8-2016

Genetic Basis of Biosynthesis and Cytotoxic Activity of Medicago truncatula Triterpene Saponins

Brynn Kathleen Lawrence
University of Arkansas, Fayetteville

Genetic Basis of Biosynthesis and Cytotoxic Activity of
Medicago truncatula Triterpene Saponins

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Cell and Molecular Biology

By

Brynn Kathleen Lawrence
Ouachita Baptist University
Bachelor of Science in Biology, 2014

August 2016
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Dr. Kenneth L. Korth
Thesis Director

_____   _____
Dr. Sun-Ok Lee   Dr. David S. McNabb
Committee Member   Committee Member

Abstract

Saponins are a large family of specialized metabolites produced in many plants. They can have negative effects on a number of plant pests and are thought to play a role in plant defense. With current and possible future uses in industry and agriculture, saponins have also been shown to be hypocholesterolemic, hypoglycemic, immunostimulatory, antioxidative, anti-inflammatory, and cytotoxic. In spite of their usefulness, our understanding of the genetic basis for saponin biosynthesis is still incomplete. We generated recombinant populations with parents from genetically distinct accessions of *Medicago truncatula*, with either high or low accumulation and varying profiles of saponins. Primers for a PCR-based parental test were developed from single-nucleotide polymorphisms in the sequence encoding CYP72A68, a cytochrome P450 enzyme involved in the biosynthesis of *M.truncatula* sapogenins. Comparison of translated CYP72A68 amino acid sequences across accessions revealed high similarity, and comparison with CYP72A proteins from other plant species suggests similar protein functioning across the accessions. The F_2 generation plants from one recombinant population were screened for hemolytic saponin accumulation levels in leaf extracts by measuring cleared zones in blood agar plates. Three distinct phenotypes were observed. Plants in the F_2 generation showed either a high or low hemolysis phenotype matching those of the parent plants, or an intermediate level of hemolysis. The high:medium:low phenotypic ratio for 141 plants was 1:3:1. Quantitative RT-PCR showed a correlation between expression of CYP716A12, CYP72A67, and CYP72A68 genes, all encoding cytochrome P450 enzymes involved in synthesis of hemolytic sapogenins, and the three hemolysis phenotypes. Concurrently, we found that treatment of Caco-2 human colon cancer cells with saponin extracts from four *M. truncatula* accessions resulted in decreased cell proliferation over time, and that this effect did not appear to be mediated through apoptosis induction. The saponin extracts were analyzed by HPLC-MS to identify individual saponins that could contribute to cytotoxic activity. Therefore, accession differences in *M. truncatula* saponin accumulation result from differential regulation of saponin synthesis gene expression and inheritance of these differences depends on more than a single gene with dominant and recessive alleles. Saponin extracts are shown to have a negative impact on cancer cell lines.

Acknowledgements

I am extremely grateful to my advisor Dr. Ken Korth for his guidance and mentorship that have pushed me to grow as a scientist and made completion of this thesis possible. I would also like to thank my committee members Dr. Sun-Ok Lee and Dr. David McNabb for always making themselves available to provide additional direction and support. I thank my current and past lab members for their friendship and for being excellent people who were a joy to work with. Special thanks go to Xuan Gu, Cindi Brownmiller, and Lacy Nelson for patiently training me, contributing to my research, and working so hard to make sure I always had the supplies and support I needed to do my best work. Thank you to my former professors at Ouachita Baptist University, especially Dr. Ruth Plymale and Dr. Nathan Reyna, for so excellently preparing me for graduate level work and continuing to be such helpful mentors. Finally, I am extremely grateful to my family and friends, and especially my incredible husband, for their constant encouragement and for giving me the confidence and motivation to undertake such a challenging task.

