University of Massachusetts Amherst [ScholarWorks@UMass Amherst](https://scholarworks.umass.edu/)

[Masters Theses](https://scholarworks.umass.edu/masters_theses_2) Contract Contra

August 2014

Metabolic Modeling of Secondary Metabolism in Plant Systems

Lisa M. Leone University of Massachusetts Amherst

Follow this and additional works at: [https://scholarworks.umass.edu/masters_theses_2](https://scholarworks.umass.edu/masters_theses_2?utm_source=scholarworks.umass.edu%2Fmasters_theses_2%2F27&utm_medium=PDF&utm_campaign=PDFCoverPages)

Part of the [Biochemical and Biomolecular Engineering Commons](http://network.bepress.com/hgg/discipline/241?utm_source=scholarworks.umass.edu%2Fmasters_theses_2%2F27&utm_medium=PDF&utm_campaign=PDFCoverPages), and the [Molecular, Cellular, and](http://network.bepress.com/hgg/discipline/236?utm_source=scholarworks.umass.edu%2Fmasters_theses_2%2F27&utm_medium=PDF&utm_campaign=PDFCoverPages) [Tissue Engineering Commons](http://network.bepress.com/hgg/discipline/236?utm_source=scholarworks.umass.edu%2Fmasters_theses_2%2F27&utm_medium=PDF&utm_campaign=PDFCoverPages)

Recommended Citation

Leone, Lisa M., "Metabolic Modeling of Secondary Metabolism in Plant Systems" (2014). Masters Theses. 27.

[https://scholarworks.umass.edu/masters_theses_2/27](https://scholarworks.umass.edu/masters_theses_2/27?utm_source=scholarworks.umass.edu%2Fmasters_theses_2%2F27&utm_medium=PDF&utm_campaign=PDFCoverPages)

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu.](mailto:scholarworks@library.umass.edu)

METABOLIC MODELING OF SECONDARY METABOLISM IN PLANT SYSTEMS

A Thesis Presented

by

LISA M. LEONE

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE IN CHEMICAL ENGINEERING

May 2014

Department of Chemical Engineering

METABOLIC MODELING OF SECONDARY METABOLISM IN PLANT SYSTEMS

A Thesis Presented

By

LISA M. LEONE

Approved as to style and content by:

Susan Roberts , Chair

Michael Henson, Member

Jeffrey Blanchard, Member

T. J. Mountziaris, Department Head

Chemical Engineering

DEDICATION

This thesis is dedicated to the memory of Babur Z. Khalique, whose passing was more absurd and untimely than my own.

EPIGRAPH

Myself when young did eagerly frequent Doctor and Saint, and heard great Argument About it and about: but evermore Came out of the same Door as in I went.

With them the Seed of Wisdom did I sow, And with my own hand labour'd it to grow: And this was all the Harvest that I reap'd— "I came like Water, and like Wind I go."

- Omar Khayyam

ACKNOWLEDGMENTS

I would like to thank my advisors, Susan Roberts and Mike Henson, for supervising my work and teaching me so much about what academic science is about.

I would like to thank my department head Lakis, and also Marie Wallace, for helping me navigate the bureaucracy necessary to complete my degree.

I would like to thank my labmates in the Roberts lab (Sarah, Rohan, Whitney and Marty) as well as my colleagues in the Chemical Engineering department, for making the grad school experience a little more bearable.

I would like to thank my parents, Robert and Ellen Leone, for supporting me in all my decisions, including my decision to leave graduate school. I would also like to thank my future in-laws, Nan Jacobson Wise and John Wise, for being supportive of me during my transitive period.

Most of all I would like to thank my fiancée, Adam Wise, for helping me to be brave enough to leave a program that was making me miserable, for helping me in every way imaginable to transition to the working world, and for being the most amazing person I've ever met.

ABSTRACT

METABOLIC MODELING OF SECONDARY METABOLISM IN PLANT SYSTEMS

MAY 2014

LISA M. LEONE, B.S., DREXEL UNIVERSITY

M.S.Ch.E, UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Susan C. Roberts

In the first part of this research, we constructed a Genome scale Metabolic Model (GEM) of *Taxus cuspidata*, a medicinal plant used to produce paclitaxel (Taxol®). The construction of the *T. cuspidata* GEM was predicated on recent acquisition of a transcriptome of *T. cuspidata* metabolism under methyl jasmonate (MJ) elicited conditions (when paclitaxel is produced) and unelicited conditions (when paclitaxel is not produced). Construction of the draft model, in which transcriptomic data from elicited and unelicited conditions were included, utilized tools including the ModelSEED developed by Argonne National Laboratory. Although a model was successfully created and gapfilled by ModelSEED using their software, we were not able to reproduce their results using COBRA, a widely accepted FBA software package. Further work needs to be done to figure out how to run ModelSEED models on commonly available software.

In the second part of this research, we modeled the MJ elicited/defense response phenotype in *Arabidopsis thaliana*. Previously published models of *A. thaliana* were tested for suitability in modeling the MJ elicited phenotype using publicly available computation tools. MJ elicited and unelicited datasets were compared to ascertain differences in metabolism between these two phenotypes. The MJ elicited and unelicited datasets were significantly different in many respects, including the expression levels of many genes associated with secondary metabolism. However, it was found that the expression of genes related to growth and central metabolism were not

generally significantly different for the MJ+ and MJ- datasets, the pathways associated with secondary metabolism were incomplete and could not be modeled, and FBA methods did not show the difference in growth that was expected. These results suggest that behavior associated with the MJ+ phenotype such as slow growth and secondary metabolite production may be controlled by factors not easily modeled with transcriptome data alone.

Additional research was performed in the area of cryosectioning and immunostaining of fixed *Taxus* aggregates. Protocols developed for this work can be found in Appendix B.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PLANT SECONDARY METABOLISM

Much of this text was taken from (Leone & Roberts 2013) with permission from Springer.

Paradigms of Anti-Cancer Drug Discovery

From 1981 to 2002, 74% of the anticancer agents approved by the FDA were inspired by, derived from, or true natural products (NPs) (Wilson & Danishefsky, 2006). Biologically-derived compounds often exhibit "privileged structures" in terms of biological activity (Evans et al. 1988). Unfortunately, these structures are often chirally complex and pose unique challenges for large-scale commercial production. For a significant number of NPs, a route to chemical synthesis does not exist, and for the vast majority of NPs, total chemical synthesis methods are not commercially viable due to structural complexity, low yields and environmental concerns (Kolewe et al. 2008). In the 1970s and 1980s, a lack of immediate production strategies for some NPs entering clinical trials led to a perceived "supply crisis" for certain drugs, including the anticancer compounds paclitaxel and camptothecin. These situations led to reluctance in the pharmaceutical industry for investment in NPs.

The ease of chemical synthesis in simple, non-NP-derived molecules is perhaps one reason why synthetic combinatorial libraries became popular in drug development (Ortholand & Ganesan 2004). The total chemical synthesis of a synthetic combinatorial library's components as a starting point ensures that every compound will have a relatively simple (and known) route for eventual large-scale production. A disappointing result of this cost-saving methodology is that the systematic exploration of organic chemistry space has been limited to known and easily synthesizable structures, which are not necessarily biologically relevant (Lipkus et al. 2008). Among the millions of compounds screened in academia and industry, only a relatively small

number of compounds from high-throughput screens have been shown to exhibit activity against drug targets, and even fewer have overcome toxicity issues and moved to clinical trials (Newman & Cragg 2007). Because of these difficulties, synthetic combinatorial libraries have shifted over the past decade from large collections of simple compounds to a diversity-centered exploration of a smaller chemistry space, aided by computational technologies (Schnur et al. 2011).

With this new focus, the drug industry has reverted back to a paradigm in which natural organisms have unique advantages over a chemist's bench. Plants, marine organisms and microbes still represent an extremely diverse and relatively untapped space for lead discovery. The NP research of the past 30 years has made rapid progress in alleviating possible supply issues through a variety of advanced technologies, including plant cell culture (PCC) and heterologous expression in microbial systems, as discussed below.

Natural Product Diversity and Production

Figure 1 – Anti-cancer natural products derived from plants.

Naturally occurring plants, microbes and fungi have been the source of a staggering amount of NP diversity throughout the history of drug discovery. In the case of plants, over 10,000 alkaloids and 23,000 terpenoids have been characterized, and these most certainly are an underestimation

of the capabilities of plants, of which only a small portion of the world's species are estimated to have been sampled (Cheng et al. 2007). Many of these NPs (Figure 1) have medicinal qualities.

Most plant NPs fall into the class of secondary metabolites (SMs), so called because they do not necessarily serve primary metabolic functions in the growth and maintenance of their native plant. Biosynthesis of plant SMs is complex, involving many different precursors from primary metabolism. Some of these pathways are interconnected, for example, all of the compounds within the alkaloid and terpenoid classes of NPs originate in part from the common precursor isopentenyl diphosphate (IPP) and its allylic isomer dimethylallydiphosphate (DMAPP). Biosynthesis of many SMs often requires dozens of enzymes acting in concert, and as a result is under strict metabolic control. The abbreviated biosynthetic pathways of paclitaxel and the Vinca alkaloids in *Taxus* spp. and *Catharanthus roseus*, respectively, are shown in [Figure 2.](#page-16-0) While SMs do not play a role in primary metabolism in that they do not directly pertain to plant growth or maintenance, many plant NPs have shown activity in nature as insecticides and anti-fungals, and are thought to confer evolutionary advantages to their native plant (Hartmann 2007).

Figure 2 – Biosynthetic pathways of paclitaxel and Vinca alkaloids in *Taxus* and *C. roseus* plants respectively.

The "screening" hypothesis of NP diversity postulates that the chances of any single NP possessing a potent bioactivity, as a fungicide or insecticide for example, is very low. Therefore, it is advantageous for plants to produce many slightly different metabolites in small concentrations, in the hopes that at least one will be useful against an external stressor, such as herbivorous insects (Firn & Jones 2003). In this perspective, a single plant can be compared to a naturally occurring drug discovery program, in which many compounds with slightly different functionalities are produced from simple starting structures. Plants accomplish this incredible diversity at minimal biosynthetic cost by employing promiscuous enzymes, using branched and matrix biosynthetic pathways, and generally producing minute quantities of these SMs (Fischbach & Clardy 2007).

Different amounts of pure product are needed at different stages of development. After a promising compound has been identified, milligram quantities are needed for structural determination, multiple grams are needed for clinical development, and multiple kilograms are needed for initial clinical trials (Cragg, Boyd, Grever, Schepartz, & Grever, 1995; Koehn & Carter, 2005). If a novel compound is found to be effective in clinical trials and approved for use, supply will need to quickly ramp up to many kilograms for commercialization. As shown in the story of paclitaxel (see below), supply issues can be a major hurdle for NPs. Supply issues are especially untenable in the treatment of terminal diseases such as cancer, where patients cannot wait months or years for a supply route to be established. Fortunately, the field of NP synthesis has matured since the paclitaxel supply crisis of the 1990s and there now exist many options for supply depending on the properties of the NP and plant source.

Natural Harvest

Natural harvest is the method by which many plant NPs have initially been collected, and for intermediate to long term production it is an option when 1) the plant producing the compound of interest is fast growing and easily cultivable, as in the case of *C. roseus*, or 2) when the product of interest is produced in relatively large quantities *in planta*, as in the case of podophyllotoxin in *Podophyllum hexandrum* rhizomes (Farkya et al. 2004). For many naturally occurring compounds neither of these conditions ismet, and environmental concerns can also arise when a plant is endangered or vital to a local ecosystem. Additionally, in the case of novel NPs found through strategies such as combinatorial biosynthesis, the native plant will generally not produce the compound at all.

Chemical Synthesis

Generally a default for the chemical industry, chemical synthesis is an option when the compound of interest is relatively simple and easy to synthesize, as in the case of flavopiridol, an anti-cancer NP in clinical trials that is currently supplied via total synthesis (Naik et al. 1988). Complete chemical syntheses have been published for many NPs of interest, such as paclitaxel (Nicolau et al. 1994; Holton et al. 1994), but they often do not represent a viable, scaleable production method due to low yields, complexity and number of steps, as well as the use of environmentally

unfriendly chemicals and solvents. Chemical semi-synthesis can be used to create NP derivatives with more favorable pharmacological properties, as in the case of the camptothecin derivatives irinotecan and topotecan, and in some cases, can be used to derive a NP from precursors that occur naturally at high levels. For example, a paclitaxel intermediate, 10-deacetylbaccatin III, was chemically modified to produce a semi-synthetic version of paclitaxel in the 1990s (Denis et al. 1988) and was used as the primary supply route for paclitaxel by Bristol-Myers Squibb for over a decade.

