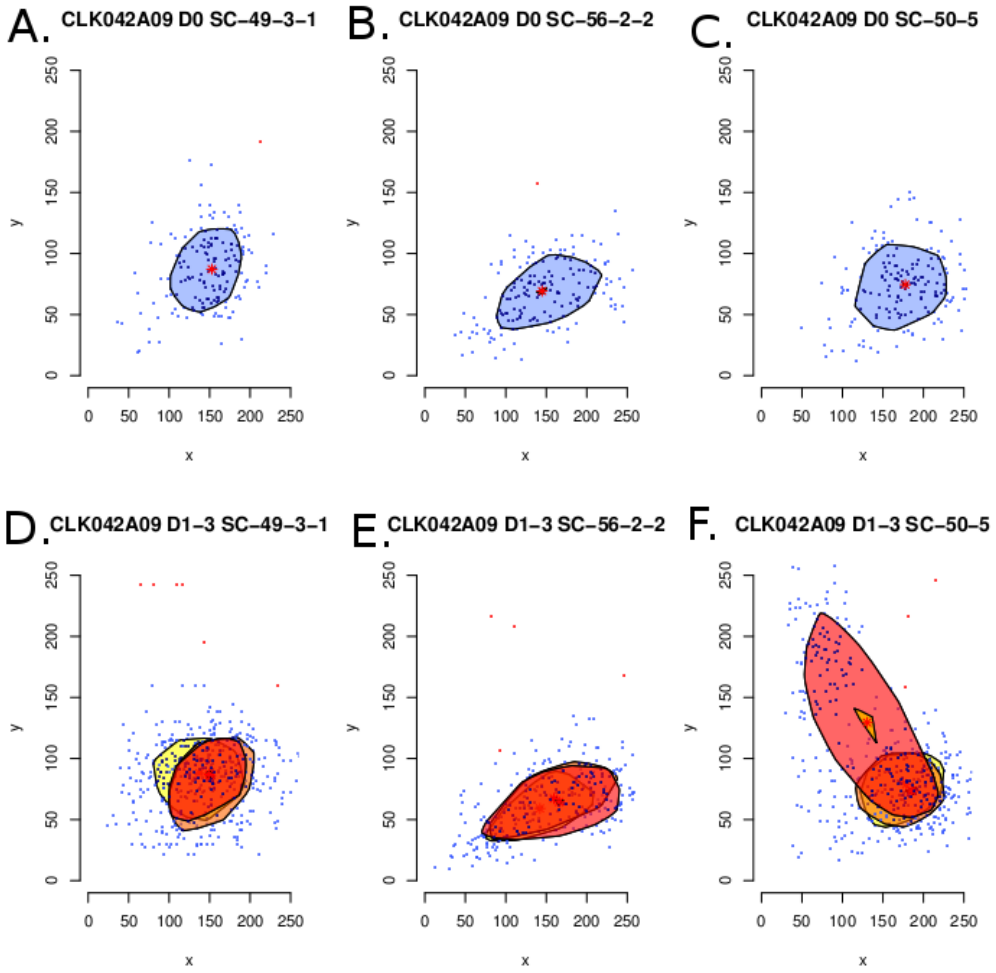

Figure 5.10: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 12 μM (yellow), 25 μM (brown) and 50 μM (red) for the lines shown.

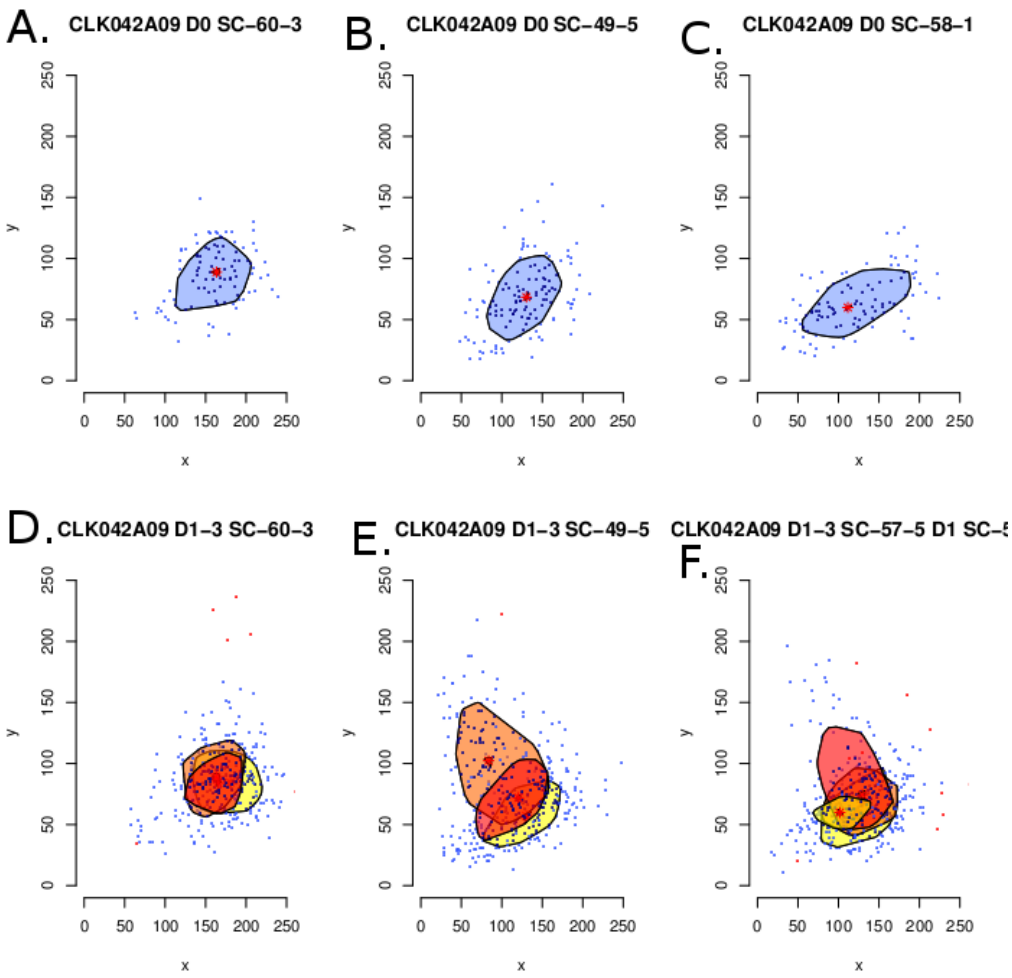

Figure 5.11: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 12 μM (yellow), 25 μM (brown) and 50 μM (red) for the lines shown.