Table of Contents

List of Figures

List of Tables

Chapter 1

***Medicago truncatula* triterpene saponins**

Introduction

Specialized metabolism in plants

Plants are under constant attack by a wide variety of pests, including bacteria, fungi, nematodes, insects, and even mammals. Without the ability to physically escape these pests, plants have had to develop their own unique methods of defending themselves. One of the key defense mechanisms plants utilize is specialized metabolism, so named because the resulting compounds are not found in all plant species and are therefore not vital to the primary metabolism all plants share (Pichersky and Lewinsohn, 2011). It is estimated that plants as a whole produce at least 200,000 different compounds through specialized metabolism, although no single plant species has the ability to produce the entire range of these compounds (Mithöfer and Boland, 2012). Instead, individual plant species have evolved the ability to produce the specific compounds they need to deal with the pressures of their particular ecological situation (Pichersky and Lewinsohn, 2011). In addition to the diversity of specialized metabolites found in the plant kingdom, individual plants species are able to produce a diverse range of specialized metabolites. This provides the plant with resistance to multiple pests and the opportunity to produce metabolites that can work synergistically to increase efficacy (Howe and Jander, 2008). When used against *Spodoptera litura,* the monoterpenoid plant essential oils thymol and *trans*-anethole are over two-times more effective than would be expected if they worked in a purely additive manner (Hummelbrunner and Isman, 2001). Specialized metabolism can take place constitutively in a plant or be induced by pest attack (Howe and Jander, 2008). The induced response to pest attack occurs both locally at the wound site and systemically throughout the plant, and both constitutive and induced metabolism can provide direct or indirect defense (Kessler and Baldwin, 2002). Direct defenses work alone to negatively affect the pest's ability to harm the plant by being toxic, antidigestive, or unpalatable, while indirect defenses work by attracting predators and parasitoids of the attacking herbivorous pest, usually through the production of volatile organic compounds (VOCs) (Mithöfer and Boland, 2012). Virtually all higher plants exhibit this

behavior; for example, *Lotus japonicus* produces a mixture of VOCs in response to spider mite damage that attracts the predatory mite *Phytoseiulus persimilus* (Ozawa et al., 2000).

The most important classes of specialized metabolites are the terpenoids, glucosinolates, flavonoids, phenolics, and alkaloids (Kliebenstein, 2012). Most relevant to this research, terpenoids derived from 5-carbon isoprenoid units are the most diverse class of specialized metabolites, including over 40,000 different described compounds. Terpenoids play many roles in the plants that produce them. Mono-, sesqui- and homo-terpenes serve as major components of the VOC blends many plants produce to attract pest parasites and predators (Mithöfer and Boland, 2012). Terpenoids can also act in a more direct way to simply repel pests. Transgenic *Arabidopsis thaliana* plants that overexpressed a terpene synthase gene and had increased levels of the monoterpene linalool were shown to significantly repel *Myzus persicae* aphids in a dual-choice assay (Aharoni et al., 2003). Mono- and sesqui-terpenes present in the resin produced by many conifers act as microbial and insect toxins (Phillips and Croteau, 1999). Some terpenoids, including the monoterpenes carvacrol and D-limonene, have even been shown to play an allelopathic role, interfering with basic plants functions such as respiration, seed germination, and the nitrogen cycle to deter growth and survival (Maffei et al., 2011). Besides their role in plant defense, terpenoids have proven to be useful to humans, exhibiting a variety of medicinal properties. Taxol, the widely used breast and ovarian cancer treatment, and artemisinin, an anti-malarial agent, as well as many promising compounds in clinical trials, such as D-limonene and β-sitosterol, are all terpenoids (Croteau et al., 2006; Pollier et al., 2011b; Singh and Sharma, 2014). With their chemical diversity and varied biological activities, specialized metabolites, and terpenoids in particular, represent a vast collection of compounds that could be exploited to help solve many of the problems humans face today.