Heterologous Synthesis in Microbes

Transfer of plant NP pathways to microbial hosts has advanced rapidly in the past decade. The driving force behind this development is the perceived advantage of microbes, such as *Escherichia coli* and *Saccharomyces cerevisiae*, as production organisms for fermentation, because they are well-studied model systems; additionally, a mature infrastructure exists for the fermentation of these microbes at large scales. One of the first success stories in this field was heterologous expression in yeast (Ro et al. 2006) and *E. coli* (Dietrich et al. 2009) of enzymes to synthesize precursors to artemesinin, an anti-malarial agent from the plant *Artemisia annua*. Under a tripartite partnership of Amyris Biotechnologies, the Institute for One World Health and the pharmaceutical company Sanofi-Aventis, the fermentation for artemesinin is currently moving into the commercial phase and is expected to begin production in 2012 (Hale et al. 2007; Solomon 2011). *Efforts* to transfer metabolic pathways for isoflavones (Leonard et al. 2008), taxanes (Ajikumar et al. 2010), and alkaloids (Nakagawa et al. 2011) are currently underway. However, major hurdles to microbial biosynthesis still exist, including incomplete NP biosynthesis pathway knowledge and access to cloned genes, as well as the difficulty in expressing plant enzymes in microbial systems.

Despite much research, knowledge of the biosynthetic pathways of many important plant NPs remain incomplete *in planta*. For example, in the paclitaxel pathway there are thought to be approximately 19 steps, of which several remain completely unknown, and among the known steps, enzyme identification is incomplete (Croteau et al. 2006). The formation of the oxetane ring in particular is essential to the clinical function of paclitaxel (Kingston 1994), but enzyme candidates for this reaction have not yet been found. A special difficulty in identifying biosynthesis genes is that genes in the same biosynthetic pathway are not clustered in plants as in microbial systems, making identification difficult unless the entire plant genome is sequenced (Schäfer & Wink 2009). Next-generation sequencing techniques are starting to alleviate this challenge, but in the present and immediate future, lack of plant genetic knowledge and tools has been a major bottleneck in the transfer of plant pathways to microbes.

Another challenge that must be addressed on a case by case basis for microbial transfer is enzyme engineering. Adapting enzymes for optimal function in a non-native host can be difficult, particularly for plant cytochrome P450s (CYP450s). CYP450s and their associated reductase partners are heme-thiolate enzymes constituting one of the largest super families of proteins, and are found in both prokaryotes and eukaryotes. In eukaryotes, CYP450s are membrane bound, while in prokaryotes they are soluble (Werck-Reichhart & Feyereisen 2000). CYP450s are responsible for many of the unusual oxidative reactions seen in the biosynthesis of plant SMs, making them essential for heterologous expression of many plant NPs. Numerous challenges exist in the expression of plant CYP450s, including: improper folding, lack of an appropriate membrane binding site in prokaryotes, and difficulties in maintaining adequate NADPH pools for reductase activity. These difficulties can be overcome on a case by case basis via protein engineering and related strategies, and multiple functioning plant CYP450s have been successfully expressed in *E. coli* (Chemler & Koffas 2008).

Plant Cell Culture

Heterologous production of plant NPs in microbes will continue to develop for the foreseeable future, but it is important to situate these breakthroughs in the context of current pharmaceutical supply. Microbial fermentation has been touted as a cost-effective route for large-scale production of high value plant NPs, but as mentioned above, it is not possible for certain NPs, particularly for those with undefined biosynthetic routes. Plant cell culture (PCC) is an attractive option for immediate production of plant NPs when heterologous production is infeasible.

Environmentally-friendly PCC has been used industrially for plant NP supply since 1984, when Mitsui Petrochemical Industries, Ltd. began using a cell suspension of *Lithospermum erythrorhizon* to produce the pigment shikonin on a 750-L bioreactor scale (Georgiev et al. 2009). Today, PCC is utilized commercially to produce a wide variety of pharmaceuticals, pigments and food additives at scales up to 75,000-L, as in the case of Phyton Biotech, Inc.'s paclitaxel process.

Plant SMs, which form the majority of plant NPs, are often under strict and specific metabolic control; for example, some plant NPs are synthesized in response to exogenously applied elicitors such as methyl jasmonate, salicylic acid or nitric oxide, which are known to be key signaling molecules in the plant defense response (Zhao et al. 2005). Some plant NPs require specialized cell types to synthesize the desired NP; for example, the biosynthesis of vindoline, one intermediate in the vinblastine and vincristine pathway (Figure 2) appears to require three different mature cell types and multiple intracellular compartments (Kutchan 2005). Overcoming limitations imposed by metabolic control is a major area of research, with efforts placed towards activating inherent secondary metabolism via elicitation, as well as metabolic engineering and expression of transcription factors to rationally control plant metabolism.

Plant Cell Suspension Culture

Plant cell suspension culture (PCSC) is a subset of PCC in which cells from an intact plant or plant embryo are "de-differentiated" using phytohormones and grown in liquid media. Plant cells are totipotent, in that given correct signals they can differentiate and develop into any plant cell type or organ; this also gives them the ability to be maintained in a de-differentiated state. Given their constant growth, PCSCs must be subcultured into fresh media every 1-4 weeks depending on their specific growth rate. Importantly, cryopreservation techniques have been developed for a number of medicinally-relevant plant species to preserve vitality without repeated subculture (Mustafa et al. 2011). However, recovery of PCSCs from cryopreserved cultures is a timeconsuming process and does not always result in a predictable performance (Harding 2004). As a result of incomplete separation after cell division, plant cells in suspension grow as aggregates ranging in size from a few cells $(100 \mu m \text{ diameter})$ to thousands of cells $(2 \mu m \text{ diameter})$ (Kolewe et al. 2008). In addition, plant cells are more sensitive to shear than microbial cells, owing to their large, water filled vacuoles. A number of novel impeller configurations and bioreactor types have been developed to meet the unique needs of PCSCs – most notably the recent invention of disposable, gas-permeable-bag wave-type bioreactors – although many plant cell types have been grown successfully in commonly available stirred tank bioreactors at culture volumes up to 70 m³ (Eibl & Eibl 2007).

A major area of research in PCSC is the variability in metabolite accumulation of dedifferentiated plant cells that have been subcultured repeatedly over a period of years. It is known that the callus culture method by which de-differentiated PCSCs are formed generally creates a heterogeneous population of cell types in suspension (Senger et al. 2006). The de-differentiated state itself sometimes leads to genomic instability over repeated subculture (Baebler et al. 2005). Additionally, the aggregated nature of plant cells in suspension may cause epigenetic changes, affecting product accumulation through unknown mechanisms (Patil et al. 2011). The relationship between aggregate size and NP accumulation has been studied in a number of different plant

systems, but results are varied, in part due to inconsistent experimental methodologies. In a recent study using a more accurate whole-culture measure of metabolite synthesis, paclitaxel accumulation in de-differentiated PCSC increased as aggregate sized decreased, suggesting aggregation as a process parameter that should be optimized (Kolewe et al. 2011).

A recent development in PCSC is the suspension culture of cambial meristematic cells (CMCs). *In planta*, cells in the cambial meristem layer of tissue can grow indefinitely and become any cell type. Because of this feature they have been likened to plant stem cells and defined as "undifferentiated" in contrast to the "de-differentiated" cells discussed previously. Suspension cultures of *Taxus* CMCs were made by creating callus culture from isolated *Taxus* cambial meristem tissue. The CMC cultures grew faster, formed smaller aggregates and displayed greater genetic stability than de-differentiated cultures (Lee et al. 2010).

Hairy Root Culture

CMCs are an attractive new option for plant cell cultures; however, some plant NPs are produced preferentially in more differentiated cell types and may be produced in small quantities, or not at all, in truly un-differentiated cells (Roberts & Kolewe, 2010). Hairy root culture (HRC) is another subset of PCC in which root tissue from a plant of interest is genetically transformed by the soil bacterium *Agrobacterium rhizogenes*, leading to branched "hairy roots" that can grow indefinitely without exogenously supplied phytohormones needed for PCSCs. Because HRCs are differentiated tissue, they frequently have the ability to produce SMs that de-differentiated PCSCs cannot, and in some cases can even produce plant NPs that are not normally found in roots of intact plants. For instance, HRCs of *A. annua* accumulated artemisinin, whereas in whole plants artemisinin is produced only in the aerial sections (Kim et al. 2002). HRCs also appear to be more genetically stable than de-differentiated cultures and do not lose biosynthetic potential upon repeated subculture (Georgiev, Ludwig-Muller, & Bley, 2010).

Despite their proven biosynthetic capabilities, HRCs have not yet been used commercially for production of any plant NPs, mostly due to the lack of proven, large-scale bioreactors. The branched and corporeal nature of HRCs make them difficult to culture in stirred tank bioreactors, and nutrient transport limitations result in slow growth and low productivity. Novel bioreactor designs may improve HRC, and among the most promising are mist bioreactors. This novel bioreactor type sprays exposed roots with a nutrient solution and it has been shown to increase growth and metabolite productivity at high root densities (Weathers et al. 2008).

Yield Enhancement Strategies

Apart from variability in product accumulation, another major area of research in PCC is enhancement of biosynthetic capabilities. This goal is realized through a number of strategies, including chemical elicitation of secondary metabolism, selection of elite cell lines, and metabolic engineering. While plant NPs are generally produced in extremely small quantities under normal conditions, in some cases SMs may constitute 20-60% of a plant's dry weight (Verpoorte et al. 1999). Activation of secondary metabolism using elicitors such as methyl jasmonate has been shown to significantly increase production of plant NPs, almost 50-fold in the case of paclitaxel in *T. baccata* PCSC (Yukimune et al. 1996). Elicitation pathways are highly conserved in plants, and elicitors such as methyl jasmonate have been shown to activate secondary metabolism in a wide variety of plants through similar pathways (Gundlach et al. 1992; van der Fits & Memelink 2000).

A useful technique in cell culture to improve yields is to select individual cells that have high biosynthetic productivity and culture them as "elite" cell lines. This technique is generally difficult to accomplish with plant cell lines, as cultures are composed of genetically heterogeneous aggregates that may respond differently to selection pressures. Furthermore, due to the unstable nature of de-differentiated cells, elite cell lines may lose their increased capacity over time. A successful example of using selection pressure to create stable, elite cell lines is establishment of *Lavendula vera* cultures producing high amounts of the phenolic plant NP rosmarinic acid. Selection was accomplished by feeding toxic amino acid analogues in the media with the expectation that only cells expressing high amounts of an enzyme common to both amino acid and rosmarinic acid biosynthesis would be able to detoxify the amino acid analogues and remain viable (Georgiev, Pavlov, & Ilieva, 2006). This approach was effective in creating genetically stable, high-producing cell lines, but is limited to cases in which an effective screening procedure exists. Additionally, the heterogeneous and aggregated nature of PCSCs may dilute the effect of small numbers of high-producing cells because they are not separated from large numbers of low-producing cells. In cases where it is possible to disassociate individual cells from aggregates, labeling of the plant NP of interest and flow cytometric cell sorting is a more robust method for isolating and re-culturing elite cell lines (Naill & Roberts 2005b; Naill & Roberts 2005a)

Manipulation of biosynthetic pathways at a genetic level offers significant potential for increasing yields of plant NPs in PCC. Recent technological developments, such as 454 pyrosequencing, have lowered the barriers to identifying genes in biosynthetic pathways, and deliver the promise of making "non-model" organisms amenable to metabolic analysis (Bräutigam & Gowik 2010). Once biosynthetic pathway genes are known, *Agrobacterium tumefaciens* transformation can be used to stably introduce or silence genes in plants, although some types of plants such as trees and crop plants can be recalcitrant to *Agrobacterium* transformation. Engineering of *Agrobacterium* strains for increased virulence as well as manipulation of plant culture conditions have resulted in a wider applicability of this technique (Gelvin 2003). Transcription factors are also targets of metabolic engineering efforts. For example, simultaneous up-regulation of the gene *G10H* and the *ORCA*3 transcription factor in HRCs of *C. roseus* resulted in a 6.5 fold increase in catharanthine, a precursor of vinblastine (Ni et al. 2011). In many cases, however, transcription factor

engineering is not effective at increasing production of the desired NP, and new models of plant metabolism may better our understanding of how to manipulate metabolism on a systems level (Stitt et al. 2010).

One promising technology to better understand plant metabolism and ways to improve it are genome scale metabolic models (GEMs). These models take into account all known metabolic reactions within a cell and apply optimization conditions (usually maximization of growth) to simulate fluxes for the metabolic reactions. GEMs have been applied successfully in the metabolic engineering of microbes such as *E.coli* for the production of chemicals such as succinic acid (Lee et al., 2005). Using a GEM, researchers can model fluxes for all biosynthetic reactions and can identify bottlenecks in pathways or competing pathways that reduce flux to the target molecule. Genetic engineering techniques can then be used to upregulate or knockout genes of interest and improve overall flux to the target molecule. GEMs have been created for multiple plant species including the model plant species *Arabidopsis thaliana*, as well as food crops such as corn and rice (Saha et al. 2011; Poolman et al. 2013).

Creating GEMs that can be used for genetic engineering of plant secondary metabolites is more complicated than for bacterial systems. First, plant secondary metabolic pathways are incomplete and poorly studied, so it may be difficult in many cases to create an accurate metabolic model that will include secondary metabolic pathways. Second, plants are highly compartmentalized and require more complex models than prokaryotes to accurately model fluxes. Third, because plants are multi-cellular and different cell types may have different functions, optimization conditions may not be as simple as maximization of growth. For optimization of secondary metabolites in particular, some molecules are only produced in specialized cell types or under stress conditions, in which optimization of growth would not give accurate accounting of fluxes.