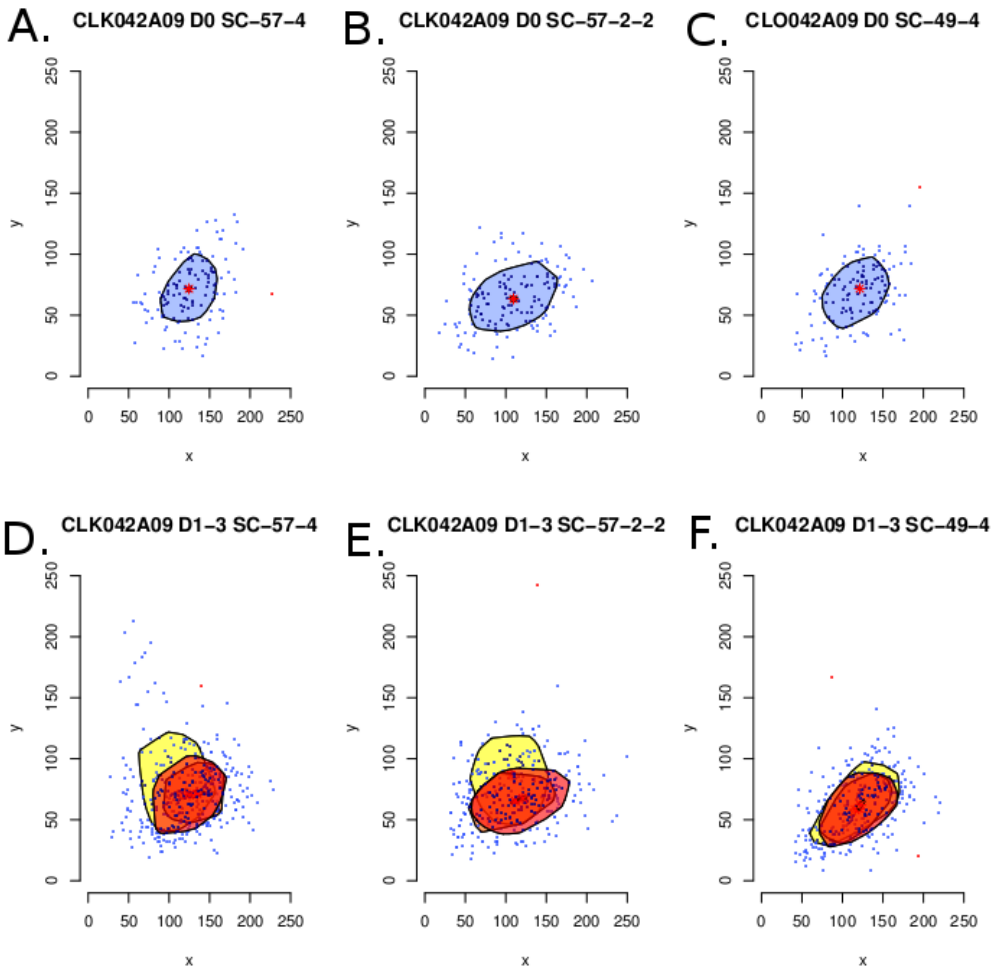

Figure 5.12: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 12 μM (yellow), 25 μM (brown) and 50 μM (red) for the lines shown.

rescued during the first resistance screen (Figure 5.2C).

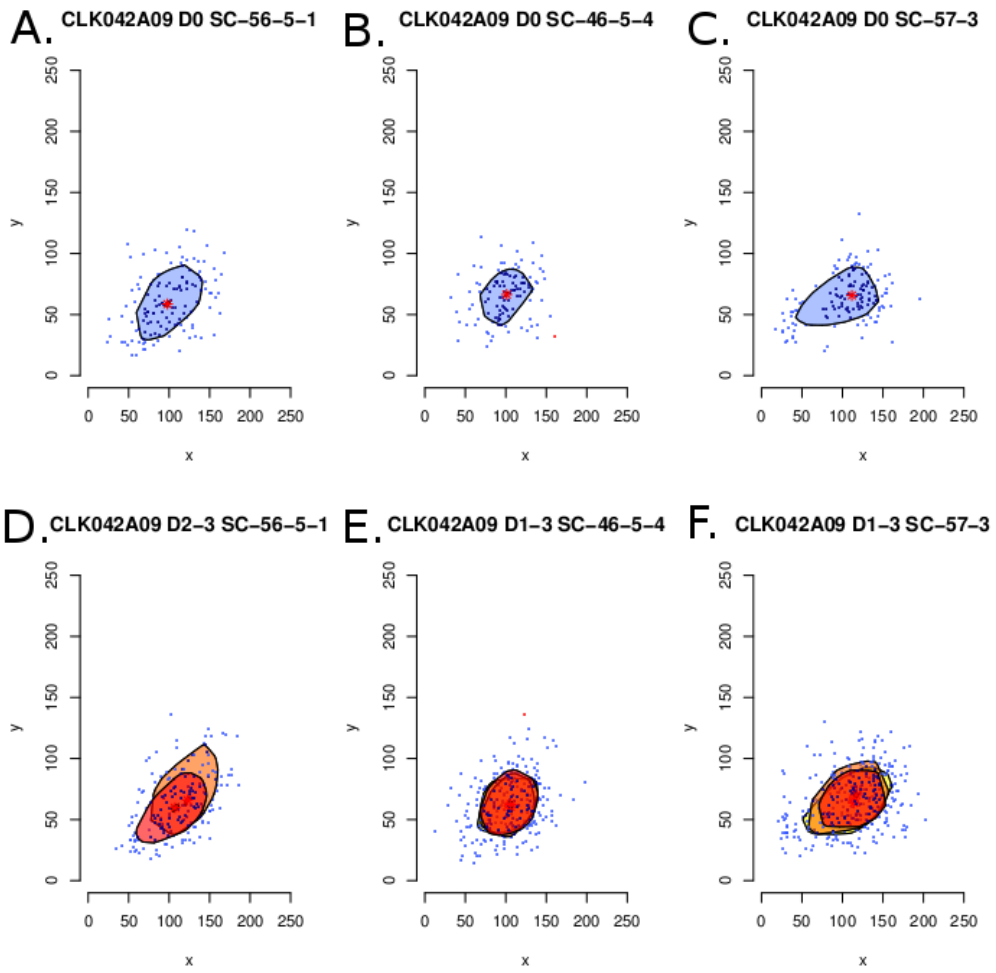

Figure 5.13: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 12 μM (yellow), 25 μM (brown) and 50 μM (red) for the lines shown.

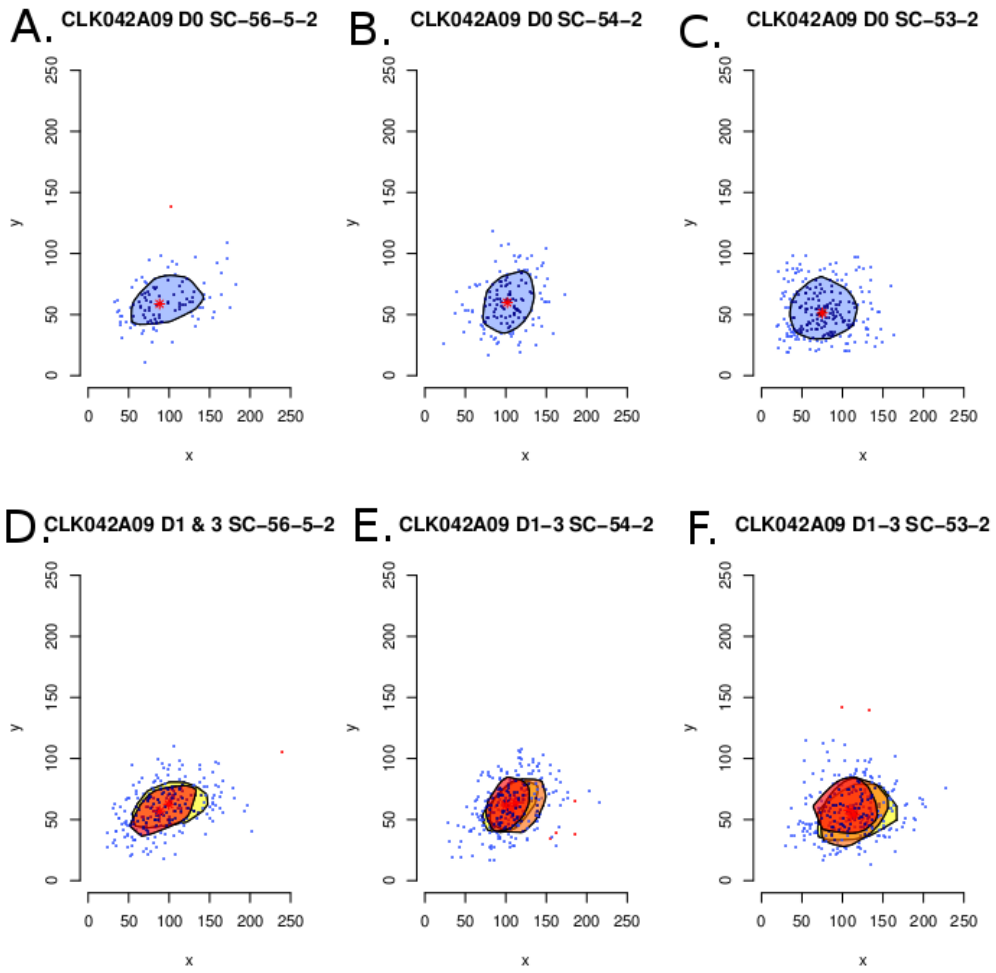

Figure 5.14: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 12 μM (yellow), 25 μM (brown) and 50 μM (red) for the lines shown.