Saponins: basic structure and properties

Saponins are the most complex group of terpenoids in terms of diversity of family members. These structurally and functionally diverse compounds are made up of a carbon-rich aglycone backbone, often called a sapogenin, glycosidically linked to a polar sugar moiety (Lacaille-Dubois, 2007). There are two types of saponins, differentiated by the chemistry of their backbone structure. The aglycone component of steroidal saponins has a 27-carbon backbone, and that of triterpene saponins has a 30-carbon backbone (Weng et al., 2011). Within these two groups there is opportunity for further

differentiation. There are nine reported classes of triterpene backbones with a total of 100 individual backbones currently known (Moses et al., 2014a; Thimmappa et al., 2014). Saponins can also be grouped by the number of attached saccharide side chains. Monodesmosidic saponins contain one saccharide side chain attached at the C-3 position, while bidesmosidic saponins contain a second saccharide side chain at either the C-26 or C-28 position (Francis et al., 2002). Tridesmoside saponins with three saccharide side chains attached at different positions exist but are rare (Augustin et al., 2011). Some of the possible monosaccharide components of the side chain include glucose, xylose, galactose, rhamnose, glucuronic acid, and methylpentose (Francis et al., 2002). The multiple possibilities in aglycone types, saccharide side chain composition, and position of saccharide side chain attachment give rise to the immense chemical diversity found among saponins (Francis et al., 2002). The combination of the lipophilic backbone with the hydrophilic sugar moiety makes saponins amphiphilic and gives them detergent, wetting, emulsifying, and surfactant properties (Güçlü-Üstündağ and Mazza, 2007). In fact, saponins were named for their ability to create stable soap-like foams in aqueous solutions; "sapo" is the Latin word for soap (Weng et al., 2011). A selection of steroidal and triterpene aglycone backbone are depicted in Figure 1.
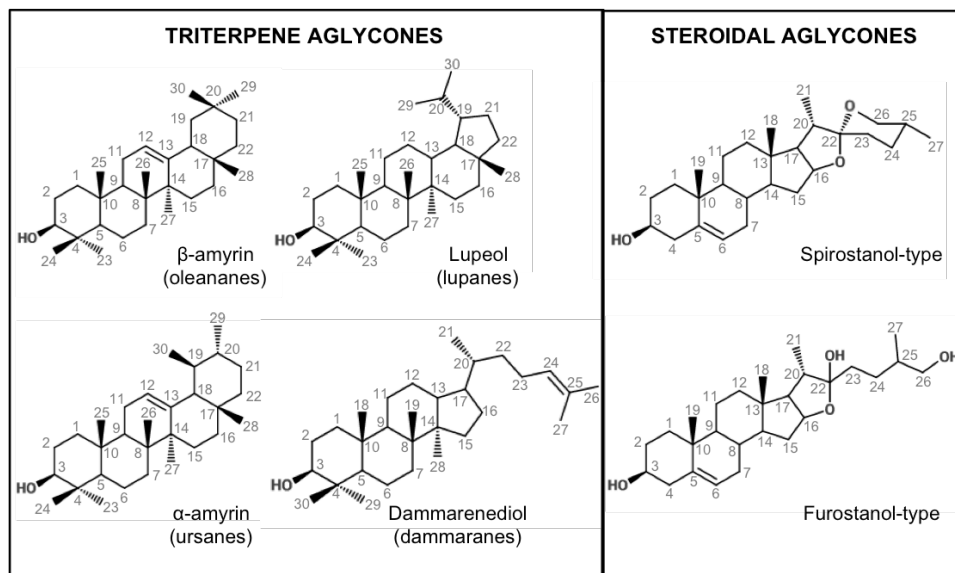
Figure 1. Triterpene and steroidal saponin aglycone chemical structures. Carbon numbering of furostanol- and spirostanol-type steroidal saponin aglycones and triterpene aglycones representative of oleanane-, lupane-, ursane-, and dammarane-type sapogenins. Figure adapted from Moses et al. (2014).