Because so little is known about fluxes in plant cells, especially for specialized cell types and cells undergoing stress responses, measurement of fluxes using Metabolic Flux Analysis (MFA) is a strong component of the plant metabolic engineering toolbox (O'Grady et al. 2012). MFA is used to experimentally measure certain fluxes in central metabolism by taking advantage of patterns in protein labeling using isotopically labeled carbon substrates. This technique has been used in the medicinal plant *C. roseus* to discern central carbon fluxes using NMR (Sriram et al. 2007). Limitations of MFA include expense of labeled substrates and the difficulty of interpreting results via NMR or GC/MS, especially for compartmentalized organisms and photosynthesizing organisms (Niklas et al. 2010).

The Paclitaxel Story

Paclitaxel is one of the most successful drugs in the history of chemotherapeutics, with annual reported sales in 2000 exceeding 1.5 billion USD (Expósito et al. 2009); however, the early history of paclitaxel was fraught with complications, and at many points it was almost discarded as a drug lead. The story of paclitaxel (summarized in Table 2) began in 1962, when a sample of bark from the Pacific yew, *T. brevifolia,* was collected by USDA workers. An extract tested positive for activity in the KB cytotoxicity assay. Following the positive result, Dr. Monroe Wall fractionated the sample and isolated the cytotoxic agent, paclitaxel, in 1971. Paclitaxel had favorable but unremarkable cytotoxic activity against common cell-based cytotoxicity screens; however, interest in paclitaxel was ignited in 1979, when Dr. Susan Horwitz showed that paclitaxel had a unique mechanism of action against tumor cells. Whereas previous spindle poisons (e.g., the vinca alkaloids vinblastine and vincristine) act by rapidly depolymerizing microtubules and preventing spindle formation in mitosis, paclitaxel instead stabilizes the tubulin polymers, preventing the cell from properly assembling its spindle and continuing through mitosis.

Paclitaxel was a poor drug lead due to low aqueous solubility, structural complexity precluding easy chemical synthesis, and the lack of a large and renewable supply of the compound. However, interest from the research community, due to its novel mode of action, helped to move it past major difficulties during its initial development, including several deaths in Phase I clinical trials due to allergic reactions with Cremophor, an emulsifying agent in its formulation. Although at many points in paclitaxel's history the chances of it emerging as a successful chemotherapeutic appeared slim, clinical trials continued after the formulation was re-worked, and extremely promising Phase II results were obtained against refractory ovarian cancer (Cragg, Schepartz, Suffness, & Grever, 1993).

Even as paclitaxel enjoyed success in the clinic, there was still a major hurdle to its development as a large-scale chemotherapeutic. While early research and initial clinical trials had relied on collection and extraction of bark from wild-growing *T. brevifolia*, the collection and extraction processes were environmentally damaging. *T. brevifolia*'s properties as a SM producer were a perfect storm of unfortunate specifications: the tree was slow growing, paclitaxel accumulated only in the bark, and yields were very low. In contrast to harvest of the vinca alkaloids from fastgrowing *C. roseus*, natural harvest of paclitaxel was unsustainable – 16,000 lbs of bark from approximately 2000 yew trees were required to produce 1 kg of paclitaxel (Cragg & Snader 1991). In the early 1990s the paclitaxel supply problem was finally recognized as a crisis. In 1991 the NCI entered into a Cooperative Research and Development Agreement (CRADA) with Bristol-Myers Squibb (BMS) to heavily fund research into alternate supply routes. Meanwhile, the destruction of trees became a high profile environmental issue and in 1992 federal legislation (The Pacific Yew Act) was passed to manage the survival of the *T. brevifolia*. Concurrently in 1992, paclitaxel was approved by the FDA to treat ovarian cancer and demand for the drug sharply accelerated. By 1993 NCI was supporting 35 grants on paclitaxel research, with funding of \$4.6 million, in addition to BMS's independent research (Cragg et al., 1993).

The most immediate relief for the supply issues was a semi-synthetic method that was first developed in 1986 and later re-worked to be more efficient (Denis et al. 1988). 10 deacetylbaccatin III was extracted from the needles of various *Taxus* species and converted to paclitaxel via chemical methods. Needle harvest was more environmentally friendly than bark harvest, leaving the tree viable and intact, and in 1994 the semi-synthetic method was approved by the FDA to supply paclitaxel, now trademarked Taxol®. However, the semi-synthesis method was expensive and environmentally unfriendly due to a number of harsh chemical solvents used. Research into production alternatives continued to be a hot topic during the 1990s, and in 1994 a total synthesis of paclitaxel was first reported (Nicolau et al. 1994; Holton et al. 1994). Paclitaxel is a complex molecule containing over 11 chiral centers and a unique oxetane ring chemistry, and total organic syntheses have still not resulted in any cost-effective or industrially-viable processes.

Plant cell culture (PCC) was first funded by the NCI as a supply route for plant-derived anticancer agents as early as 1977, but the contracts were terminated prematurely in 1980 due to lack of interest at the time (Cragg et al., 1993). A number of PCC projects were funded by the NCI in response to the paclitaxel crisis, and the first patent for plant cell suspension culture of *T. brevifolia* for paclitaxel production was issued in 1991, with reported yields of 1-3 mg/L (Christen et al. 1991). Phyton Biotech, Inc. was formed in 1990 near Cornell University (Ithaca, NY) in response to a renewed interest in PCC technologies to supply paclitaxel. The company grew quickly and in 1993 acquired Phyton GmbH and a 75,000 L cGMP PCC facility in Heidelberg, DE. Phyton Biotech licensed their PCC process to BMS in 1995, and continued to improve the paclitaxel process, recently filing a patent for strategies to increase broth titers to 900 mg/L. (Bringi et al. 2007). The current worldwide supply of paclitaxel is provided by Phyton Biotech, which is now a subsidiary of DFB Pharmaceuticals, Inc. Recent research in *Taxus* PCC

focuses on improving paclitaxel yield and minimizing production variability by better understanding paclitaxel biosynthesis and regulation on a genetic level, as well as the influence of key process variables such as aggregation (Wilson & Roberts, 2011).

Research Objectives

The work in this thesis is an exploration of the use of genome scale metabolic models in understanding plant secondary metabolism, with an aim towards future genetic manipulations of secondary metabolite producing plants to improve yields. In Chapter 2, a general overview of genome scale models is given, and the process of creating a metabolic model from a transcriptome for a plant species of interest (*T. cuspidata*) is outlined. In Chapter 3, a published genome scale metabolic model of *A. thaliana* is used in conjunction with transcriptome data under varying conditions to model changes in metabolism. Different methods for integrating transcriptome data with metabolic models are explored and discussed.

CHAPTER 2

METABOLIC MODEL OF *TAXUS CUSPIDATA*

Background

Flux Balance Analysis

Genome-scale metabolic models (GEMs), also referred to as metabolic reconstructions, represent the current start-of-the-art for capturing whole cell metabolism by quantifying steady-state fluxes through a network of enzymatic reactions. This technique has been used to model systems with a range of complexity from *Escherichia coli* (Edwards & Palsson 2000) to *Homo sapiens* (Bordbar & Palsson 2012). The primary element of a GEM is the reaction network described by a stoichiometric matrix relating the individual reactions and their participating metabolites. The term "genome-scale" describes the process by which the stoichiometric matrix is obtained, usually from an annotated genome of the organism (Borodina & Nielsen 2005). The stoichiometric matrix is invariably underdetermined with more unknown reaction fluxes than metabolite balances, meaning that the solution space is infinite. To determine unique solutions, the stoichiometric matrix is combined with a suitably chosen objective function that represents the "cellular objective" and presumably captures various unmodeled regulatory processes. The resulting optimization problem is a classic linear program, and the process of solving the model is termed Flux Balance Analysis (FBA).The default objective function is maximization of the growth rate, which requires that the biomass composition be specified in terms of the reaction fluxes to the biomass precursors (Feist & Palsson 2010). This objective function has been successfully applied to numerous microbial models (Edwards & Palsson 2000; Duarte et al. 2004).

To date, curated GEMs exist for over 80 organisms and draft models exist for thousands of other organisms. Once available, a GEM can be used to analyze and engineer metabolism through a set of computational methods collectively termed "constraint-based modeling"(Orth et al. 2010) and available through software platforms such as COBRA (Becker et al. 2007).

Creation of Metabolic Networks

Before FBA can be performed, a metabolic network must be constructed to provide the stochiometric matrix. In practice, many of these metabolic networks have been constructed from existing databases of curated biochemical and genomic data, such as AraCyc for Arabidiopsis (Poolman et al. 2009) and RiceCyc for rice (Poolman et al. 2013). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database(Kanehisa & Goto 2000) is also a useful resource in model building, as it provides information on pathway directionality and GPR correlations for many organisms that may not possess curated databases. The ability to create a GEM for any organism relies on the ability to accurately recognize enzymes encoded in the genome or transcriptome, annotate those enzymes with correct metabolic functionality, and build a complete enzymatic reaction, including cofactors.

Application of genome-scale modeling to plant species has proven to be difficult for a number of reasons, most notably the lack of genomic information and the complexity of plant metabolic pathways. Despite these challenges, GEMs have recently been developed for several plant species including *A. thaliana* (Poolman et al. 2009; de Oliveira Dal'Molin et al. 2010; Mintz-Oron et al. 2011) and important food and biofuel crops such as maize (Saha et al. 2011), rice (Poolman et al. 2013), and the bioenergy crops sugarcane and sorghum (Dal'Molin et al. 2010). Although plants are complex multi-cellular systems that have disparate tissues and organ functions, these models typically apply the maximum growth objective for resolving unknown reaction fluxes. Existing plant GEMs have been used to answer basic questions about plant biology or to facilitate

improvement of crop species, and as yet there have been no models created for medicinal plants such as *T. cuspidata*.

Results

Illumina Data Translation

In previous work with collaborating groups, a comprehensive transcriptome for *T. cuspidata* cultured cells using the 454 and Illumina sequencing technologies was developed. For 454 sequencing, cells were MJ elicited or unelicited and harvested every 24 hours post-elicitation over a time period spanning 22 days of the culture period. Equal amounts of RNA from each culture were then pooled and submitted as a single sample for sequencing. A base transcriptome library was generated by sequencing the pooled and fragmented RNA sample on one full PicoTiterPlate (PTP) using the 454 Genome Sequencer FLX Titanium System™(Roche, Branford, CT) (Patil 2013).

In addition, MJ elicited and unelicited *T. cuspidata* RNA isolated from cultures at time points of 18 and 72 hours were created for Illumina sequencing. The Illumina HiSeq 2000 platform (Illumina, Inc. San Diego, CA) was used for these samples. Contigs were generated using both the 454 and Illumina sequencing libraries. CLC genomics workbench (CLC Bio, Aarhus, Denmark) was used to generate contigs using the A, C, G, T voting method to resolve conflicts (Patil 2013). 48,614 contigs with >200 bp and >50x coverage were generated. Annotation was performed using Blast2GO default parameters (Conesa et al. 2005).

This transcriptome represents the full biosynthetic capabilities of *T. cuspidata* in both the MJ elicited and unelicited state, and was used to develop the metabolic model. A Matlab script was developed to translate the transcriptome. This script had four steps:

1. The sequence of the contig was sent to BLAST and a BLAST report was received. The reading frame from the top BLAST hit was used in subsequent steps.

- 2. The transcriptome derived DNA contig was translated into amino acid sequence using functions from the Matlab bioinformatics toolbox and the reading frame from the top BLAST hit.
- 3. Because the contig sequence often had beginning or ending 'nonsense' sequences, the correct portion of the amino acid sequence was determined via homology with the top hit of the BLAST report.
- 4. The annotations associated with each contig from Blast2GO as previously described were added to the contig in FASTA format.

This script was applied to the 29,000 contigs from the *Taxus* sequencing project that had significant homology with previously annotated proteins and thus had BLAST reports.

Draft Model

A draft GEM of *T. cuspidata* primary metabolism was created by the ModelSEED using transcriptomics data from 454 sequencing, as described above, and recently developed plant modeling tools available in the ModelSEED metabolic reconstruction pipeline (Henry et al. 2010). A key novelty of the *Taxus* model is that its future use was intended to be a study of secondary metabolism. Most GEMs (especially in higher organisms such as plants) do not include secondary metabolism, largely due to lack of genomic information. For example, in the diterpenoid paclitaxel biosynthetic pathway of *Taxus*, multiple enzymes catalyzing reactions in the latter half of the 19 step pathway remain unknown (Hampel et al. 2009). The intended focus of this research was on increasing the synthesis of precursors to general terpenoid metabolism that are included in the model, such as isopentenyl diphosphate (IPP), its allylic isomer dimeythlallyl diphosphate (DMAPP), and the first committed metabolite in the diterpenoid (paclitaxel) pathway, geranylgeranyl diphosphate (GGPP) (Gräwert et al. 2011; Roberts 2007). By determining the regulatory points in primary metabolism to increase precursor concentration and combining with targeted gene manipulations within secondary metabolism (e.g., regulatory steps in the paclitaxel biosynthetic pathway (Nims et al. 2006), increases in yields to paclitaxel may be achievable.