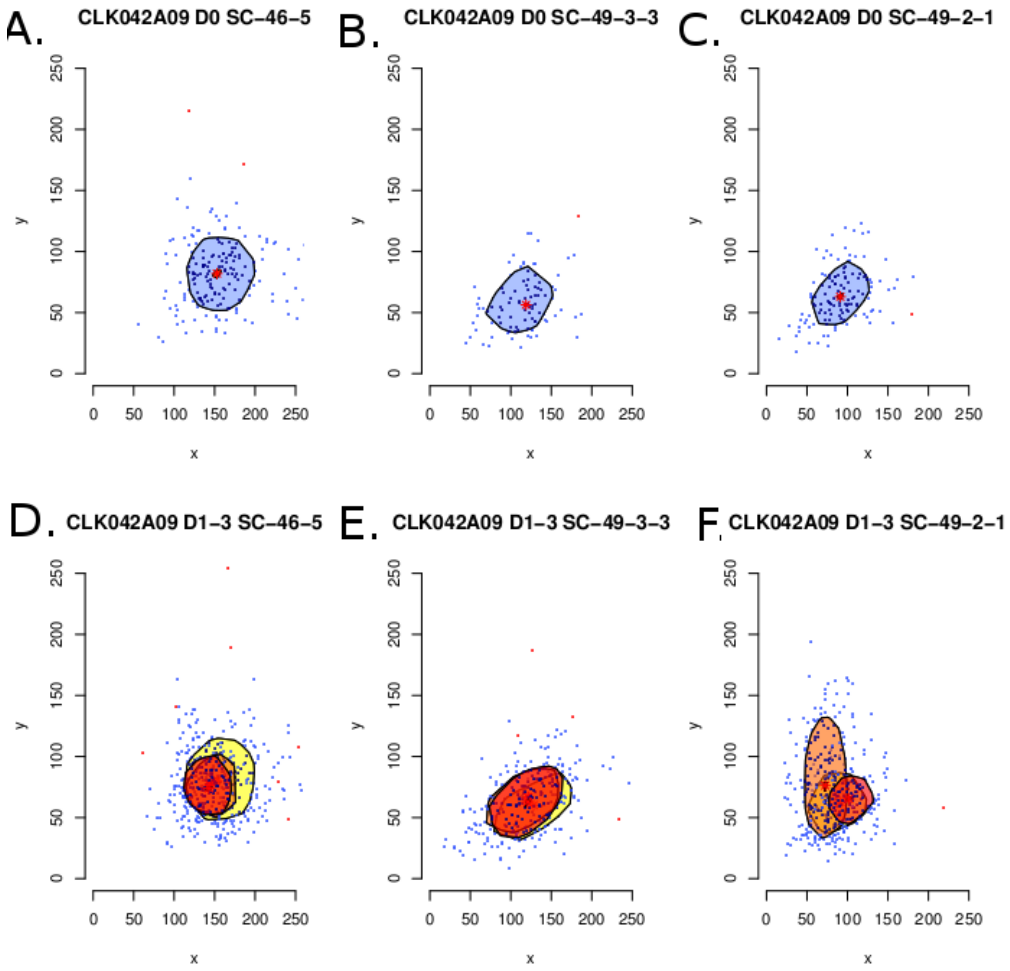

Figure 5.15: Dose curve analysis of putative mutants identified by EMS screening. A-C. DMSO control. D-F. Doses are indicated 12 μM (yellow), 25 μM (brown) and 50 μM (red) for the lines shown.

5.4.4 Secondary search for resistant candidates to CLK024F02 dansyl block A and CLK042A09 dansyl block A

A second resistance screen was conducted to identify more resistant individuals to CLK024F02 dansyl and CLK042A09 dansyl (Figure 5.2B). A number of candidates were found, rescued and grown to maturity to acquire seed for analysis (Figure 5.2B and C). Due to limited availability of the probes CLK024F02 dansyl and CLK042A09 dansyl the progeny of the rescued Ler-EMS seedlings were first interrogated using the respective terminal acetylene fragments for any signs of cross-resistance to the acetylene fragment (Figure 5.10)-(Figure 5.15).

The line CLK042A09D-SC-49-3-1 has mean values centered around (150, 100) for all doses of CLK042A09 examined (Figure 5.10A and D). Line CLK042A09D-SC-56-2-2 has very tall hypocotyls for all doses examined with a reduction of root values showing at 30 μM (Figure 5.10B and E). The performance of CLK042A09D-SC-50-5 on 30 μM CLK042A09 shows a large variance not observed for the DMSO control (Figure 5.10C and F). The outliers for this sample at 30 μM almost segregate the bagplot into two regions (Figure 5.10F). In one extreme, the bagplot for CLK042A09D-SC-50-5 has a number of individuals with shorter roots and long hypocotyls (Figure 5.10F, top half of baghull) and the other has long roots with short hypocotyls (Figure 5.10F, bottom half of baghull).

All doses of CLK042A09 caused a reduction of the hypocotyl and root bagplot domains of CLK042A09D-SC-60-3 with no reduction in the bivariate mean (Figure 5.11A and D). The line CLK042A09D-SC-49-5 has long roots on 15 μM wihich shorten when the dose is doubled (Figure 5.11E). The shape and span of the baplot for CLK042A09-SC-57-5 at 30 μM (Figure 5.11F) is similar to CLK042A09D-SC-49-5 (Figure 5.11E) at a

lower dose. Several outliers can be seen at the forefront of the bagplot for CLK042A09D-SC-57-5 at a high dose (Figure 5.11F).

In the next panel the line CLK042A09D-SC-57-4 appears to have long roots at 5 μM of CLK042A09 which shortens to a similar domain for 15 and 30 μM (Figure 5.12A and D). Performance of the line CLK042A09-SC-57-2-2 at 30 μM (Figure 5.12B and E) looks promising, but is suspicious since lengths at 15 μM were so short (Figure 5.12E). The next line, CLK042A09D-SC-49-4 appears similar to previous with a slightly better performance at 15 μM (Figure 5.12F). The subtle differences between these lines can be elucidated with future analysis. In comparison to other panels (Figure 5.13)-(Figure 5.14A-F) the bagplots may indicate resistance trends worth investigating.

5.4.5 Clear phenotypes that are difficult to interpret

In the first scenario CLK042A09D-SC-46-5 has a gradual compression of both hypocotyl and root domains which are nested in each other (Figure 5.15D). As we increase the dose *Arabidopsis* has a gradual reduction in overall length. In the second example, the mean is slightly higher than control for all doses and the low dose has outliers with taller hypocotyls (Figure 5.15B and E). An increase of the dose from 5 to 15 μM causes the expected reduction in values which do not change further when doubling the dose (Figure 5.15E). The third and last scenario is a peculiar one where the middle dose outperforms the control values and the high dose has compressed the baghull of control values (Figure 5.15F). The mechanics of the resistance performances in these rescued lines of *Arabidopsis* may be resolved with similar resistance analysis using closely spaced doses.

5.5 Summary and Discussion

5.5.1 Sequencing options to identify the causative resistance allele to bioactives

A resistant candidate for the 1,2,3-triazole probes was not identified during screening. A number of options are available to perform candidate gene sequencing once a resistant individual is identified. Illumina array based sequencing [87] was the method chosen to identify the SNP for CLK003D03. The traditional Sanger di-deoxy sequencing technique [88] and other high throughput sequencing platforms such as 454 are available to identify genetic lesions caused by EMS-mutagenesis. The approach chosen should consider the number of candidates to test, budget, and extend of secondary genetic lesions.

5.5.2 Considerations before Illumina sequencing

SHOREMAP [87] has been used to find genetic lesions in *Arabidopsis*. Austin et al. [89] demonstrated that we can identify causative SNPs by Illumina sequencing of allelic pooled F2 populations. Rough mapping was used to determine linkage of the resistance allele to a rough chromosomal location. In addition, tests were made to determine a sufficient level of sufficient genetic coverage when sequencing pools of samples using Illumina to accomplish successful SNP identification [89]. In addition, the F2 line may be selfed several generations prior to mapping to remove secondary lesions. This consideration may become the norm since Austin et al. and others have reported genetic heterogeneity at the SNP level in homozygous inbred populations and even between lab strains of the Columbia-0 ecotype.

5.5.3 Quantitative phenotype metrics offer a solution to evaluate individuals and a population *en masse*

Strict quantitative phenotype evaluation of individuals may clarify the decision making process when planning costly whole genome sequencing of resistant mutant, mutant candidates or pools. Especially when the bioactive causes a subtle phenotype or the probe is very rare. Furthermore, sound quantitative data may be used for statistical analysis of mutant pools to discover causative SNPs.