Saponins are mainly produced by plants, but are also found in some marine organisms, insects, and bacteria (Thimmappa et al., 2014; Weng et al., 2011). Among plants, saponins are most prevalent among angiosperms (Magnoliophyta), with dicotyledonous plants accumulating primarily triterpene saponins, and monocotyledonous plants producing mostly steroidal saponins (Moses et al., 2014a). In addition, both wild plants and cultivated crops produce saponins. Triterpene saponins are more common in cultivated crops, while steroidal saponins are predominant in herbs and other plants with health-promoting properties (Francis et al., 2002). The main dietary source of saponins is the legumes, including soybeans, chickpeas, peanuts, lentils, peas, and other beans, while cereals and grasses, with the exception of oats, are usually deficient in saponins (Güçlü-Üstündağ and Mazza, 2007; Moses et al., 2014a). Due to their biological activity against many plant pathogens, pests, and herbivores and the ability of jasmonates to trigger their production, saponins are commonly considered to play a part in plant defense (Augustin et al., 2011). Saponin accumulation varies across plant organs and tissue types and can fluctuate according to the season, as well as a plant's species, cultivar, and developmental stage. In addition, biotic stresses, such as infection or herbivory, and abiotic stresses, such as light, temperature, humidity and nutrient starvation, can affect saponin accumulation levels (Moses et al., 2014a). In the licorice plant *Glycyrrhiza glabra*, the medically important triterpene saponin glycyrrhizin was produced in higher amounts as a result of severe drought stress (Nasrollahi et al., 2014).

Steroidal and triterpene saponin synthesis originates in the mevalonate (MVA) pathway with the conversion of acetyl-CoA into the 5-carbon isopentyl pyrophosphate (IPP). Some of the IPP molecules are isomerized into dimethylallyl pyrophosphate (DMAPP), and then 2 IPP molecules and 1 DMAPP molecule condense to form 15-carbon farnesyl pyrophosphate (FPP). Two FPP molecules condense to form squalene, which is then epoxidized to form the linear 30-carbon 2,3 oxidosqualene. The successive variable cyclization of 2,3 oxidosqualene results in the production of a range of aglycone backbones and serves as the point of differentiation between steroidal and triterpene saponins, as well as between the different classes of triterpene saponin backbones. All types of the aglycone backbones then undergo a series of distinct oxygenation and glycosylation reactions catalyzed by different tailoring enzymes to produce the finished saponin product (Augustin et al., 2011; Moses et al., 2014a; Thimmappa et al., 2014).

The most extensively studied property of saponins is their ability to permeabilize membranes, the molecular basis for their hemolytic activity (Augustin et al., 2011). The primary model for saponin-membrane interaction involves the spontaneous incorporation of saponins into the outer layer of the bilayer membrane driven by the lipophilic aglycone's affinity for the lipid interior of the membrane. The saponins then form 1:1 complexes with membrane sterols, and the complexes aggregate to form plaques (Augustin et al., 2011; Lorent et al., 2014). The sterical nature of the plaques results in membrane curvature that can lead to the formation of pores that increase membrane permeability (Armah et al., 1999) or vesiculation that carries away sterols (Elias et al., 1978). Research by Lin and Wang, 2009, presented an alternative model of saponin-membrane interaction in which incorporated saponins migrated to sterol-rich lipid rafts to complex with the cholesterol present, instead of creating sterol-rich plaques. Saponin-membrane interaction is influenced by characteristics of both the membrane and the saponins involved. Regarding the membrane, saponin interaction is affected by the presence or absence of cholesterol (Armah et al., 1999; Gögelein and Hüby, 1984), the concentration of cholesterol (Nishikawa et al., 1984; Rosenqvist et al., 1980), and the structure of the cholesterols present (Steel and Drysdale, 1988; Walker et al., 2008). On the saponin, aglycone structure; saccharide side chain number, length, and composition; and the nature of aglycone-sugar and sugar-sugar linkages affects membrane interaction (Moses et al., 2014a). The structures of some saponins, including the oleanane-type triterpenoid soyasaponins in *Medicago truncatula*, do not afford for any hemolytic activity (Yoshiki et al., 1998). Saponins are a chemically diverse family of specialized metabolites produced by many plants, including several important food crops, as a means of plant defense. They are composed of a hydrophobic aglycone backbone and a polar sugar moiety resulting in amphipathic properties that contribute to their ability to permeabilize membranes.