The data used for draft metabolic model reconstruction was a combination of transcriptomes from MJ elicited and unelicited cultures. As described above, of the roughly 40,000 contigs generated from sequencing, 29,000 were found to have

statistically significant BLAST hits, allowing preliminary annotation. These annotated sequences were translated into amino acid sequences using the open reading frame (ORF) associated with top BLAST hits. The resulting annotated protein fragments were used by the ModelSEED to automatically generate Gene-Protein-Reaction (GPR) associations for all contigs corresponding to enzymes recognized in the ModelSEED database. A total of 924 reactions were generated directly from the annotated contigs by the ModelSEED. Many of these reactions referenced multiple contigs, which may be a result of multiple isoforms or subunits of a particular enzyme.

The ModelSEED software was used to gap-fill the draft model by adding reactions necessary to achieve *in silico* growth. Prior to gap-filling, growth was not possible due to dead-end reactions in the metabolic network and the inability to synthesize all biomass components. The gap-filling procedure required specification of the biomass equation and definition of the culture media. The biomass equation used in our draft model was that previously developed for an *A. thaliana* GEM (Poolman et al. 2009). A standard heterotrophic plant cell culture media [\(Table 1\)](#page-35-0) was included in the model by adding bounded exchange reactions for available media components.

Phosphate	-100
NH ₃	-100
Sulfate	-100
H+	-100
Nitrate	-100
Sucrose	-100
Mg	-100

Table 1 – Media components and bounds in *T. cuspidata*

Using a mixed-integer linear programming methodology to minimize the number of gap-filled reactions, the ModelSEED software predicted that growth could be achieved by adding 37 noncontig-associated reactions to the model [\(Table 2\)](#page-35-1). The gap-filled draft model contained 1001 reactions and 948 metabolites. By performing FBA with a maximum growth objective, successful prediction of the biomass growth rate as a function of the sucrose uptake rate was achieved using ModelSEED software [\(Figure 3\)](#page-36-0). The model data were provided by Argonne National Laboratory using their software.

Total Reactions	1001
Contig-associated	924
Gapfilled	37
Transport/exchange	40
Metabolites	948

Table 2. Draft GEM model statistics.

Figure 3 - Relationship between biomass growth rate (g dry weight/L/Day) and sucrose uptake rate (g/L/Day) from experiment (points) and draft model

Unfortunately, at the time of this research, the software that ModelSEED made publicly available was not useable. At that time, ModelSEED distributed beta versions of their Linux-based software that were accessible on Windows systems and internet interfaces such as Kbase. The software could not predict growth on "full" media, which was defined to contain all possible building blocks a cell would need. This "full" media contained more chemical components than the media used to culture *T. cuspidata* experimentally and should have been more than sufficient to enable model growth. The lack of growth using the new software distributions from ModelSEED suggested problems with the software and not with the model.

It was attempted to convert the model, which was stored in SBML format, to COBRA format and run it in Matlab using the well-known COBRA toolbox (Becker et al. 2007). COBRA has built-in functions to facilitate translation of models stored in SBML, 'translate sbml.m' and 'convertsbmltocobra.m'. These functions in the past had required adjustment to avoid errors, e.g., removing code from translate_sbml.m that tried to access features in a model that were not present, such as notes. It is important to remember that the softwares involved in FBA are

generally written by overworked biology graduate students, not computer scientists, and as such they are full of bugs and not robust. The COBRA 'translate_sbml.m' function worked on the ModelSEED SBML without issue to create a Matlab structure for the model, and COBRA's function 'convertSBMLToCobra.m' also worked without issue to create a COBRA model in Matlab. However, when it was attempted to run FBA on the model in COBRA using the 'optimizeCbModel.m' function, the model would not produce growth. It was found that other models downloaded in SBML format from the ModelSEED repository and converted to COBRA format would similarly not produce biomass. For example, the *E. coli* model iJR904 (Reed et al. 2003) would successfully produce biomass when downloaded from the website of the Palsson group in SBML format, converted to COBRA and ran with 'optimizeCbModel.m', but when iJR904 was downloaded in SBML format from the ModelSEED website, it would experience the same problem as the *Taxus* model, in that it would successfully convert to a COBRA model, but would not produce biomass when 'optimizeCbModel.m' was run. The SBML codes of the original iJR904 model and the ModelSEED version were compared, but the problem with the ModelSEED version could not be diagnosed. The fact that the iJR904 model from the ModelSEED repository experienced the same problem as the *Taxus* model suggests that the root cause may be an inability of ModelSEED models to be accurately converted to COBRA format.

Discussion

In this study, an automated software was used in conjunction with transcriptome data to create a metabolic model of a plant. To our knowledge, this was the first time that a metabolic model was created using a transcriptome instead of a genome, and this was also the first genome-scale model of a medicinal plant, as opposed to a model plant such as *A. thaliana* or a crop species such as *Zea mays*. The use of a transcriptome from next-generation sequencing significantly reduced the barrier to entry for genome scale metabolic models of non-model species. However, despite automation, genome-scale metabolic models still require significant manual refinement, and the

lack of actual experimental knowledge of *T. cuspidata* hampered efforts to make the model more accurate. Additionally, because much of the software is built by labs for their own use, it is still quite difficult to use existing software packages such as those put out by the ModelSEED. The COBRA toolbox represents an attempt to standardize FBA and make it more widely available, but it is probably that available software will still be sub-par until commercial packages become available. As it stands, the *Taxus* metabolic model requires significant improvement to become useful.

CHAPTER 3

GENOME SCALE METABOLIC MODELING OF ARABIDOPSIS THALIANA UNDER METHYL JASMONATE ELICITED CONDITIONS

Introduction

Many important plant-derived compounds fall under the category of secondary metabolites in that they are not essential for the normal functioning of the native plant. Secondary metabolites are largely thought to participate in plant defense, protecting plants from external threats such as fungal diseases and insect herbivory (Bourgaud 2001). Secondary metabolism is activated by key compounds produced endogenously by the plant, termed elicitors. The study of elicitation and the plant defense response is currently an extremely active area of research in plant biology (De Geyter et al. 2012). Of the known compounds involved in elicitation, jasmonic acid and structurally similar compounds, collectively referred to as jasmonates, are major players in the most dynamic defense response pathway. Methyl jasmonate (MJ) is a stable esterified jasmonate that is present in nature and often applied exogenously *in vitro* to elicit the defense response in plants. MJ itself does not induce the defense response, but must be conjugated to isoleucine *in planta* to bind to transcription factors, such as JAZ (Fonseca et al. 2009). Elicited plants or cell cultures exhibit changes in diverse metabolic pathways, including decreases in ATP production related to mitochondrial membrane perturbation (Ruiz-May et al. 2011), decreased participation in cell cycle (Swiatek et al. 2002), and increased production of secondary metabolites (Gundlach et al. 1992; Mueller et al. 1993). An elicited plant can be thought of as having an altered phenotype while it is exhibiting the defense response.

Because jasmonates significantly induce secondary metabolite formation, they are a typical component of bioprocesses involving cultured plant cells. Paclitaxel accumulation in particular is greatly enhanced upon MJ elicitation (Yukimune et al. 1996). Previous studies have shown that MJ treatment results in upregulation of paclitaxel biosynthesis genes in *Taxus cuspidata* cultured cells (Nims et al. 2006). Other studies in model systems such as *Arabidopsis thaliana* have focused on transcriptional regulation of the jasmonate response at a systems level (Pauwels et al. 2008), but there have been no systematic studies on how metabolic flux is affected by MJ elicitation.

Genome-scale metabolic models (GEMs), also referred to as metabolic reconstructions, are tools used to quantify steady-state fluxes through a network of enzymatic reactions for a particular organism. Because they can model systemic changes in an organism's metabolism, they are an ideal tool to study MJ elicitation. In general, GEMs have an inherent advantage over timedependent models (kinetic models) in that knowledge of rate parameters for each reaction is not necessary – only the stoichiometry of each reaction must be obtained. After a GEM is constructed, it can be used to analyze and engineer metabolism through a set of computational methods collectively termed "constraint-based modeling" (Orth et al. 2010). Flux Balance Analysis (FBA) is currently the most popular of these GEM methods. Given a metabolic network and a "cellular objective function," FBA methods tune flux through each reaction until optimal flux through the objective function is achieved. Common objective functions for FBA have included maximization of cellular growth, minimization of ATP usage, production of secondary metabolites, or combinations of the afore-mentioned functions. An active area of research in the field is the integration of 'omics data (transcriptomics, metabolomics and proteomics) into FBA methods to increase accuracy and to model complex cellular programs where the cell's regulatory logic may be unknown. This application, referred to here as "phenotype simulation" was of most interest to us in modeling the MJ elicited state in the model plant species *A. thaliana*. *A. thaliana*

was chosen for our study because there exist multiple, continually improved GEMs of *A. thaliana* metabolism (Poolman et al. 2009; de Oliveira Dal'Molin et al. 2010; Saha et al. 2011; Mintz-Oron et al. 2011). Despite progress and improvement on *A. thaliana* GEMs, there problems still exist with plant GEMs that do not necessarily apply to microbial systems. Plants are complex and multi-cellular systems that may have disparate tissue and organ functions. The most common objective function (unlimited growth) may not apply to plant tissues under many circumstances, especially under circumstances like MJ elicitation where experimental evidence shows that growth is inhibited.

Phenotype simulation is useful in this circumstance because transcriptomic data from MJ elicited *A. thaliana* cultures (Pauwels et al. 2008) can be used to augment FBA predictions and model MJ elicitation phenomena such as growth inhibition and secondary metabolite production if these phenomena are reflected in gene expression changes. The phenotype simulation algorithms that were chosen in this case to model MJ elicitation were GIMME and E-Flux. GIMME can be thought of as the simplest method of phenotype simulation. A cellular objective is set, and reactions are assigned Boolean ON/OFF values based on the expression level of the gene corresponding to that reaction compared to a threshold expression value (Becker $\&$ Palsson 2008a). FBA is run on the augmented matrix, and if a reaction necessary for the objective function has been turned off, GIMME may turn it back on. E-Flux operates differently in that instead of turning reactions ON or OFF, it changes the upper and lower bounds of each reaction in accordance with its expression value (Colijn et al. 2009). Highly expressed reactions will have larger upper and lower bounds, allowing more flux through if the objective function demands it. Lowly expressed reactions will not be able to produce large amounts of flux, even if it would assist in the optimization of the objective function.

In applying GIMME and E-flux to MJ elicitation modeling, our goal was that these phenotype simulation methods would capture the growth inhibition and secondary metabolite production

that are experimentally observed in *A. thaliana* (Pauwels et al. 2009). If growth-associated genes had low expression, they would be turned OFF in GIMME or have small bounds in E-flux, and the modeled growth activity would be lessened. If genes related to secondary metabolism were increased, flux would be shifted to these pathways, resulting in both secondary metabolism flux and growth inhibition.

Materials and Methods

Model

The previously published *A. thaliana i*RS1597 model containing 1234 genes, 1591 reactions, and 1916 metabolites was used in this analysis (Saha et al. 2011). The model includes reactions for secondary metabolism. The model was obtained in SBML format and converted subsequently to COBRA and TIGER compliant models using those respective software toolboxes (Becker et al. 2007; Jensen et al. 2011).

Data

Previously published data on MJ elicitation were used in this analysis (Pauwels et al. 2008). Suspension cultures of *A. thaliana* (L.) Heynh. (ecotype Columbia-0) were grown for 7 days, diluted 10-fold in fresh medium and at 16 hours were either elicited with 50 μM MJ or mock-elicited with equal volume of DMSO. Data were collected at 0.5, 2, and 6 hours post-treatment. Data used in this analysis were from MJ+ and MJ- cultures 6 h after treatment. Normalized data were obtained from the ArrayExpress database (www.ebi.ac.uk/arrayexpress; accession no. E-ATMX-13). For each condition, two independent experiments were used to generate transcripts for microarray profiling, and one biological repeat underwent RT-qPCR with multiple reference gene normalization on 4 reference genes to validate the data. The reference genes At1g69280, At4g17300,

At3g25800, and At1g04300 were of comparable intensity to the genes of interest and exhibited minimal variability in the RT-qPCR analysis. The data were background corrected and normalized using the BioConductory Affy Package in the following manner: "The Affy mas5calls function assigned probe sets present, absent, or marginally detectable. Probe sets with fewer than two present calls during the time course were considered not reliably detected and were removed from further analysis. With the BioConductor Limma package, probe sets were selected with a significantly altered intensity value at one or more time points after MeJA elicitation. The MeJA-treated and mock-treated cultures were compared pairwise at the three time points (30 min, 2 h, and 6 h). The decideTests function assigned significance to ttests while correcting for multiple testing of probe sets and contrasts with the false discovery rate method. The corrected P value threshold was set to 0.05." (Pauwels et al., 2008, SI). Only processed expression data corresponding to genes included in the *A. thaliana* model were considered in our analysis. In the study, 598 of the 22,746 probe sets were found to be differentially regulated by MJ. Of these genes, 75 were differentially regulated in the first 0.5 hour, rising to 495 at the 6 hour time point. Of the 328 genes expressed only at the 6 hour time point, 60% were down regulated in the MJ+ condition. Genes were clustered according to up or down regulation at different time points and further clustered according to function or class using MAPMAN software (Nagel et al. 2005). The cluster analysis revealed that genes associated with the phenylpropanoid pathway were up regulated by MJ addition, while genes associated with mitosis and cell cycle progression were down regulated. Following the cluster analysis, additional experiments were performed to explore the findings. One result of interest from this paper is that MJ

31

addition to *A. thaliana* cell cultures increased flux through the monolignol pathway (related to the phenylpropanoid pathway), which was confirmed via HPLC detection of oligolignols, including lignin. Another result of interest was that MJ inhibited the growth cycle and stopped cell cycle progression in the G_2 phase, which was confirmed via flow cytometry (Pauwels et al. 2008).