5.6 Conclusion

My goal was to identify resistant *Arabidopsis* candidates that could be used to perform map based target identification strategies for the rare probes CLK024F02 dansyl block A and CLK042A09 dansyl block A. Several resistance screens were performed and the results were visualized using a bivariate bagplot. A number of candidates display trends of resistance for both or either of hypocotyl and root values, but these were not explored further.

The phenotypes of CLK024F02 dansyl and CLK042A09 dansyl were previously unknown. Since these probes can be modified to diazirine containing photo-affinity ligands (Chapter 3 and Chapter 4) they satisfied our primary objective in a probe, despite the modest phenotype. Therefore, we quantitatively demonstrated that individuals can be found with dose specific trends of resistance. These trends may not be apparent or discernible by eye. Therefore, by systematizing the collection and display of these values in bagplots we can describe the range and behavior of novel candidates from EMS pools. A threshold for resistance will be a necessary parameter for further resistance screening for either of hypocotyl or root values, and tests should be done to confirm if

dose specific resistance can be linked to a resistance allele. Seed for select individuals will be submitted to the Arabidopsis Biological Resource Center (ABRC)[90] to support external validation.

Chapter 6

Target identification efforts via covalent capture

6.1 Abstract

The diazirine ring was explored for use as a photo-affinity module in our fluorescent and bioactive probes. 1,2,3-Triazole conjugates were prepared from dansyl diazirine block B in the TBTA catalyzed click reaction (Chapter 2: Organic Synthesis). The bioactive candidate CLK042A09 dansyl diazirine was the focus of our investigation. A preliminary experiment was conducted to label *Arabidopsis thaliana* protein extracted from etiolated seedlings. Non-specific labeling of a small molecular weight protein was observed for CLK042A09 dansyl diazirine and the inactive analog CLK039G03 dansyl diazirine. Experiments to label generic protein such as Bovine Serum Albumin (BSA) was conducted to learn more regarding the stoichiometry of ligand-protein binding. These experiments were inconclusive. To simplify our experiment we chose to label single amino acids with the probe CLK042A09 dansyl diazirine. In sum, our results suggested activation of the diazirine probe by ultraviolet photolysis at 365 nm was insufficient. En-

IC50	High (nM)	Low (μM)	Low (μM)		
Target abundance	Low	High	Low		
	Likelihood	of	success	Range	Probe
Affinity capture	Good	Poor	Poor		Biotin tag
Coomassie stain	Maybe	Poor	Unlikely	30-100 ng	None
Fluorescence	Good	Poor	Unlikely	?	Fluoresce
Silver stain	Good	Poor	Unlikely	5-10 ng	None
Chemiluminescence	Good	Good	Maybe	0.25-1.0 ng	Biotin tag

Table 6.1: Hypothetical situations regarding detection of a unknown protein target for a novel bioactive small molecule.

hancement of photolysis conditions using a Mercury arc lamp was confirmed by 1HNMR. Probe was rare and this was the extent of examination permissible.

6.2 Introduction

Our strategy was to identify bioactive probes that can be modified into diazirine functionalized versions for target identification strategies without altering potency. The probe CLK042A09 dansyl diazirine was more potent than CLK024F02 dansyl diazirine and selected for use (synthesized in Chapter 2: Organic Synthesis and described quantitatively in Chapter 4: Phenotypes of Bioactives). Given the novelty of the probe there was no *a priori* knowledge of the target. Nonetheless, a number of situations can be hypothesized regarding the potential biological target(s) of CLK042A09 dansyl diazirine along with the most logical means to detect these proteins (Table 6.1). Due to space limitations the most desirable category of probe, an IC50 of high (nM) with high target abundance has a high chance of success for affinity capture and coomasie stain. A number of experiments were performed to identify the target protein of CLK042A09 dansyl diazirine *in vivo*. Factors such as target abundance and potency were of primary concern, but out of our hands.

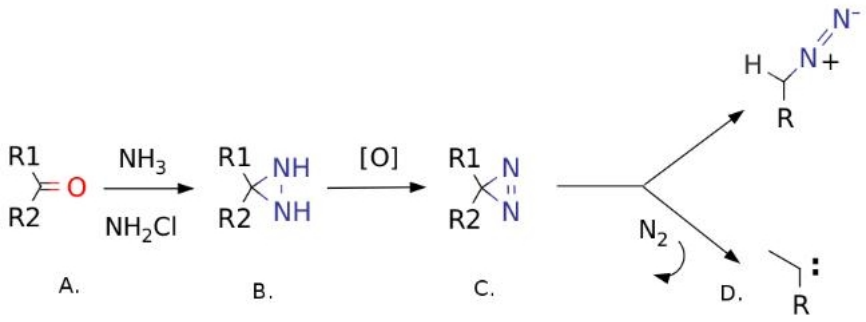

Figure 6.1: Synthesis of diazirine from precursors.

6.2.1 Diazirine is a small photo-activatible probe

Diazirine rings may be prepared in three steps for a myriad of functional groups (Figure 6.1A-C) [55, 52]. Diazirine is formed from starting material in the presence of ammonia and chloroamine which is then oxidized to form an energetic diazirine ring [52, 53] [91]. The product may be photolysed using UV light of 365 nm to liberate nitrogen to produce a carbene *in situ* (Figure 6.1D.)[92]. The short lived carbene is used to label protein(s). Non ideal conditions may cause the generation of an undesired diazo (Figure 6.1D) that produces re-arrangement products [93, 55].

6.2.2 Diazirine for covalent capture of small molecule targets

Covalent capture of protein targets *in vivo* or *in vitro* is a complementary route to target identification of novel small molecules. Precending chapters lay the groundwork for use of diazirine functionalized photo affinity ligands (PALs) of CLK024F02 dansyl and CLK042A09 dansyl.

6.2.3 Promising diazirine probes for use in covalent capture

Dose curve analysis determined that both CLK024F02 dansyl diazirine and CLK042A09 dansyl diazirine both retained activity profiles similar to the block A versions. These results were consistent with the expectation that subtle changes on the acetamide would not disturb bioactivity.The analog to CLK042A09, CLK039G03 dansyl block A, was demonstrated as inactive (Figure 4.21). Therefore, the compounds CLK042A09 dansyl diazirine and CLK039G03 may be used in combination during biochemical experiments as positive and negative controls.

6.3 Results

6.3.1 Experiments to covalently label phosphate buffered saline (PBS) extracted *Arabidopsis* protein

Encouraged by the bioactivity of the diazirine functionalized probes we performed biochemical labeling experiments on *Arabidopsis* proteins extracted with phosphate buffered saline in Laemmli buffer. The PALs were added to a final concentration of 100 μM and added to the protein sample. Protein samples were split into two pools, one for UV exposure, and the other as a control. Samples were irradiated with UV light for 20 minutes on ice, and run on a 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (12%-SDS-PAGE).

Experiments resulted in the labeling of a small abundant 35-55 kDa sized protein for all chemical classes. Early attempts with the diazirine probe heed caution for the interpretation of photolysis experiments where small abundant proteins are sequestered non-specifically [55]. Removal of small abundant proteins may shed light on the observed non-specific labelling. Subsequent experiments were more conservative with the

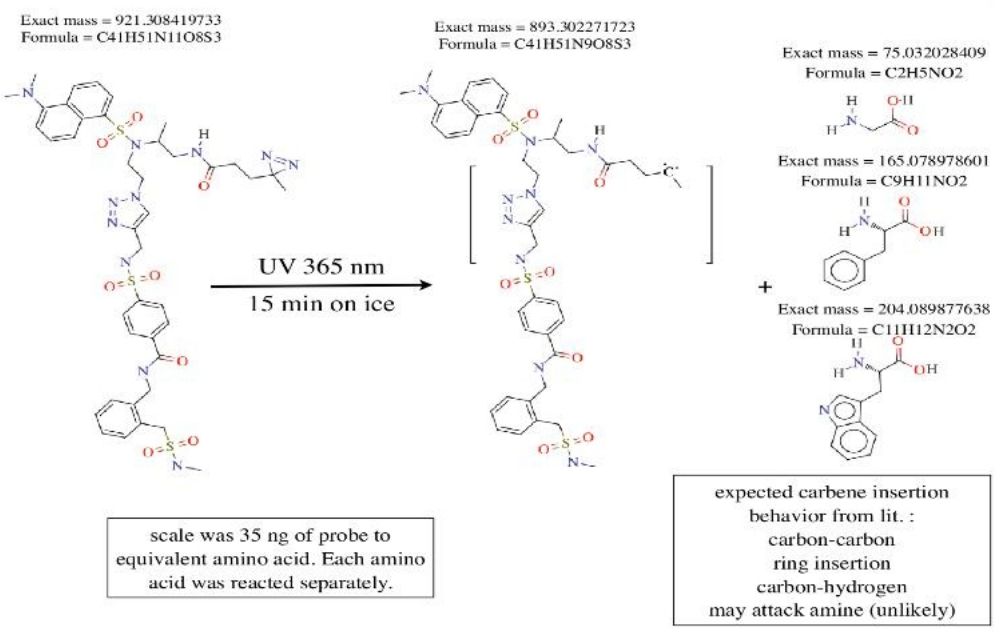

Figure 6.2: Scheme showing expected result for small scale labeling experiment with CLK042A09 dansyl diazirine.

goal to improve photolysis conditions and improve product formation.