Applications of saponins: industry

Saponins are currently used in many industrial applications for the emulsifying, foaming, and detergent properties provided by their amphipathic character (Güçlü-Üstündağ and Mazza, 2007). *Yucca schidigera* is the most common commercial source of steroidal saponins, while *Quillaja saponaria*, or the soapbark tree, is the most common commercial source of triterpene saponins (Francis et al., 2002). The bark of *Q. saponaria*, a native of central and south Chile, contains 5% triterpenoid saponins, and a

conservative estimate put the annual market value of the quillaja extract industry at 7 million USD in 1999. The indigenous people of Chile were the first to recognize the benefits of quillaja bark, placing it in boiling water and soaking it overnight before using the extract to wash their hair and clothes (San Martín and Briones, 1999). Quillaja extract toxicology studies in rats found no ill effect from consuming even up to 1.5 g/kg/day over 2 years (Drake et al., 1982), and quillaja extracts, as well as a saponin extracts from a small group of plants, have been approved for use as food additives in the United States, Japan, and the European Union (Güçlü-Üstündağ and Mazza, 2007). Saponin extracts are added to carbonated beverages because of their foaming properties, while their emulsifying properties are exploited in the production of mayonnaise, whipping cream, soy sauce, and candies with oil-soluble flavorings (San Martín and Briones, 1999). Saponins, which are generally considered to be bitter (Price et al., 1985) but can be sweet (Kennelly et al., 1996), are also used as flavoring agents. The saponin glycyrrhizin is thought to be primarily responsible for the sweetness of licorice, which is regularly added to candy, chewing gum, beverages, baked foods, and herbs and seasonings (Güçlü-Üstündağ and Mazza, 2007). There has been interest in the use of oleanolic acid as a flavoring agent in diet beverages to reduce aftertaste and enhance sweetness (Kang et al., 1999). Because of their varied biological activities and low toxicity, saponins could also be used as antimicrobials or anti-yeast agents in food. Finally, saponins have been used in the removal of cholesterol from dairy products (Richardson and Jimenez-Flores, 1994) and may be used in the elimination of fat globule membranes from whey protein concentrates to improve their quality for use in processed foods in the future (Hwang and Damodaran, 1994).

Saponins are also used as additives in the cosmetics industry. Due to their surface-active and detergent properties, they have been added to many different personal cleansing agents including liquid soap, toothpaste, mouth wash, shower gel, lotions, and foam bath (Güçlü-Üstündağ and Mazza, 2007). Quillaja extracts have been shown to be compatible with many of the chemicals commonly used in shampoos including sodium laurel sulfate, methylparaben, glucamate, and propylparaben and are already widely used as secondary surfactants in many shampoos in Chile (San Martín and Briones, 1999). Saponins have also been marketed as active ingredients in cosmetic products that claimed to delay aging of the skin and prevent acne. The ability of saponins to form micelles in aqueous solution can increase the solubility of other compounds, and this property could be exploited to increase the solubility of

6

important ingredients in cosmetic, food, and pharmaceutical products (Güçlü-Üstündağ and Mazza,

2007). Saponins have also been utilized in several other miscellaneous industrial applications, including

as a wetting agent in the production of film, as a wastewater treatment in Japan to enhance oxygen

transfer and fat degradation, as a component of light weight composites used in construction, and as a

surfactant useful for washing soil contaminated with naphthalene and hexachlorobenzene (Roy et al.,

1997; San Martín and Briones, 1999). Saponins, especially from *Y. schidigera* and *Q. saponaria*, are

currently used in many industrial applications, primarily as emulsifying, foaming, detergent, and flavoring

agents in food and cosmetics New uses for saponins continue to be proposed, including as antimicrobials

in food and for increasing solubility of other compounds.