GIMME and Essential Genes

The TIGER toolbox for Matlab (Jensen et al. 2011) was used to implement GIMME (Becker & Palsson 2008a). TIGER's implementation of GIMME requires a metabolic model containing an objective function, a set of expression values corresponding to the model's Gene-Protein-Reaction (GPR) matrix, and a threshold value for expression. Because a biologically relevant "threshold value" for GIMME could not be determined, this algorithm was used instead to determine "essential genes" in the MJ elicited and unelicited states. This was accomplished by setting a threshold higher than any one expression value, thus setting all reactions OFF, and then determining which genes GIMME deemed necessary to turn back on. All genes turned ON by GIMME were deemed "essential." This process was dependent on the particular profile of an expression dataset, leading to a GIMME-determined "essential gene set" for each condition tested. To test the bounds of which genes could be "essential" or "non-essential" 100 sets of gene expression data were created containing randomized values for each gene. Using the Gurobi5 solver and TIGER toolbox, GIMME was applied to each set, and the numbers of "essential" and "non-essential" genes were collected for each case, as well as the number of reactions carrying flux in each case.

FBA and Phenotype Simulations

To perform FBA, we utilized the *GIMME.m* command in the TIGER toolbox for GIMME, and the *optimizeCbModel.m* command in the COBRA toolbox for E-Flux and all other instances of FBA. The Gurobi5 solver was used in all cases. Elicited (MJ+) and unelicited (MJ-) data were integrated with the *A. thaliana* GEM using the GIMME and E-Flux algorithms. In the case of GIMME, phenotype simulation was carried out using the method described above, in which $MI+$ and MJ- expression sets were given thresholds exceeding values in either data-set, and "essential" genes were collected for both data-sets. The E-Flux algorithm (Colijn et al. 2009) was implemented using Matlab scripts and Excel worksheets that set bounds for each reaction based on the expression value of the gene corresponding to that reaction. The highest bounds were 1000 mmol/gDW/hr and assigned to the highest expression value, while lower expression values were given bounds corresponding to their ratio with the highest expression value. Bounds for MJ+ and MJ- E-Flux simulations can be found in Appendix A. When multiple genes were acting on a reaction, the appropriate expression value was determined following the Boolean logic rules. In the case of AND logic (more than one gene is needed for the reaction), the minimum expression value was used, and in the case of OR logic (more than one gene is available for the reaction, but only one is needed), the maximum expression value was used.

Results

GIMME

Using the GIMME algorithm, the capabilities of the *A. thaliana* GEM were explored. For 100 randomly generated expression data-sets, the number of essential genes and reactions carrying flux were determined in a GIMME simulation with a growth objective function. We found that in all cases, a core set of 116 genes were activated, while an additional set of 121 genes were turned on in at least one condition. The average number of genes activated was 153, and the average number of reactions carrying flux was 543. There was not a significant relationship between number of genes activated and number of reactions carrying flux [\(Figure 4\)](#page-46-0). This is because there are promiscuous relationships among both genes and reactions, i.e., one reaction may be

influenced by multiple genes, and one gene may influence multiple reactions. These results served to establish the bounds of the *A. thaliana* modeling system.

Figure 4 Analysis of *A. thaliana* metabolic network requirements. (a) Histogram of essential genes required to produce biomass using 100 randomly generated expression sets. The red asterisk represents the number of essential genes in MJ+ and MJ- datasets with same analysis. (b) Histogram of total number of reactions carrying flux in aforementioned simulations.

To model MJ- phenomena using GIMME, we initially performed simulations with our control dataset, MJ-, to determine what a biologically relevant threshold would be. We thought that there might be a threshold at which many genes not essential for growth would be turned OFF. However, we found that over 15% of genes in both datasets had expression levels below the means of the two expression datasets, which were 6.40 for MJ- and 6.39 for MJ+. Averages were

calculated using the entire microarray dataset, not just genes included in the metabolic model. One of the lowly expressed essential genes, AT1G32780, was in the bottom 10% of the distribution, precluding a meaningful threshold value. A full list of MJ- and MJ+ expression values for each gene can be found in Appendix A. A list of essential genes below mean expression values can be found in Table 2. These lowly expressed genes may represent problems with the model, the dataset, or our algorithmic approach. This initial modeling with our control dataset suggested that a biologically relevant threshold for GIMME could not be determined for our dataset and model.

AT5G52100	4.97	5.14	chloroplast NAD(P)H dehydrogenase	$1.6.99 -$
AT5G42250	6.09	6.46	S-(hydroxymethyl)glutathione	1.1.1.284 1.1.1.
			dehydrogenase / alcohol dehydrogenase	
AT1G67550	6.17	6.39	urease	3.5.1.5
AT4G29890	6.29	6.25	choline monooxygenase	1.14.15.7
AT1G17050	6.29	5.96	all-trans-nonaprenyl-diphosphate	2.5.1.84
			synthase	

Table 3. Lowly expressed genes in MJ+ and MJ- conditions.

Because our random phenotypic modeling had shown that determination of "essential" genes proceeded in a path dependent manner dependent on the expression data-set, we postulated that there might be differences in the number or type of essential genes in the MJ- and MJ+ dataset. We performed GIMME simulations using MJ- and MJ+ datasets, with a threshold set to initially turn OFF all genes. After essential genes were determined, we binned essential and non-essential genes according to their expression level in each dataset [\(Figure 5\)](#page-49-0).

Figure 5. Essential vs. non-essential genes. Blue bars represent MJ- (unelicited control), red bars represent MJ+ (elicited). Expression levels of essential genes have higher expression levels (>10), while non-essential genes have lower expression levels (<5).

It was hypothesized that because higher growth is observed in MJ- cultures experimentally, essential genes in MJ- cultures would have higher expression values and be represented in those bins [\(Figure 5a](#page-49-0)). However, we did not observe that trend, in fact the MJ+ expression set had generally higher expression among essential genes. It was also hypothesized that because MJ+

cultures are experimentally observed to produce secondary metabolites that are not connected to growth, that there would be a higher proportion of highly expressed genes (expression values above 10) in MJ+ cultures, and this was in fact the case [\(Figure 5b](#page-49-0)).

E-Flux

GIMME operates with a simplistic methodology that genes are OFF or ON, but in reality low gene expression does not necessarily correlate to an absence of a gene, just low amounts. In our next round of modeling, E-Flux was used to better simulate more fine-grained behavior. E-Flux operates by widening or narrowing the bounds a particular reaction can carry flux through depending on the expression of the gene controlling that reaction. In this way, a more nuanced result pertaining to cellular growth becomes apparent. In the determination of essential fluxes using GIMME, reactions carrying small amounts of flux to one essential precursor are given the same amount of weight as central reactions that carry large amounts of flux to multiple essential precursors, even though the latter would necessitate much more expression than the former during growth conditions. It was postulated that the growth inhibition observed in MJ+ cultures could be due to inhibition of a few key enzymes especially essential for growth. E-Flux can help simulate whether some genes are important in carrying large amounts of flux.

Using this algorithm, it is possible that if a few genes were especially important bottlenecks to growth, and more highly expressed in the MJ- dataset, that growth flux would be higher in the MJ- set compared to the MJ+ set. FBA was performed using a growth objective function, and resulting fluxes from MJ- and MJ+ datasets were compared to fluxes from both normal FBA and to each other. In the reaction pertaining to biomass formation (growth flux), there was very little difference between MJ- and MJ+ conditions [\(Table 4\)](#page-51-0).

Table 4 Growth in E-Flux simulations vs. FBA. Growth was calculated as the flux through the "biomass" reaction

In terms of overall flux distributions, the MJ+ and MJ- flux distributions are well correlated (R^2 = 0.9629) and are in fact nearly identical. These results are contrary to experimental data showing that the MJ- condition grew faster than the MJ+ condition. When the MJ+ flux distribution was plotted against the MJ- flux distribution, the trendline had a slope of 1.01, showing that the flux distributions are well correlated with each other [\(Figure 6\)](#page-52-0). In contrast, when these same MJ- and MJ+ flux distributions were individually plotted against the flux distribution from an FBA simulation, they were poorly correlated, with trendlines had slopes of 0.45 and 0.46 for MJ- and MJ+ respectively, and R^2 values of 0.412 and 0.403, respectively [\(Figure 6\)](#page-52-0).

Figure 6. E-flux comparison of fluxes from MJ- to MJ+ (a) MJ- (x-axis) E-flux derived fluxes plotted against MJ+ (y-axis) E-flux derived fluxes. The sets of fluxes are well correlated (R^2 = 0.962), although experimentally these conditions exhibited divergent behavior. (b) MJ- (blue **cross**) and MJ+ (red **x**) e-flux derived fluxes plotted against FBA derived fluxes (x-axis). MJ- and MJ+ are not well correlated with FBA fluxes $(R^2 = 0.412$ and $R^2 = 0.4028$ respectively) but the Eflux derived datasets are both extremely similar to each other. Both trendlines are right on top of each other.

Because the E-flux simulations did not yield divergent results as hypothesized, we further considered the expression values in MJ- vs. MJ+ datasets. The "controlling" relative expression value for each reaction in the MJ- and MJ+ conditions is plotted against the FBA-derived fluxes in [Figure 7.](#page-53-0) As can be seen, the differences in MJ- vs. MJ+ datasets are very small in every reaction that carries flux in FBA simulation. In most cases the MJ+ values are equal to or higher than the MJ- values, suggesting that MJ- expression levels are not higher for growth-associated genes although MJ- cultures exhibited higher growth experimentally.

Figure 7. Relative gene expression values of MJ- (blue cross) and MJ+ (red x) plotted against FBA derived fluxes. Many reactions with high expression are not included in the model, and it can be seen that of the highly expressed genes that exhibit high positive or negative flux in the model, there are not significant differences in MJ- vs. MJ+ datasets.

To explore whether these results are statistically significant, we cross-referenced the statistically significantly up or down regulated genes in (Pauwels et al. 2008). Of the 22,000 genes included

on the chip, 581 were up or down regulated with a p-value below 0.05. Of these 581 genes, 51 were included in our model. Of these 51, 15 participated in 17 reactions that carried flux above or below zero in the FBA simulation [\(Table 5\)](#page-56-0).

Table 5 – List of 15 genes that were found to be significantly responsive to MJ in (Pauwels et al. 2008) and also carried non-zero fluxes in 17 reactions in the FBA model of *A. thaliana* metabolism. Many of the genes (e.g. Sinapyl alcohol NADP oxidoreductase, p coumaroyl CoA caffeoyl CoA 3 hydroxylase) are associated with the lignin pathway (Humphreys & Chapple 2002). All of the genes were positively up regulated at the 6 hour time point.

Discussion

Gene Essentiality

Experimental studies of gene essentiality are often used to validate *in silico* metabolic networks of microbial metabolism. However, there are difficulties common to all metabolic networks in determining which genes are essential or non-essential. In a study comparing *in silico* determined essential genes to experimentally determined essential genes in multiple microbes, it was found that a few factors contributed strongly to false negative predictions of essential genes for growth, i.e., a gene is predicted to be non-essential, but it really is essential (Becker & Palsson 2008b).

Incorrectly predicted essential genes tend to be less well connected in the model, implying that there is incomplete knowledge of their multiple functions. They are more likely to be blocked reactions, suggesting that there is incomplete knowledge of metabolically proximal genes. They are also less likely to be connected to important overcoupled metabolites, such as ATP. This study suggests that these incorrectly predicted genes may be involved in metabolic processes that have not yet been completely characterized or defined (Becker & Palsson 2008b).