6.3.2 Attempts to covalently label BSA

Investigation of the efficiency of the probe CLK042A09 dansyl diazirine to covalently label samples of Bovine Serum Albumin (BSA). Photolysis conditions and electrophoresis conditions were identical to the previous experiment with 20 minute UV exposure at 3 cm distance with a 20 watt source. These conditions were sufficient for activation of our PAL [92]. Despite previous efforts labeling protein in PBS with Laemmli buffer, we were not able to label BSA. The PAL dye was seed to run through the gel ahead of the smallest 66.5 KDa band. It was suspected that the exclusion of Laemmli buffer lead to inefficient labeling, but an insufficient amount of probe remained to continue the investigation.

6.3.3 Attempts to covalently label select amino acids

The remainder of CLK042A09 dansyl diazirine was used in three separate small scale reactions to label glycine, tryptophan, and phenylalanine. The samples were irradiated with UV light for 20 minutes at a distance of less than 1 cm, and the product was subjected to ESI LC/MS (Figure 6.2. Mass analysis confirmed the the presence of starting material, but this is inconclusive. No specific new product was detected.

6.3.4 Enhancement of photolysis conditions using a Mercury arc lamp and optical filter

Rather than exploring a longer duration of photolysis we sought to use a more energetic UV light source. An inverted Leica microscope and an optical filter was used to selectively transmit light of 365 nm from an Olympus Mercury arc lamp. The sample was dissolved in D6-DMSO in an NMR tube and 1HNMR was recorded before and after exposure to UV light for 20 minutes. 1HNMR analysis determined full diazirine photolysis.

6.4 Summary

The type of alkyl diazirine tested, NHS-diazirine, is one of a family of related photo-affinty ligands. Our results can not be generalized to all the variants, especially since the literature has documented subtle enhancements of labeling through replacement of the terminal methyl group on alkyl diazirine to trifluoromethyl [94, 95] [34, 96]. One report of a *p*-[(3-trifluoromethyl)diaizirine-3-yl]benzoic acid was shown to have a half-life of approximately two minutes after exposure to UV light [97]. In addition, a myriad of covalent labels, available as NHS-esters, may have been explored using our

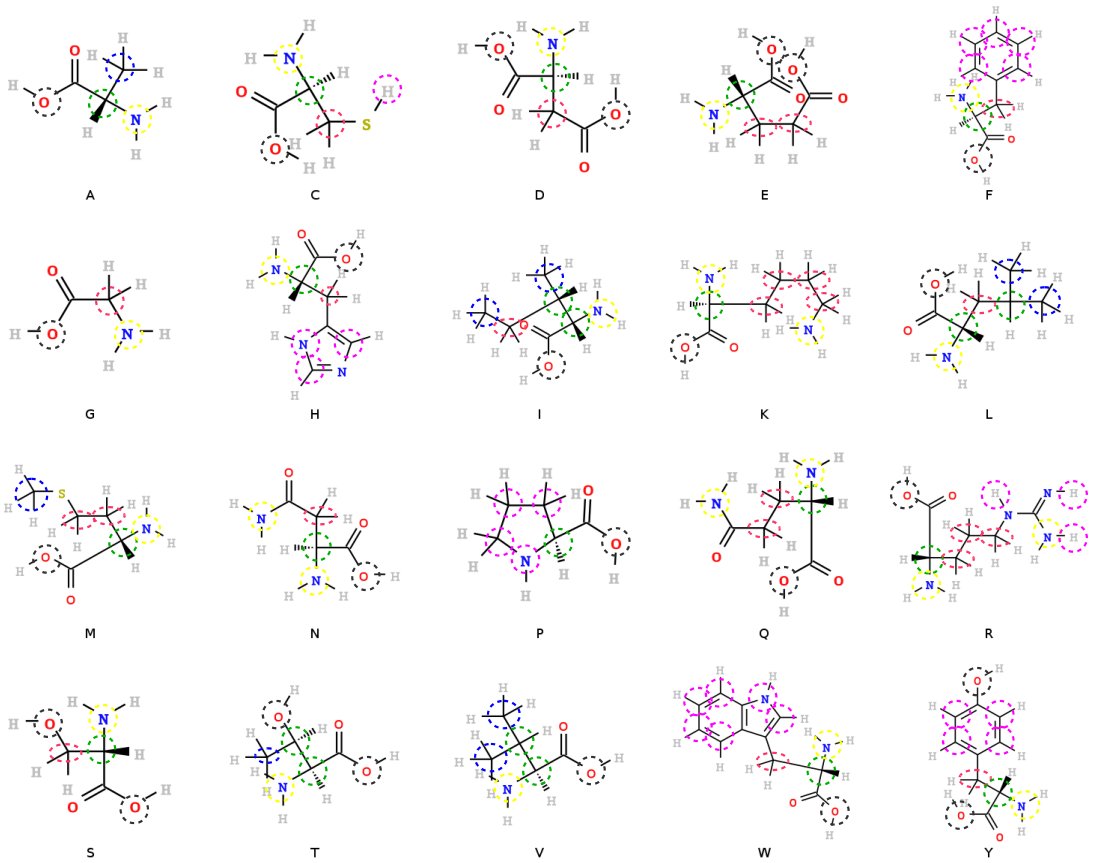

Figure 6.3: Twenty common amino acids are shown with potential sites for reactivity circled. In this original figure different colors represent potential insertion events. It is unknown how the sites in each amino acid would react to a covalent probe as a monomer or as peptide. This figure is an attempt to exhaustively enumerate groups in each amino acid as candidates for alteration by a photoaffinity probe.

building block B such as the epoxide group (K-trap) and benzophenone [98].

6.5 Discussion

6.6 Covalent probes and amino acid reactivity

We generated a number of unique probes and attempted to use them in the most crude way possible. A systematic study of the reactivity of each amino acid, as monomers, and thereafter as peptides would be the best starting point to characterize

the CLK024F02 and CLK042A09 dansyl diazirine covalent probes. A number of sites exist in amino acids that may react with a covalent probe. A systematic study could determine the philicity of a probe to a particular bond or group (Figure 6.3). These studies could provide deep insight into this problem.

6.6.1 Biotin affinity probes

The precursor to the diazirine functionalized 1,2,3-triazoles can potentially react with iodo-acetyl-PEG-biotin thus providing a affinity group rather than a covalent group. The biotinylated molecule may be used in biochemical pull down efforts for target enrichment, and the diazirine probe may be used afterward to label the enriched pools. Furthermore, a biotin probe may be used in combination with phage display to enrich antigens [99], and has been successful in *Arabidopsis* to find FK506 binding proteins using a biotinylated FK506 probe [100].

6.7 Conclusion

Labelling studies may be more successful using a Mercury arc lamp with optical filter. The use of NMR to confirm photolysis prior to mass analysis may reveal more information without complicating the investigation. A systematic set of experiments must be conducted to produce a model of how *in situ* created carbenes from photolyzed PALs can be efficiently utilized. Tests with different solvents, additives, and conditions must be conducted [96]. The use of microprobe NMR may aid in the minimization of probe used per experiment.