Applications of saponins: agriculture

Interest in exploiting saponins in agriculture, to improve both the cultivation of crops and of

livestock, has been growing due to their biological activity. While the use of synthetic chemicals is still the

primary method of protecting crops by deterring unwanted pests and competing plants, concerns about

the environment, animal and human health, and decreasing effectiveness due to pest resistance have led

many to seek alternative methods (Tava and Odoardi, 1996). Saponins have been shown to have

insecticidal, nematicidal, allelopathic, fungicidal, antibacterial, antiviral, and mulluscidal properties and

could be used to protect crops by either application of extracts or plant material or by bioengineering

increased or novel saponin production (Augustin et al., 2011; Francis et al., 2002). The incorporation of

partially purified alfalfa (*Medicago sativa*) saponin mixtures into an artificial diet was able to increase

mortality for the European grape moth (*Lobesia botrana*) and the summer fruit tortrix moth (*Adoxophyes

orana*) and inhibit larval growth and development for European corn borer (*Ostrinia nubilalis*), *Tenebrio

molitor*, and *Spodoptera littoralis* larvae (Tava and Odoardi, 1996), while saponin preparations from a

variety of legume seeds were able to inhibit development of the Azuki bean beetle (*Callosobruchus

chinensis*) when ingested (Applebaum et al., 1969). Soil amendment with saponin-rich dried *M. sativa*

plant material was able to significantly reduce populations of root-knot nematode (*Meloidogyne incognita*),

potato cyst nematode (*Globodera rostochiensis*), and carrot cyst nematode (*Heterodera carotae*) infesting

tomato, potato, and carrot plants, respectively, in both potted plant and field trials (D'Addabbo et al.,

2011). *M. sativa* medicagenic acid glycosides can act as either plant growth inhibitors or stimulators

7

depending on whether they are present at a high or low concentration, respectively (D'Addabbo et al., 2011). Saponins extracted from the roots of *M. sativa* were able to significantly inhibit mycelium growth of the fungal plant pathogens *Alternaria zinniae, Botrytis cinerea, Botrytis tulipae, Phoma narcissi, Phoma poolensis,* and *Rhizoctonia solani in vitro* and inhibited development of *Phoma narcissi* on *Hippeastrum* leaves without causing leaf damage when administered as a preinoculation spray (Saniewska et al., 2013).

Similar to the increasingly negative public opinions on the use of synthetic chemicals in crop cultivation, the use of antibiotics as feed additives to enhance livestock performance has come under increasing scrutiny due to fears of generating antibiotic resistant pathogenic bacteria. Plant-derived products, such as saponins, show promise as possible alternatives along with other non-antibiotic growth promoters like organic acids and probiotics (Windisch et al., 2008). In fact, Nutrafito Plus, a proprietary blend of saponin-rich yucca and quillaja plant powders, was able to improve growth, feed conversion, and other performance parameters in broiler chickens to a similar level as seen with antibiotic growth promoters (Cheeke, 2009). However, due to the chemical diversity of saponins and their varying biological activities in the many bodily environments of different animal species, research on the use of saponins as feed additives presents a complicated story, showing both positive and negative effects of different saponins on livestock.