It is very easy to create large populations of mutant cell lines in an organism such as *E.coli* and experimentally determine viability under many different conditions. It is much harder to do so in a polyploidal, multi-cellular organism that undergoes multiple growth stages, such as *A. thaliana* and all plants in general. The first difficulty is that plants are notoriously difficult to genetically transform, even in a model species such as *A. thaliana*, due to time consuming regeneration procedures. Stable genetic transformations of plant species only became widely used in the late 1990's, in contrast to bacterial transformations that have been common since the 1970's (Chang et al. 1994)*.* Another key difficulty in plant species is their complicated growth cycle and multicellularity. A mutation that would be classified experimentally as "essential" could just affect the gametophyte or embryo and not be truly metabolically lethal (Lloyd & Meinke 2012). Most plant genetics studies occur in many different types of systems within Arabidopsis, for example, genetic studies can be performed on full grown plants, seedlings and embryos. Very few studies of essential genes occur in plant cell culture, although they are perhaps the most straightforward system where distinctions could be made in purely metabolic gene essentiality, rather than embryo, gametophyte, or seedling specific mutations. This makes it difficult to establish whether a mutation is actually metabolically lethal or just lethal to embryos or seedlings. Another complication is the general redundancy of plant genetics. There are duplicate or triplicate genes for the same genetic process that may be expressed in different growth stages or tissues, or all in the same cell (Hanada et al. 2009).

In this work we have determined a core set of 116 genes (Appendix A) that are predicted to be essential under all conditions. The natural continuation of this work would be to knock-out these genes in *A. thaliana* suspension cell culture and see whether these mutations result in lethality. This would suggest that these genes are metabolically essential, because in cell culture there is no ambiguity between growth phases and cell type – there is only one type of cell, and only undifferentiated growth.

Phenotype Simulation (GIMME and E-Flux)

Our initial results with GIMME suggested that binary thresholding using expression data was not sufficient to model growth in *A. thaliana*, even in the absence of complicated behaviors due to elicitation. The presence of even a few essential genes with low expression did not allow growth. This behavior, though unfortunate for the purpose of modeling MJ elicitation, is good for improving our understanding of plant metabolism. If the paradigm of transcriptome/GEM integration is that high expression leads to high flux of controlled reactions, then there is something wrong with our model or something wrong with our paradigm.

Model inaccuracies could include incorrect annotation of genes or incorrect annotation of reactions. If the genes predicted to perform an essential function in reality perform another, nonessential function, that would explain why they have low expression in the system. If a reaction is mapped to a certain gene with low expression, but one gene with high expression is key to its function, and is left out, then the reaction will show that it is only controlled by the gene with low expression, and model dysfunction will result. Our list of lowly expressed but essential genes [\(Table 3\)](#page-48-0) found by applying GIMME to the MJ- (control) dataset may point in the direction of model improvement by pinpointing reactions that have low expression but are nonetheless essential to the function of the current model.

Recent research suggests that inaccuracies in the transcriptome/GEM integration paradigm may also be likely. Studies in mouse fibroblasts showed that the amount of accumulated protein was more dependent on the rate of translation than the amount of transcripts (Schwanhäusser et al. 2011). The relationship of transcript abundance vs. translation of protein can also be very slow in plants (up to several days to replace significant amounts of proteins), so that the steady state assumption of FBA may be violated in many systems (Piques et al. 2009). Degradation rates are another potent variable that is entirely unmodeled by FBA mixed with transcriptomic data integration. In plant suspension cultures, rates of protein degradation may vary by as much as 60 fold, depending on variables such as protein function and location (Li et al. 2012). The emerging paradigm in plants is that signaling pathways pertaining to features such as circadian rhythms and light abundance have much more rapid turnover than general proteins pertaining to metabolism, which helps to reduce expenditure of proteins on rapidly fluctuating/cycling environments (Stitt 2013).

In the specific case of methyl jasmonate elicitation, our comparison of MJ- and MJ+ expression values against FBA fluxes [\(Figure 6\)](#page-52-0) suggests that our approach, based on differences between MJ- and MJ+ expression for important growth regulating genes, may need modification. A major difficulty with the E-Flux model is the complexity of the rules governing the Gene-Protein-Reaction (GPR) matrix. Consider the example of AT5G36220, a gene that was found to be significantly upregulated upon MJ elicitation [\(Table 5\)](#page-56-0). AT5G36220 corresponds to the reaction "ascorbate acceptor oxidoreductase" in our model. 70 other genes also correspond to "ascorbate acceptor oxidoreductase" in the model. The Boolean rules for this reaction dictate that any of these genes can drive the reaction. In the MJ- condition, AT5G36220 has a relative expression value of 5.26, while in the MJ+ condition AT5G36220 has a relative expression value of 9.41. The 71 genes that correspond to "ascorbate acceptor oxidoreductase" have expression values ranging from a minimum of 2.76 and 2.77 in the MJ- and MJ+ datasets respectively, and

maximum values of 9.51 and 10.66 the MJ- and MJ+ datasets respectively. The maximum values were used to determine the E-Flux bounds, as given by the Boolean rules. This means that the significant MJ responsiveness of AT5G36220 is overshadowed by 70 other reactions that can drive the reaction. This is a challenging problem with no simple solution. In this case, the complexity and vagueness of the model obscured the effect of MJ on this reaction. Further improvements to the model that may help alleviate this would be more specificity connecting genes to reactions, and fewer non-relevant associations.

An additional problem with the dataset itself is that the study we took it from only collected data until 6 hours post elicitation. This may represent only an initial inductive, wave of transcription that involves signaling proteins but not metabolically involved enzymes. Datasets that include later time points post-elicitation may give more divergent results. Another difficulty is the incompleteness of our model, especially in reactions pertaining to secondary metabolism. The authors of the paper whose dataset we used in our analysis noted that MJ "primarily triggered gene activation," and that only a small fraction of the genes having differential expression were down regulated (Pauwels et al. 2008). Our analysis shows that even genes strongly associated with growth metabolism are not down regulated by MJ elicitation, in that genes with high fluxes (of positive and negative magnitude) in our model do not show significant differences comparing MJ- to MJ+. It is also noted in the original paper, and in subsequent work in the Roberts lab, that MJ has a strong inhibitory effect on the plant growth cycle, and results in the arrest of cells in G2 phase (Patil et al. 2012). It was our thought that this growth arrest would lead to reduced transcription of enzymes involved in synthesis of biomass components. However, it appears that either transcriptional control of growth associated enzymes does not begin until later in the MJ induced cascade, or that the levels of these proteins are controlled by mechanisms other than transcription. A proteomic investigation of enzymes in a longer time-study of MJ elicitation might help to elucidate the mechanism by which MJ induction controls growth metabolism.

Another drawback to our current modeling approach was that most of the genes strongly induced by MJ even at the 6 hour time point are secondary metabolites. It is thought that the secondary metabolism activation may divert flux from growth specific enzymes and depress growth without needing to down regulate transcription of those enzymes. Unfortunately, secondary metabolism related enzymes are not well represented in our current metabolic network. Enzymes pertaining to secondary metabolism are variable and species specific, and are not as well studied as genes pertaining to phenomena such as growth and reproduction. Their pathways are in many cases incomplete, resulting in either blocked reactions in our metabolic network, or lack of transcriptomic information for pathway genes that are represented. If these secondary metabolism genes were better studied, and included in our model, it is possible that forcing flux through them using an approach like E-Flux might yield better modeling of the growth inhibition of MJ+ cultures and also allow us to better study how this flux is diverted.

APPENDICES

APPENDIX A

TABLES FOR GENOME SCALE METABOLIC MODELING OF *ARABIDOPSIS THALIANA* **UNDER METHYL JASMONATE**

ELICITED CONDITIONS

AT5G67590	10.05681986	10.35804973
AT5G67630	8.948104786	8.902883101
NP_174083	n/a	n/a
NP_188430	n/a	n/a
NP_192901	n/a	n/a
NP_194697	n/a	n/a
NP_200227	n/a	n/a
NP_563822	n/a	n/a
NP_565650	n/a	n/a
AT2G44160	10.45995286	10.67263346
AT1G06290	10.33396862	10.25641075
AT1G31180	8.083627336	8.182029957
AT1G62640	4.091637039	4.07828681
AT2G28850	10.09980178	10.72279456
AT2G44160	8.083627336	8.182029957

Table 6 – List of genes included in *A. thaliana* model and relative expression values (from (Pauwels et al. 2008)). Values of "n/a" indicate that the gene was not in the expression dataset for two possible reasons, one being that the gene was not included on the ATH1 chip, and the other being that the probe corresponding to the gene was discarded from the analysis because of technical issues with the microarray.

Table 7 – Table of upper and lower bounds for MJ- and MJ+ conditions in E-Flux simulations.

APPENDIX B

CRYOSECTIONING AND IMMUNOFLUORESCENCE PROTOCOLS

Paraffin Embedding

Fixation and paraffin embedding

- 1. Prepare 37% (w/v) formaldehyde fresh from powder by dissolving in heated water and adding a few drops of NaOH, alternatively 37% formaldehyde that has been stored frozen can be diluted to 4% immediately prior to use
- 2. Make fixative solution of 4% (w/v) formaldehyde and 0.1% (v/v) glutaraldehyde
- 3. Remove aggregates from media by using filter, place aggregates in scintillation vial and cover with excess of fixative solution
- 4. Place uncapped vials in vacuum chamber and apply vacuum for 1 minute
- 5. Cap vials and let sit 24 hours at room temperature
- 6. Remove fixative and apply a graded ethanol series to dehydrate the sample
	- a. Add 30%, 50%, 75% and neat solutions of histological grade ethanol for 30 minutes each, with 3 applications of neat solution
- 7. Immerse sample in pure xylene for 2 hours
- 8. Pipette out solven and add molten paraffin to sample, let sit for 1 hour in incubator at 58

C

- 9. Pour off paraffin and fresh paraffin to sample, let sit for 1 hour
- 10. Transfer aggregates to square sample mold and let solidify to room temperature

Sectioning

- 1. Using a microtome, section wax blocks with embedded sample to 10 μm slices
- 2. Float slices in water bath and then place on lysine-coated slides
- 3. Allow to dry on heated drying block for > 2 hours

Dewaxing and Rehydration

- 1. Place slides in a coplin jar, add 35 ml 100% xylene, incubate for 10 minutes, empty coplin jar, add new 100% xylene and incubate another 10 minutes
- 2. Add 95% ethanol to slides, incubate 10 minutes
- 3. Add 75% ethanol diluted in PBS, incubate 5 minutes
- 4. Add 50% ethanol in PBS, incubate 5 minutes
- 5. Add PBS, incubate 5 minutes
- 6. Remove slides from coplin jar and dip in distilled water to remove PBS salts

Staining

- 1. Incubate slides at RT in coplin jar with 0.01% calcufluor (diluted in water) in dark for 4 hours
- 2. Dip slides in distilled water to rinse
- 3. Stain with PI (1:100 from stock) lay slides flat in dark for 10 minutes
- 4. Put 1-3 drops mounting media (10:1 glycerol to PBS mixsture) and coverslip
- 5. Blot edges with Kim wipe
- 6. Image with microscope

Cryosectioning with sucrose gradient

Solution and Sample preparation

- 1. 24 hours before fixation, add stock solution of Fluorescent Brightener 28 (calcufluor) to live cells to final concentration in media of 0.01% and incubate in the dark for 24 hours
- 2. Make fixative solution (10 ml for each sample vial, this makes 50 ml total)
	- a. 5.4 mls 37% formaldehyde (freshly prepared from powder or frozen stock)
	- b. 20 μL 25% glutaraldehyde
	- c. add nanopure water to reach 50 ml
- 3. Make 2.3 M sucrose, dissolved into water
- 4. Make 0.1 M Sorensen buffer
- 5. Make PBS from packet or from recipe

Cell Fixation

- 1. Filter aggregates and place cells in scintillation vial
- 2. Rinse briefly with PBS to remove media, pipette off PBS
- 3. Pipette 10 ml of prepared fixative into each sample vial
- 4. Place open sample vials in vacuum chamber and turn on vacuum for 1 min
- 5. Remove sample vial from vacuum, cap and Incubate 12-24 hours at 4 C

Cryoprotection

- 1. Pipette out fixation solution from scintillation vials, add 10 ml PBS to cells and let sit for 15 min
- 2. Prepare iterative dilutions of 2.3 M sucrose solution into 0.1 M Sorensen buffer: 25, 33, 50, 66, 75 and 100%
- 3. Pipette out PBS and pipette 25% sucrose solution into vial, let sit for one hour
- 4. Pipette out 25% sucrose solution and pipette in 33% sucrose solution, let sit for one hour, repeat until 100% solution is reached
- 5. For 100% sucrose solution, let sit in refrigerator overnight
- 6. Freeze cells within 24 hours

Freezing

- 1. In a square plastic mold, submerge cells in OCT freezing solution
- 2. Submerge mold in liquid nitrogen for 10 seconds until hardened
- 3. Cover with aluminum foil and freeze at -80 C until use

Cryosectioning without sucrose gradient

Solution and Sample preparation

- 1. 24 hours before fixation, add stock solution of Fluorescent Brightener 28 (calcufluor) to live cells to final concentration in media of 0.01% and incubate in the dark for 24 hours
- 2. Make fixative solution (10 ml for each sample vial, this makes 50 ml total)
	- a. 5.4 mls 37% formaldehyde (freshly prepared from powder or frozen stock)
	- b. 20 μL 25% glutaraldehyde
	- c. add nanopure water to reach 50 ml
- 3. Make 2.3 M sucrose, dissolved into water
- 4. Make PBS from packet or from recipe