More work is necessary to characterize the bioactive PALs CLK024F02 dansyl diazirine and CLK042A09 diazirine for successful biochemical target capture. A number

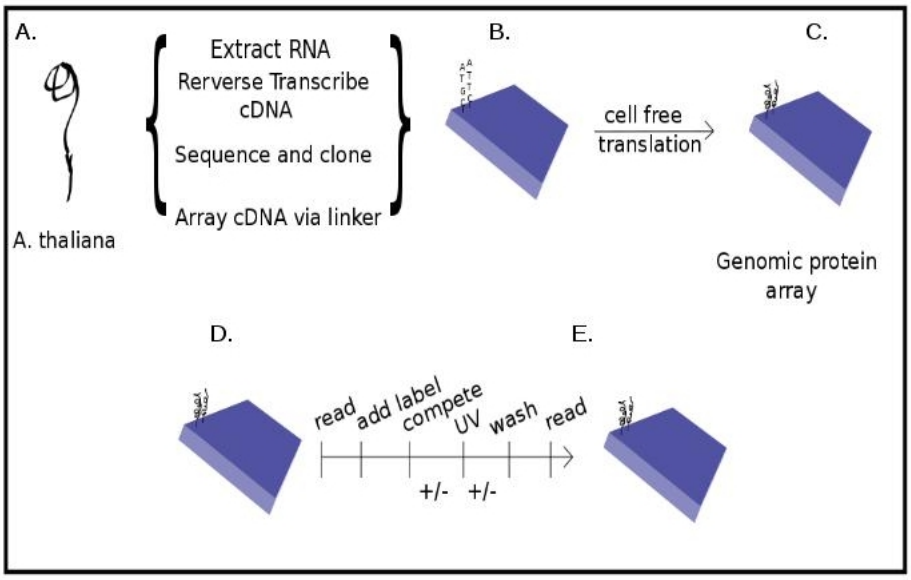

Figure 6.4: Illustration demonstrating the application of photo affinity ligands (PALs) for unbiased genomic protein studies using protein chips from the Arabidopsis Biological Resource Center (ABRC). A-C. Preparation of the chip. D-E. Proposed steps to label proteins using our PALs.

of issues remain as primary concerns for the investigator. One, target abundance is of significance for our probe with a modest IC50 (Figure 4.13 and 4.17). In these instances detection could be enhanced through chemiluminescent detection biotin using avidin-linked horse radish peroxidase (Figure 6.3).

Cheap protein chips are available from the ABRC with standardized coverage genomic translation products (Figure 6.4) [101]. On these protein arrays cross-linked cDNA *in vitro* translated using cell free translation systems leaving protein adhered to the chip. There is potential to use this platform for the identification of covalently labelled proteins, and discount a number of concerns that arise from labeling of unpurified protein from a whole organism.

Even though this study was inconclusive regarding identifying the target, an example has been presented in *Arabidopsis in vivo* where activity is preserved when an acetyl group of the acetamide is replaced by that diazirine acylating agent (Figure

4.17) and (Figure 4.22).This thesis is a demonstration that we can develop photo affinity ligands from fragments of existing probes with modest changes to potency, and a little ingenuity.

Bibliography

[1] A. Saghatelian and B.F. Cravatt. Assignment of protein function in the postgenomic era. *Nature chemical biology*, 1(3):130–142, 2005.

[2] T.F. Chan, J. Carvalho, L. Riles, and XF Zheng. A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (tor). *Proceedings of the National Academy of Sciences*, 97(24):13227–13232, 2000.

[3] C.M. Grozinger, E.D. Chao, H.E. Blackwell, D. Moazed, and S.L. Schreiber. Identification of a class of small molecule inhibitors of the sirtuin family of nad-dependent deacetylases by phenotypic screening. *Journal of Biological Chemistry*, 276(42):38837–38843, 2001.

[4] Y. Zhao, X. Dai, H.E. Blackwell, S.L. Schreiber, and J. Chory. Sir1, an upstream component in auxin signaling identified by chemical genetics. *Science Signalling*, 301(5636):1107, 2003.

[5] X. Dai, K. Hayashi, H. Nozaki, Y. Cheng, and Y. Zhao. Genetic and chemical analyses of the action mechanisms of sirtinol in arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):3129–3134, 2005.

[6] K. Hayashi, A.M. Jones, K. Ogino, A. Yamazoe, Y. Oono, M. Inoguchi, H. Kondo, and H. Nozaki. Yokonolide b, a novel inhibitor of auxin action, blocks degradation of aux/iaa factors. *Journal of Biological Chemistry*, 278(26):23797–23806, 2003.

[7] T. Asami, T. Nakano, H. Nakashita, K. Sekimata, Y. Shimada, and S. Yoshida. The influence of chemical genetics on plant science: shedding light on functions and mechanism of action of brassinosteroids using biosynthesis inhibitors. *Journal of plant growth regulation*, 22(4):336–349, 2003.

[8] N. Raikhel and M. Pirrung. Adding precision tools to the plant biologists' toolbox with chemical genomics. *Plant physiology*, 138(2):563–564, 2005.

[9] P. McCourt and C. Benning. Arabidopsis: A rich harvest 10 years after completion of the genome sequence. *The Plant Journal*, 61(6):905–908, 2010.

[10] C. Town, R. Schmidt, and I. Bancroft. Comparative genome analysis at the sequence level in the brassicaceae. *Genetics and Genomics of the Brassicaceae*, pages 171–194, 2011.

[11] S.J. Clough and A.F. Bent. Floral dip: a simplified method for agrobacterium-mediated transformation of arabidopsis thaliana. *The Plant Journal*, 16(6):735–743, 1998.

[12] E.M. Knee, L. Rivero, D. Crist, E. Grotewold, and R. Scholl. Germplasm and molecular resources. *Genetics and Genomics of the Brassicaceae*, pages 437–467, 2011.

[13] M. Koornneef, M. Reymond, and C. Alonso-Blanco. Natural variation in arabidopsis thaliana. *Genetics and Genomics of the Brassicaceae*, pages 123–151, 2011.

[14] M.A. Lysak and M.A. Koch. Phylogeny, genome, and karyotype evolution of crucifers (brassicaceae). *Genetics and Genomics of the Brassicaceae*, pages 1–31, 2011.

[15] C.F. Quiros and M.W. Farnham. The genetics of brassica oleracea. *Genetics and Genomics of the Brassicaceae*, pages 261–289, 2011.

[16] N. Ramchiary and Y.P. Lim. Genetics of brassica rapa l. *Genetics and Genomics of the Brassicaceae*, pages 215–260, 2011.

[17] I. Parkin. Chasing ghosts: comparative mapping in the brassicaceae. *Genetics and Genomics of the Brassicaceae*, pages 153–170, 2011.

[18] O. Savolainen and H. Kuittinen. Arabidopsis lyrata genetics. *Genetics and Genomics of the Brassicaceae*, pages 347–372, 2011.

[19] G. Theißen. The genetics of capsella. *Genetics and Genomics of the Brassicaceae*, pages 373–387, 2011.

[20] F.B. Pickett and D.R. Meeks-Wagner. Seeing double: appreciating genetic redundancy. *The Plant Cell*, 7(9):1347, 1995.

[21] N. Bouché, D. Bouchez, et al. Arabidopsis gene knockout: phenotypes wanted. *Current opinion in plant biology*, 4(2):111, 2001.

[22] S. Cutler and P. McCourt. Dude, where's my phenotype? dealing with redundancy in signaling networks. *Plant physiology*, 138(2):558–559, 2005.

[23] D.C. Boyes, A.M. Zayed, R. Ascenzi, A.J. McCASKILL, N.E. Hoffman, K.R. Davis, and J. Görlach. Growth stage–based phenotypic analysis of arabidopsis a model for high throughput functional genomics in plants. *The Plant Cell Online*, 13(7):1499–1510, 2001.

[24] S.B. Primrose and R. Twyman. *Principles of genome analysis and genomics*. Wiley-Blackwell, 2009.

[25] G.C. Briggs, K.S. Osmont, C. Shindo, R. Sibout, and C.S. Hardtke. Unequal genetic redundancies in arabidopsis: a neglected phenomenon? *Trends in plant science*, 11(10):492–498, 2006.

[26] S.Y. Park, P. Fung, N. Nishimura, D.R. Jensen, H. Fujii, Y. Zhao, S. Lumba, J. Santiago, A. Rodrigues, T.F. Chow, et al. Abscisic acid inhibits type 2c protein phosphatases via the pyr/pyl family of start proteins. *Science Signalling*, 324(5930):1068, 2009.

[27] T.J. Mitchison et al. Towards a pharmacological genetics. *Chemistry & biology*, 1(1):3, 1994.

[28] L. Burdine and T. Kodadek. Target identification in chemical genetics: the (often) missing link. *Chemistry & biology*, 11(5):593, 2004.

[29] A. Adibekian, B.R. Martin, C. Wang, K.L. Hsu, D.A. Bachovchin, S. Niessen, H. Hoover, and B.F. Cravatt. Click-generated triazole ureas as ultrapotent in vivo–active serine hydrolase inhibitors. *Nature Chemical Biology*, 7(7):469–478, 2011.

[30] D.P. Walsh, Y.T. Chang, et al. Chemical genetics. *Chemical Reviews-Columbus*, 106(6):2476–2530, 2006.

[31] C.M. Dobson. Chemical space and biology. *Nature*, 432(7019):824–828, 2004.

[32] S.L. Schreiber. The small-molecule approach to biology. *Chem. Eng. News*, 81(9):51–61, 2003.

[33] S.L. Schreiber et al. Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorganic and Medicinal Chemistry*, 6(8):1127–1152, 1998.

[34] A. Blencowe and W. Hayes. Development and application of diazirines in biological and synthetic macromolecular systems. *Soft Matter*, 1(3):178–205, 2005.

[35] M.R. Bond, H. Zhang, P.D. Vu, and J.J. Kohler. Photocrosslinking of glycoconjugates using metabolically incorporated diazirine-containing sugars. *Nature protocols*, 4(7):1044–1063, 2009.

[36] M.C. Pirrung. *Molecular diversity and combinatorial chemistry: principles and applications*, volume 24. Elsevier Science Limited, 2004.

[37] R.J. Simon, R.S. Kania, R.N. Zuckermann, V.D. Huebner, D.A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, and C.K. Marlowe. Peptoids: a modular approach to drug discovery. *Proceedings of the National Academy of Sciences*, 89(20):9367–9371, 1992.

[38] B. Dörner, G.M. Husar, J.M. Ostresh, and R.A. Houghten. The synthesis of peptidomimetic combinatorial libraries through successive amide alkylations. *Bioorganic & Medicinal Chemistry*, 4(5):709–715, 1996.

[39] T.R. Chan, R. Hilgraf, K.B. Sharpless, and V.V. Fokin. Polytriazoles as copper (i)-stabilizing ligands in catalysis. *Organic letters*, 6(17):2853–2855, 2004.

[40] F. Himo, T. Lovell, R. Hilgraf, V.V. Rostovtsev, L. Noodleman, K.B. Sharpless, and V.V. Fokin. Copper (i)-catalyzed synthesis of azoles. dft study predicts unprecedented reactivity and intermediates. *Journal of the American Chemical Society*, 127(1):210–216, 2005.

[41] S.M. Wignall, N.S. Gray, Y.T. Chang, L. Juarez, R. Jacob, P.G. Schultz, R. Heald, et al. Identification of a novel protein regulating microtubule stability through a chemical approach. *Chemistry & biology*, 11(1):135–146, 2004.

[42] H.S. Moon, E.M. Jacobson, S.M. Khersonsky, M.R. Luzung, D.P. Walsh, W. Xiong, J.W. Lee, P.B. Parikh, J.C. Lam, T.W. Kang, et al. A novel microtubule destabilizing entity from orthogonal synthesis of triazine library and zebrafish embryo screening. *Journal of the American Chemical Society*, 124(39):11608–11609, 2002.

[43] D. Williams, D.W. Jung, S.M. Khersonsky, N. Heidary, Y.T. Chang, and S.J. Orlow. Identification of compounds that bind mitochondrial f1f0 atpase by screening a triazine library for correction of albinism. *Chemistry & biology*, 11(9):1251–1259, 2004.

[44] L. Ni-Komatsu, J.K. Leung, D. Williams, J. Min, S.M. Khersonsky, Y.T. Chang, and S.J. Orlow. Triazine-based tyrosinase inhibitors identified by chemical genetic screening. *Pigment cell research*, 18(6):447–453, 2005.

[45] H.C. Kolb, MG Finn, and K.B. Sharpless. Click chemistry: diverse chemical function from a few good reactions. *Angewandte Chemie International Edition*, 40(11):2004–2021, 2001.

[46] N.J. Agard, J.M. Baskin, J.A. Prescher, A. Lo, and C.R. Bertozzi. A comparative study of bioorthogonal reactions with azides. *ACS Chemical Biology*, 1(10):644–648, 2006.

[47] J.J. Irwin and B.K. Shoichet. Zinc-a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182, 2005.

[48] A. Benalil, B. Carboni, and M. Vaultier. Synthesis of 1, 2-aminoazides. conversion to unsymmetrical vicinal diamines by catalytic hydrogenation or reductive alkylation with dichloroboranes. *Tetrahedron*, 47(38):8177–8194, 1991.

[49] HD Dakin and R. West. A general reaction of amino acids. *Journal of Biological Chemistry*, 78(1):91–104, 1928.

[50] JD Hepworth. Aminoacetone semicarbazone hydrochloride. *Org. Synth*, 45(1):3, 1976.

[51] P. Kocienski. *Protecting Groups (softcover)*. Thieme, 2005.

[52] R.A. Moss, W. Ma, D.C. Merrer, and S. Xue. Conversion of obstinate nitriles to amidines by garigipati's reaction. *Tetrahedron letters*, 36(48):8761–8764, 1995.

[53] D.M. Henze, O. Kreye, S. Brauch, C. Nitsche, K. Naumann, L.A. Wessjohann, and B. Westermann. Photoaffinity-labeled peptoids and depsipeptides by multicomponent reactions. *Synthesis*, 2010(17):2997–3003.

[54] A.L. MacKinnon, J.L. Garrison, R.S. Hegde, and J. Taunton. Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *Journal of the American Chemical Society*, 129(47):14560–14561, 2007.

[55] E. Work, W.H. Evans, C.R. Lowe, et al. *Laboratory techniques in biochemistry and molecular biology*, volume 7. Elsevier, 1979.

[56] H. Bayley. *Photogenerated reagents in biochemistry and molecular biology*, volume 12. Elsevier Science, 1983.

[57] Y. Cao, A. Charisi, L.C. Cheng, T. Jiang, and T. Girke. Chemminer: a compound mining framework for r. *Bioinformatics*, 24(15):1733–1734, 2008.

[58] C.A. Lipinski. Drug-like properties and the causes of poor solubility and poor permeability. *Journal of pharmacological and toxicological methods*, 44(1):235–249, 2000.

[59] M. Pennacchio, L.V. Jefferson, and K. Havens. Arabidopsis thaliana: A new test species for phytotoxic bioassays. *Journal of chemical ecology*, 31(8):1877–1885, 2005.

[60] T. Tojo, K. Tsuda, TS Wada, K. Yamazaki, et al. A simple and extremely sensitive system for detecting estrogenic activity using transgenic arabidopsis thaliana. *Ecotoxicology and environmental safety*, 64(2):106, 2006.

[61] A. Lehman, R. Black, J.R. Ecker, et al. Hookless1, an ethylene response gene, is required for differential cell elongation in the arabidopsis hypocotyl. *Cell*, 85(2):183, 1996.

[62] P. Westhoff. *Molecular plant development from gene to plant*. Oxford, etc.: Oxfod University Press, 1998.

[63] P. Guzman and J.R. Ecker. Exploiting the triple response of arabidopsis to identify ethylene-related mutants. *The Plant Cell Online*, 2(6):513–523, 1990.

[64] T. Ishida, S. Thitamadee, and T. Hashimoto. Twisted growth and organization of cortical microtubules. *Journal of plant research*, 120(1):61–70, 2007.

[65] Freeman Tsz-fung Chow. *LATCA: a library of biologically active small molecules for plant chemical genomics*. University of Toronto (Canada), 2007.

[66] Y. Zhao, T.F. Chow, R.S. Puckrin, S.E. Alfred, A.K. Korir, C.K. Larive, and S.R. Cutler. Chemical genetic interrogation of natural variation uncovers a molecule that is glycoactivated. *Nature chemical biology*, 3(11):716–721, 2007.

[67] J.P. Vogel, P. Schuerman, K. Woeste, I. Brandstatter, and J.J. Kieber. Isolation and characterization of arabidopsis mutants defective in the induction of ethylene biosynthesis by cytokinin. *Genetics*, 149(1):417–427, 1998.