Cell Fixation

- 1. Filter aggregates and place cells in scintillation vial
- 2. Rinse briefly with PBS to remove media, pipette off PBS
- 3. Pipette 10 ml of prepared fixative into each sample vial
- 4. Place open sample vials in vacuum chamber and turn on vacuum for 1 min

5. Remove sample vial from vacuum, cap and Incubate 12-24 hours at 4 C

Cryoprotection

- 1. Pipette out fixation solution from scintillation vials, add 10 ml PBS to cells and let sit for 15 min
- 2. Pipette out PBS and pipette sucrose solution into vial, let sit overnight.
- 3. Freeze cells within 24 hours

Freezing

- 1. In a square plastic mold, submerge cells in OCT freezing solution
- 2. Submerge mold in liquid nitrogen for 10 seconds until hardened
- 3. Cover with aluminum foil and freeze at -80 C until use

Flash freezing of fresh tissue

- 1. Filter aggregates and place cells in scintillation vial
- 2. Rinse briefly with PBS to remove media, pipette off PBS
- 3. In a square plastic mold, submerge cells in OCT freezing solution
- 4. Submerge mold in liquid nitrogen for 10 seconds until hardened
- 5. Cover with aluminum foil and freeze at -80 C until use

Cryosectioning

- 1. Cool cryostat, blade and metal sample chucks to -20 C
- 2. Remove frozen sample from mold and if desired, cut to desired size
- 3. Add a thick layer of OCT compound to metal sample chuck and lightly press sample onto the OCT
- 4. Allow sample to harden inside cryostat for >30 minutes
- 5. Set cryostat to thick slices $(50 \mu m)$, set chuck/sample in place and begin turning crank until blade comes into contact with sample. Make some thick slices to even out top of sample
- 6. Set cryostat to desired slice thickness $(5 \text{ to } 10 \text{ µm})$ and begin slicing
	- a. After each slice, take a warm poly-lysine coated slide and hover the slide a few mm above sample – sample will jump up to slide
	- b. Allow each slide to sit at room temperature for >1 hour to dry

Immunostaining for paclitaxel

- 1. Solution preparation:
	- a. Prepare blocking buffer: 2% nonfat milk powder (w/w%) in PBS.
	- b. Prepare primary antibody solution: 1:100 dilution of anti-taxol mouse antibody in blocking buffer
	- c. Prepare secondary antibody solution: 1:200 dilution of anti-mouse antibody in blocking buffer
- 2. In a coplin jar, immerse dried slides in blocking buffer and let sit 1 hour
- 3. Dip slides in PBS briefly to remove blocking buffer
- 4. In a box lined with damp paper towels, lay slides flat. Blot excess blocking buffer from edge of slide using a Kim wipe.
- 5. Pipette 1 ml of diluted primary antibody onto each slide. Cut a piece of parafilm to cover the slices on the slide, and lay parafilm gently over the slices (do not stretch)
- 6. Let slides incubate in the dark for1.5 hours
- 7. Remove parafilm from slides, place in coplin jar, and immerse in PBS for 15 minutes on shaker, put slides in fresh PBS and wash for 15 minutes, repeat one more time (3 total washes)
- 8. Pipette 1 ml of diluted secondary antibody onto each slide. Cut a piece of parafilm to cover the slices on the slide, and lay parafilm gently over the slices
- 9. Let the slides incubate in the dark for 1.5 hours
- 10. Remove parafilm and perform 3 PBS washes as in step 7 above
- 11. Dip slides in distilled water to remove salts
- 12. If desired, stain nuclei with PI as in paraffin embedding protocol above.
- 13. To preserve slides, add one drop of aqueous mounting medium and gently place coverslip over slices.

Figure 8 – Example of a cryosectioned *Taxus* aggregate stained with calcufluor (blue) to designate cell wall and propidium iodide (red) to denote the nucleus. This aggregate was sliced into 30 μm slices.

BIBLIOGRAPHY

- Ajikumar, P.K. et al., 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science (New York, N.Y.)*, 330(6000), pp.70–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3034138&tool=pmcentrez&ren dertype=abstract [Accessed June 10, 2011].
- Baebler, Š. et al., 2005. Establishment of cell suspension cultures of yew (Taxus \times media Rehd. and assessment of their genomic stability. *In Vitro Cellular & Developmental Biology - Plant*, 41(3), pp.338–343. Available at: http://www.springerlink.com/index/10.1079/IVP2005642 [Accessed December 10, 2010].
- Becker, S. et al., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2(3), pp.727–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17406635 [Accessed July 14, 2012].
- Becker, S. & Palsson, B.O., 2008a. Context-specific metabolic networks are consistent with experiments. *PLoS computational biology*, 4(5), p.e1000082. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2366062&tool=pmcentrez&ren dertype=abstract [Accessed July 13, 2012].
- Becker, S. & Palsson, B.O., 2008b. Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC systems biology*, 2, p.14. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2248557&tool=pmcentrez&ren dertype=abstract [Accessed July 9, 2013].
- Bordbar, A. & Palsson, B., 2012. Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *Journal of internal medicine*, 271(2), pp.131– 41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22142339 [Accessed September 7, 2012].
- Borodina, I. & Nielsen, J., 2005. From genomes to in silico cells via metabolic networks. *Current opinion in biotechnology*, 16(3), pp.350–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15961036 [Accessed July 13, 2012].
- Bourgaud, F., 2001. Production of plant secondary metabolites: a historical perspective. *Plant Science*, 161(5), pp.839–851. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168945201004903.
- Bräutigam, A. & Gowik, U., 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant biology (Stuttgart, Germany)*, 12(6), pp.831–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21040298 [Accessed July 11, 2011].
- Bringi, V. et al., 2007. Enhanced production of Taxol and Taxanes by cell cultures of Taxus species.
- Chang, S.S. et al., 1994. Stable genetic transformation of Arabidopsis thaliana by Agrobacterium inoculation in planta. *The Plant journal : for cell and molecular biology*, 5(November 1993), pp.551–558.
- Chemler, J.A. & Koffas, M.A.G., 2008. Metabolic engineering for plant natural product biosynthesis in microbes. *Current opinion in biotechnology*, 19(6), pp.597–605. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18992815 [Accessed July 20, 2011].
- Cheng, A.-X. et al., 2007. Plant Terpenoids : Biosynthesis and Ecological Functions. *Journal of Integrative Plant Biology*, 49(2), pp.179–186.
- Christen, A., Gibson, D.M. & Bland, J., 1991. Production of taxol or taxol-like compounds in cell culture.
- Colijn, C. et al., 2009. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. *PLoS computational biology*, 5(8), p.e1000489. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2726785&tool=pmcentrez&ren dertype=abstract [Accessed March 20, 2014].
- Conesa, A. et al., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, 21(18), pp.3674–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16081474 [Accessed July 16, 2012].
- Cragg, G. et al., 1993. The taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *Journal of natural products*, 56(10), pp.1657–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7903979.
- Cragg, G. & Snader, K., 1991. Taxol: The Supply Issue. *Cancer Cells*.
- Cragg, G.M. et al., 1995. Pharmaceutical Prospecting and the Potential for Pharmaceutical Crops. Natural Product Drug Discovery and Development at the United States National Cancer Institute. *Annals of the Missouri Botanical Garden*, 82(1), pp.47–53.
- Croteau, R. et al., 2006. Taxol biosynthesis and molecular genetics. *Phytochemistry reviews : proceedings of the Phytochemical Society of Europe*, 5(1), pp.75–97. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2901146&tool=pmcentrez&ren dertype=abstract [Accessed August 2, 2010].
- Dal'Molin, C.G.D.O. et al., 2010. C4GEM, a genome-scale metabolic model to study C4 plant metabolism. *Plant physiology*, 154(4), pp.1871–85. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2996019&tool=pmcentrez&ren dertype=abstract [Accessed August 21, 2012].
- Denis, J.-N. et al., 1988. A Highly Efficient, Practical Approach to Natural Taxol. *J. Am. Chem. Soc.*, 110, pp.5917–5919.
- Dietrich, J. a et al., 2009. A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3). *ACS chemical biology*, 4(4), pp.261–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19271725.
- Duarte, N.C., Herrgård, M.J. & Palsson, B.Ø., 2004. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome research*, 14(7), pp.1298–309. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=442145&tool=pmcentrez&rende rtype=abstract [Accessed July 19, 2012].
- Edwards, J.S. & Palsson, B.O., 2000. The Escherichia coli MG1655 in silico metabolic genotype : Its definition , characteristics , and capabilities. *PNAS*.
- Eibl, R. & Eibl, D., 2007. Design of bioreactors suitable for plant cell and tissue cultures. *Phytochemistry Reviews*, 7(3), pp.593–598. Available at: http://www.springerlink.com/index/10.1007/s11101-007-9083-z [Accessed June 30, 2011].
- Evans, B.E. et al., 1988. Methods for Drug Discovery: Development of Potent, Selective, Orally Effective Cholecystokinin Antagonists. *J. Med. Chem.*, 31, pp.2235–2246.
- Expósito, O. et al., 2009. Biotechnological production of taxol and related taxoids: current state and prospects. *Anti-cancer agents in medicinal chemistry*, 9(1), pp.109–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19149486.
- Farkya, S., Bisaria, V.S. & Srivastava, A.K., 2004. Biotechnological aspects of the production of the anticancer drug podophyllotoxin. *Applied microbiology and biotechnology*, 65(5), pp.504–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15378293 [Accessed November 3, 2011].
- Feist, A.M. & Palsson, B.O., 2010. The biomass objective function. *Current opinion in microbiology*, 13(3), pp.344–349. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20430689.
- Firn, R.D. & Jones, C.G., 2003. Natural products: A simple model to explain chemical diversity. *Natural Product Reports*, 20(4), p.382. Available at: http://xlink.rsc.org/?DOI=b208815k [Accessed July 21, 2011].
- Fischbach, M.A. & Clardy, J., 2007. One pathway , many products. *Nature Chemical Biology*, 3(7), pp.353–355.
- Van der Fits, L. & Memelink, J., 2000. ORCA3, a Jasmonate-Responsive Transcriptional Regulator of Plant Primary and Secondary Metabolism. *Science*, 289(5477), pp.295–297. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.289.5477.295 [Accessed October 12, 2011].
- Fonseca, S. et al., 2009. (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nature chemical biology*, 5(5), pp.344–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19349968 [Accessed July 24, 2012].
- Gelvin, S.B., 2003. Improving plant genetic engineering by manipulating the host. *Trends in biotechnology*, 21(3), pp.95–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12628361.
- Georgiev, M., Ludwig-Muller, J. & Bley, T., 2010. Hairy Root Culture: Copying Nature in New Bioprocesses. In R. Arora, ed. *Medicinal Plant Biotechnology*. Cambridge, MA: CABI, pp. 156–175.
- Georgiev, M., Pavlov, A. & Ilieva, M., 2006. Selection of high rosmarinic acid producing Lavandula vera MM cell lines. *Process Biochemistry*, 41(9), pp.2068–2071. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1359511306001942 [Accessed January 19, 2012].
- Georgiev, M.I., Weber, J. & Maciuk, A., 2009. Bioprocessing of plant cell cultures for mass production of targeted compounds. *Applied microbiology and biotechnology*, 83(5), pp.809– 23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19488748 [Accessed January 9, 2011].
- De Geyter, N. et al., 2012. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends in Plant Science*, 17(6), pp.349–359. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1360138512000532 [Accessed March 29, 2012].
- Gräwert, T. et al., 2011. Biochemistry of the non-mevalonate isoprenoid pathway. *Cellular and molecular life sciences : CMLS*, 68(23), pp.3797–814. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21744068 [Accessed September 4, 2012].
- Gundlach, H. et al., 1992. Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Plant biology*, 89(March), pp.2389–2393.
- Hale, V. et al., 2007. Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. *The American journal of tropical medicine and hygiene*, 77(6 Suppl), pp.198–202. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18165493.

Hampel, D., Mau, C.J.D. & Croteau, R.B., 2009. Taxol biosynthesis: Identification and characterization of two acetyl CoA:taxoid-O-acetyl transferases that divert pathway flux away from Taxol production. *Archives of biochemistry and biophysics*, 487(2), pp.91–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2748654&tool=pmcentrez&ren dertype=abstract [Accessed March 3, 2011].

- Hanada, K. et al., 2009. Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. *Genome biology and evolution*, 1, pp.409–14. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2817435&tool=pmcentrez&ren dertype=abstract [Accessed May 29, 2013].
- Harding, K., 2004. Genetic integrity of cryopreserved plant cells: A review. *Cryoletters*, 25(1), pp.3–22.
- Hartmann, T., 2007. From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry*, 68(22-24), pp.2831–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17980895 [Accessed July 20, 2011].
- Henry, C.S. et al., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology*, 28(9), pp.977–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20802497 [Accessed July 16, 2012].
- Holton, R.A. et al., 1994. First Total Synthesis of Taxol. 1. Functionalization of the B Ring. *J. Am. Chem. Soc.*, 116(2), pp.1597–1598.
- Humphreys, J.M. & Chapple, C., 2002. Rewriting the lignin roadmap. *Current opinion in Plant Biology*, 5, pp.224–229.
- Jensen, P.A., Lutz, K.A. & Papin, J.A., 2011. TIGER : Toolbox for integrating genome-scale metabolic models , expression data , and transcriptional regulatory networks. *BMC Systems Biology*, 5(1), p.147. Available at: http://www.biomedcentral.com/1752-0509/5/147.
- Kanehisa, M. & Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), pp.27–30. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102409&tool=pmcentrez&rende rtype=abstract.
- Kim, Y., Wyslouzil, B.E. & Weathers, P.J., 2002. Secondary metabolism of hairy root cultures in bioreactors. *In Vitro Cellular & Developmental Biology - Plant*, 38(1), pp.1–10. Available at: http://www.springerlink.com/index/10.1079/IVP2001243 [Accessed June 16, 2011].
- Kingston, D.G.I., 1994. Taxol: The chemistry and structure-activity relationships of a novel anticancer agent. *Trends in biotechnology*, 12(6), pp.222–227.
- Koehn, F.E. & Carter, G.T., 2005. The evolving role of natural products in drug discovery. *Nature reviews. Drug discovery*, 4(3), pp.206–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15729362 [Accessed July 23, 2011].
- Kolewe, M.E., Gaurav, V. & Roberts, S.C., 2008. Pharmaceutically active natural product synthesis and supply via plant cell culture technology. *Molecular pharmaceutics*, 5(2), pp.243–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18393488.
- Kolewe, M.E., Henson, M.A. & Roberts, S.C., 2011. Analysis of aggregate size as a process variable affecting paclitaxel accumulation in Taxus suspension cultures. *Biotechnology progress*, 27(5), pp.1365–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21692199 [Accessed January 18, 2012].
- Kutchan, T.M., 2005. A role for intra- and intercellular translocation in natural product biosynthesis. *Current opinion in plant biology*, 8(3), pp.292–300. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15860426 [Accessed September 2, 2011].
- Lee, E.-K. et al., 2010. Cultured cambial meristematic cells as a source of plant natural products. *Nature biotechnology*, 28(11), pp.1213–1217. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20972422 [Accessed October 28, 2010].
- Lee, S.J. et al., 2005. Metabolic Engineering of Escherichia coli for Enhanced Production of Succinic Acid , Based on Genome Comparison and In Silico Gene Knockout Simulation Metabolic Engineering of Escherichia coli for Enhanced Production of Succinic Acid , Based on Genome Com. *Applied and environmental microbiology*, 71(12), pp.7880–7887.
- Leonard, E. et al., 2008. Strain improvement of recombinant Escherichia coli for efficient production of plant flavonoids. *Molecular pharmaceutics*, 5(2), pp.257–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18333619.
- Leone, L.M. & Roberts, S.C., 2013. Accessing Anti-Cancer Natural Products by Plant Cell Culture. In F. E. Koehn, ed. *Natural Products and Cancer Drug Discovery*. Springer New York, pp. 193–211.
- Li, L. et al., 2012. Determining degradation and synthesis rates of arabidopsis proteins using the kinetics of progressive 15N labeling of two-dimensional gel-separated protein spots. *Molecular & cellular proteomics : MCP*, 11(6), p.M111.010025. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3433911&tool=pmcentrez&ren dertype=abstract [Accessed March 21, 2014].
- Lipkus, A.H. et al., 2008. Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry. *The Journal of organic chemistry*, 73(12), pp.4443–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18505297.
- Lloyd, J. & Meinke, D., 2012. A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant physiology*, 158(3), pp.1115–29. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3291275&tool=pmcentrez&ren dertype=abstract [Accessed September 29, 2013].
- Mintz-Oron, S. et al., 2011. Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proceedings of the National Academy of Sciences*, (7), pp.1–6. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1100358109 [Accessed December 20, 2011].
- Mueller, M.J. et al., 1993. Signaling in the elicitation process is mediated through the octadecanoid pathway leading to jasmonic acid. *Proceedings of the National Academy of Sciences of the United States of America*, 90(16), pp.7490–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=47167&tool=pmcentrez&render type=abstract.
- Mustafa, N.R. et al., 2011. Initiation, growth and cryopreservation of plant cell suspension cultures. *Nature protocols*, 6(6), pp.715–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21637194 [Accessed August 1, 2011].
- Nagel, A. et al., 2005. Extension of the Visualization Tool MapMan to Allow Statistical Analysis of Arrays , Display of Coresponding Genes , and Comparison with Known Responses 1. , 138(July), pp.1195–1204.
- Naik, R. et al., 1988. An antiinflammatory cum immunomodulatory piperidinylbenzopyranone from dysoxylum binectariferum : isolation, structure and total synthesis. *Tetrahedron*, 44(7), pp.2081–2086.
- Naill, M.C. & Roberts, S.C., 2005a. Culture of isolated single cells from Taxus suspensions for the propagation of superior cell populations. *Biotechnology letters*, 27(21), pp.1725–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16247682 [Accessed August 2, 2010].
- Naill, M.C. & Roberts, S.C., 2005b. Flow cytometric identification of Paclitaxel-accumulating subpopulations. *Biotechnology progress*, 21(3), pp.978–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15932283.
- Nakagawa, A. et al., 2011. A bacterial platform for fermentative production of plant alkaloids. *Nature communications*, 2(May), p.326. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3112539&tool=pmcentrez&ren dertype=abstract [Accessed July 26, 2011].
- Newman, D.J. & Cragg, G.M., 2007. Natural products as sources of new drugs over the last 25 years. *Journal of natural products*, 70(3), pp.461–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17309302.
- Ni, X. et al., 2011. Enhancement of camptothecin production in Camptotheca acuminata hairy roots by overexpressing ORCA3 gene. *Journal of Applied Pharmaceutical Science*, 01(08), pp.85–88.
- Nicolau, K.C. et al., 1994. Total synthesis of taxol. *Nature*, 367, pp.630–634.
- Niklas, J., Schneider, K. & Heinzle, E., 2010. Metabolic flux analysis in eukaryotes. *Current opinion in biotechnology*, 21(1), pp.63–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20163950 [Accessed June 21, 2011].
- Nims, E. et al., 2006. Expression profiling of genes involved in paclitaxel biosynthesis for targeted metabolic engineering. *Metabolic engineering*, 8(5), pp.385–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16793302 [Accessed August 2, 2010].
- O'Grady, J. et al., 2012. Metabolic cartography: experimental quantification of metabolic fluxes from isotopic labelling studies. *Journal of experimental botany*, 63(6), pp.2293–308. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22371075 [Accessed July 26, 2012].
- De Oliveira Dal'Molin, C.G. et al., 2010. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant physiology*, 152(2), pp.579–89. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2815881&tool=pmcentrez&ren dertype=abstract [Accessed July 28, 2012].
- Orth, J.D., Thiele, I. & Palsson, B.Ø., 2010. What is flux balance analysis? *Nature biotechnology*, 28(3), pp.245–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20212490.
- Ortholand, J.-Y. & Ganesan, a, 2004. Natural products and combinatorial chemistry: back to the future. *Current opinion in chemical biology*, 8(3), pp.271–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15183325 [Accessed October 11, 2011].
- Patil, R. et al., 2012. Contribution of taxane biosynthetic pathway gene expression to observed variability in paclitaxel accumulation in Taxus suspension cultures. *Biotechnology journal*, 7(3), pp.418–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22095859 [Accessed September 16, 2012].
- Patil, R., 2013. *Molecular and Population Level Approaches to Understand Taxus Metabolism in Cell Suspension Cultures*. University of Massachusetts Amherst.
- Patil, R. et al., 2011. Taxane biosynthetic pathway gene expression in Taxus suspension cultures with different bulk paclitaxel accumulation patterns $\&$ a molecular approach to understand variability in paclitaxel accumulation. *Biotechnology journal*, pp.1–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22095859 [Accessed January 16, 2012].
- Pauwels, L. et al., 2008. Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4), pp.1380–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2234147&tool=pmcentrez&ren dertype=abstract.
- Pauwels, L., Inzé, D. & Goossens, A., 2009. Jasmonate-inducible gene: What does it mean? *Trends in plant science*, 14(2), pp.87–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19162528 [Accessed March 10, 2013].
- Piques, M. et al., 2009. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Molecular systems biology*, 5(314), p.314. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2779082&tool=pmcentrez&ren dertype=abstract [Accessed March 21, 2014].
- Poolman, M.G. et al., 2009. A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant physiology*, 151(3), pp.1570–81. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2773075&tool=pmcentrez&ren dertype=abstract [Accessed July 20, 2012].
- Poolman, M.G. et al., 2013. Responses to Light Intensity in a Genome-Scale Model of Rice Metabolism. *Plant physiology*, 162(June), pp.1060–1072. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23640755 [Accessed May 31, 2013].
- Reed, J.L. et al., 2003. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome biology*, 4(9), p.R54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=193654&tool=pmcentrez&rende rtype=abstract.
- Ro, D.-K. et al., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086), pp.940–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16612385 [Accessed July 6, 2011].
- Roberts, S. & Kolewe, M., 2010. Plant natural products from cultured multipotent cells. *Nature biotechnology*, 28(11), pp.1175–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21057485 [Accessed January 7, 2011].
- Roberts, S.C., 2007. Production and engineering of terpenoids in plant cell culture. *Nature chemical biology*, 3(7), pp.387–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17576426.
- Ruiz-May, E. et al., 2011. Methyl jasmonate induces ATP biosynthesis deficiency and accumulation of proteins related to secondary metabolism in Catharanthus roseus (L.) G. hairy roots. *Plant & Cell Physiology*, 52(8), pp.1401–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21727181 [Accessed August 20, 2012].
- Saha, R., Suthers, P.F. & Maranas, C.D., 2011. Zea mays i RS1563 : A Comprehensive Genome-Scale Metabolic Reconstruction of Maize Metabolism. *PLoS biology*, 6(7).
- Schäfer, H. & Wink, M., 2009. Medicinally important secondary metabolites in recombinant microorganisms or plants: progress in alkaloid biosynthesis. *Biotechnology journal*, 4(12), pp.1684–703. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19946877 [Accessed November 29, 2011].
- Schnur, D.M. et al., 2011. *Chemoinformatics and Computational Chemical Biology* J. Bajorath, ed., Totowa, NJ: Humana Press. Available at: http://www.springerlink.com/index/10.1007/978-1-60761-839-3 [Accessed June 17, 2011].
- Schwanhäusser, B. et al., 2011. Global quantification of mammalian gene expression control. *Nature*, 473(7347), pp.337–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21593866 [Accessed March 20, 2014].
- Senger, R.S. et al., 2006. Development of a culture sub-population induction model: signaling pathways synergy and taxanes production by Taxus canadensis. *Biotechnology progress*, 22(6), pp.1671–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17137317.

Solomon, L.D., 2011. *Synthetic Biology: Science, Business, and Policy*, Transaction Publishers.

- Sriram, G., Fulton, D.B. & Shanks, J. V, 2007. Flux quantification in central carbon metabolism of Catharanthus roseus hairy roots by 13C labeling and comprehensive bondomer balancing. *Phytochemistry*, 68(16-18), pp.2243–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17532015 [Accessed September 14, 2012].
- Stitt, M., 2013. Systems-integration of plant metabolism: means, motive and opportunity. *Current opinion in plant biology*, 16(3), pp.381–388. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23642787 [Accessed May 27, 2013].
- Stitt, M., Sulpice, R. & Keurentjes, J., 2010. Metabolic networks: how to identify key components in the regulation of metabolism and growth. *Plant physiology*, 152(2), pp.428– 44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2815907&tool=pmcentrez&ren dertype=abstract [Accessed July 28, 2011].
- Swiatek, A. et al., 2002. Differential Effect of Jasmonic Acid and Abscisic Acid on Cell Cycle Progression in Tobacco BY-2 Cells 1. *Plant physiology*, 128(January), pp.201–211.
- Verpoorte, R. et al., 1999. Metabolic engineering of plant secondary metabolite pathways for the production of fine chemicals. *Cultures*, (Verpoorte), pp.467–479.
- Weathers, P. et al., 2008. Mist reactors : principles , comparison of various systems , and case studies. *Biotechnology and Bioengineering*, 3(October), pp.29–37.
- Werck-Reichhart, D. & Feyereisen, R., 2000. Protein family review Cytochromes P450 : a success story. *Genome biology*, 1(6), pp.1–9.
- Wilson, R.M. & Danishefsky, S.J., 2006. Small molecule natural products in the discovery of therapeutic agents: the synthesis connection. *The Journal of organic chemistry*, 71(22), pp.8329–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17064003.
- Wilson, S. & Roberts, S.C., 2011. Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. *Plant biotechnology journal*, pp.1–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22059985 [Accessed November 15, 2011].
- Yukimune, Y. et al., 1996. Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in Taxus cell suspension cultures. *Nature Biotechnology*, 14.
- Zhao, J., Davis, L.C. & Verpoorte, R., 2005. Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnology advances*, 23(4), pp.283–333. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15848039 [Accessed July 24, 2011].