[68] S. DeBolt, R. Gutierrez, D.W. Ehrhardt, C.V. Melo, L. Ross, S.R. Cutler, C. Somerville, and D. Bonetta. Morlin, an inhibitor of cortical microtubule dynamics and cellulose synthase movement. *Proceedings of the National Academy of Sciences*, 104(14):5854–5859, 2007.

[69] M. Nakamura, K. Naoi, T. Shoji, and T. Hashimoto. Low concentrations of propyzamide and oryzalin alter microtubule dynamics in arabidopsis epidermal cells. *Plant and cell physiology*, 45(9):1330–1334, 2004.

[70] Peter J Rousseeuw, Ida Ruts, and John W Tukey. The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387, 1999.

[71] Krishnamoorthy Sivakumar, Fang Xie, Brandon M Cash, Su Long, Hannah N Barnhill, and Qian Wang. A fluorogenic 1, 3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Organic letters*, 6(24):4603–4606, 2004.

[72] W. Lukowitz, C.S. Gillmor, and W.R. Scheible. Positional cloning in arabidopsis. why it feels good to have a genome initiative working for you. *Plant Physiology*, 123(3):795–806, 2000.

[73] H.E. Blackwell and Y. Zhao. Chemical genetic approaches to plant biology. *Plant Physiology*, 133(2):448–455, 2003.

[74] A. Kilaru, E.B. Blancaflor, B.J. Venables, S. Tripathy, K.S. Mysore, and K.D. Chapman. The n-acylethanolamine-mediated regulatory pathway in plants. *Chemistry & biodiversity*, 4(8):1933–1955, 2007.

[75] N.D. Teaster, C.M. Motes, Y. Tang, W.C. Wiant, M.Q. Cotter, Y.S. Wang, A. Kilaru, B.J. Venables, K.H. Hasenstein, G. Gonzalez, et al. N-acylethanolamine metabolism interacts with abscisic acid signaling in arabidopsis thaliana seedlings. *The Plant Cell Online*, 19(8):2454–2469, 2007.

[76] S.C. Kim, K.D. Chapman, and E.B. Blancaflor. Fatty acid amide lipid mediators in plants. *Plant science*, 178(5):411–419, 2010.

[77] EA Dikusar, NG Kozlov, SS Koval'skaya, LA Popova, and KL Moiseichuk. Synthesis and reactions of ethynyl derivatives of adamantanone, verbanone, and verbenone. *Russian Journal of General Chemistry*, 71(2):290–293, 2001.

[78] A. Bannigan, A.M.D. Wiedemeier, R.E. Williamson, R.L. Overall, and T.I. Baskin. Cortical microtubule arrays lose uniform alignment between cells and are oryzalin resistant in the arabidopsis mutant, radially swollen 6. *Plant and cell physiology*, 47(7):949–958, 2006.

[79] P.G. Bartels and J.L. Hilton. Comparison of trifluralin, oryzalin, pronamide, propham, and colchicine treatments on microtubules. *Pesticide Biochemistry and Physiology*, 3(4):462–472, 1973.

[80] Tobias I Baskin, Jan E Wilson, Ann Cork, and Richard E Williamson. Morphology and microtubule organization in arabidopsis roots exposed to oryzalin or taxol. *Plant and Cell Physiology*, 35(6):935–942, 1994.

[81] LC Morejohn, TE Bureau, J. Mole-Bajer, AS Bajer, and DE Fosket. Oryzalin, a dinitroaniline herbicide, binds to plant tubulin and inhibits microtubule polymerization in vitro. *Planta*, 172(2):252–264, 1987.

[82] E. Gendreau, J. Traas, T. Desnos, O. Grandjean, M. Caboche, and H. Hofte. Cellular basis of hypocotyl growth in arabidopsis thaliana. *Plant Physiology*, 114(1):295–305, 1997.

[83] E. Gendreau, H. Höfte, O. Grandjean, S. Brown, and J. Traas. Phytochrome controls the number of endoreduplication cycles in the arabidopsis thaliana hypocotyl. *The Plant Journal*, 13(2):221–230, 1998.

[84] Y. Kim, K.S. Schumaker, and J. Zhu. Ems mutagenesis of arabidopsis. *Methods in Molecular Biology-Clifton-Totowa*, 323:101, 2006.

[85] D.R. Page, U. Grossniklaus, et al. The art and design of genetic screens: Arabidopsis thaliana. *Nature Reviews Genetics*, 3(2):124–136, 2002.

[86] Jonathan Lightner and Timothy Caspar. Seed mutagenesis of arabidopsis. In *Arabidopsis protocols*, pages 91–102. Springer, 1998.

[87] S. Ossowski, K. Schneeberger, J.I. Lucas-Lledó, N. Warthmann, R.M. Clark, R.G. Shaw, D. Weigel, and M. Lynch. The rate and molecular spectrum of spontaneous mutations in arabidopsis thaliana. *Science*, 327(5961):92–94, 2010.

[88] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218):53–59, 2008.

[89] R.S. Austin, D. Vidaurre, G. Stamatiou, R. Breit, N.J. Provart, D. Bonetta, J. Zhang, P. Fung, Y. Gong, P.W. Wang, et al. Next-generation mapping of arabidopsis genes. *The Plant Journal*, 67(4):715–725, 2011.

[90] P. Lamesch, T.Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D.L. Alexander, M. Garcia-Hernandez, et al. The arabidopsis information resource (tair): improved gene annotation and new tools. *Nucleic acids research*, 40(D1):D1202–D1210, 2012.

[91] Y. Tanaka and J.J. Kohler. Photoactivatable crosslinking sugars for capturing glycoprotein interactions. *Journal of the American Chemical Society*, 130(11):3278–3279, 2008.

[92] P. Klán and J. Wirz. *Photochemistry of organic compounds: from concepts to practice*. Wiley-Blackwell, 2009.

[93] K.E. Bergmann, K.E. Carlson, and J.A. Katzenellenbogen. Hexestrol diazirine photoaffinity labeling reagent for the estrogen receptor. *Bioconjugate Chemistry*, 5(2):141–150, 1994.

[94] M. Huss, G. Ingenhorst, S. König, M. Gaßel, S. Dröse, A. Zeeck, K. Altendorf, and H. Wieczorek. Concanamycin a, the specific inhibitor of v-atpases, binds to the vo subunit c. *Journal of Biological Chemistry*, 277(43):40544–40548, 2002.

[95] H. Zhang, K.A. Lerro, T. Yamamoto, T.H. Lien, L. Sastry, M.A. Gawinowicz, and K. Nakanishi. The location of the chromophore in rhodopsin: a photoaffinity study. *Journal of the American Chemical Society*, 116(22):10165–10173, 1994.

[96] J. Brunner, H. Senn, and FM Richards. 3-trifluoromethyl-3-phenyldiazirine. a new carbene generating group for photolabeling reagents. *Journal of Biological Chemistry*, 255(8):3313–3318, 1980.

[97] Y. Hatanaka, M. Hashimoto, and Y. Kanaoka. A novel biotinylated heterobifunctional cross-linking reagent bearing an aromatic diazirine. *Bioorganic & medicinal chemistry*, 2(12):1367–1373, 1994.

[98] F. Krauth, C.H. Ihling, H.H. Rüttinger, and A. Sinz. Heterobifunctional isotope-labeled amine-reactive photo-cross-linker for structural investigation of proteins by matrix-assisted laser desorption/ionization tandem time-of-flight and electrospray ionization ltq-orbitrap mass spectrometry. *Rapid Communications in Mass Spectrometry*, 23(17):2811–2818, 2009.

[99] G.P. Smith et al. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (New York, NY)*, 228(4705):1315, 1985.

[100] A.M. Piggott, A.M. Kriegel, R.D. Willows, and P. Karuso. Rapid isolation of novel fk506 binding proteins from multiple organisms using gdna and cdna t7 phage display. *Bioorganic & medicinal chemistry*, 17(19):6841–6850, 2009.

[101] H. Zhu and M. Snyder. Protein chip technology. *Current opinion in chemical biology*, 7(1):55–63, 2003.

[102] Y.L. Angell and K. Burgess. Peptidomimetics via copper-catalyzed azide–alkyne cycloadditions. *Chem. Soc. Rev.*, 36(10):1674–1689, 2007.