One positive effect saponins have on ruminant livestock is defaunation, or the removal of protozoa through interaction with their membrane sterols, which can lead to an increase in microbial efficiency in the rumen (Cheeke, 1996). Ammonia-binding yucca extracts are believed to bind ammonia when levels are high in the rumen and then release ammonia when levels are low, maintaining a constant and adequate supply of ammonia in the rumen that is necessary for microbial metabolism, making nitrogen utilization more efficient, and lowering ammonia levels in waste, which would otherwise contribute to groundwater pollution and greenhouse gases (Cheeke, 2000; Francis et al., 2002; Pen et al., 2006). Yucca saponins were also found to selectively inhibit certain bacterial species and not others in the rumen, opening up the possibility of using different saponins to design the microbial population of the rumen to maximize animal performance (Francis et al., 2002). There has been research into the use of yucca extracts to decrease production of methane, a source of energy loss to the animal and contributor

to greenhouse gas emissions, although reports are mixed. Pen et al. (2006) found that addition of yucca extract was able to significantly reduce methane production by up to 42% *in vitro*, while Holtshausen et al. (2009) found that feeding dairy cows a low enough amount of yucca plant powder to eliminate possible negative effects resulted in a loss of the positive effect of methane reduction as well. Oral administration of saponins may stimulate animal immune systems and strengthen disease resistance; feeding sows yucca extract in late pregnancy resulted in reduced neonatal pig mortality (Cheeke, 2000). Diets supplemented with sarsaponin resulted in increased growth in steers and increased growth, feed efficiency, and intake for pigs (Mader and Brumm, 1987).

The most basic argument against the use of saponins as feed additives is that they are unpalatable to many animals because of their bitter taste. Pigs and various avian species, including turkey and chicken, have shown an aversion to feed containing even very low levels of high-saponin alfalfa meal, and cattle have often shown aversions to particular batches of alfalfa hay, presumably because of higher levels of saponins (Cheeke, 1996). This aversion can lead to reduced feed intake and has led some to call for the development of saponin-free alfalfa (Dixon and Sumner, 2003). Saponins have also been blamed for the occurrence of ruminant bloat, although this has yet to be supported by clear experimental proof (Francis et al., 2002). Early research claimed that saponins reduced surface tension due to their surface-active properties, leading to foam formation, and reduced ruminal motility, limited the rumen's ability to clear the foam produced (Milgate and Roberts, 1995). However, later work failed to find a connection between the occurrence of bloat in cattle and whether they were fed a high or low saponin alfalfa diet (Cheeke, 1996). Certain saponins have also been shown to interfere with small intestinal mucosal cells *in vitro*, increasing their permeability and inhibiting active transport (Johnson et al., 1986). This could lead to the uptake of certain macromolecules, such as allergens, that would normally be excluded and increase the turnover rate of intestinal mucosal cells, a source of energy and protein loss that could explain the growth-decreasing effects of saponins some have reported (Cheeke, 1996; Francis et al., 2002). Saponins, poorly absorbed themselves due to their binding with the intestinal mucosa, have also been reported to reduce absorption of different micronutrients, including vitamin A and E, sodium, iron, magnesium, calcium, and zinc in the diet of different animals (Cheeke, 1996; Francis et al., 2002). Finally, steroidal saponins have been implicated in livestock development of

photosensitization, a condition that occurs when UV light reacts with photodynamic compounds in the blood and produces free radicals that react with proteins and nucleic acids, causing severe skin lesions. The saponins are conjugated with glucuronic acid in the liver and excreted into the bile as insoluble calcium salt crystals that damage the bile ducts and disrupt bile excretion, resulting in the retention of phylloerythrin, a photodynamic compound, in the blood. Often, the sources of the saponins that cause this problem are weeds, including *Tribulus terrestris, Narthecium ossifragum*, and *Agave lechuguilla*, and not part of the animal's typical diet (Cheeke, 1996). Saponins show activity against a variety of plant pathogens, including insects, nematodes, and fungus, and could be used to protect crops either by application of extracts or plant material. Saponins could also be used as feed additives to improve livestock performance, but the existence of studies reporting both positive and negative effects of saponins suggests that additional research is needed.

Applications of saponins: human health

Research into the use of saponins to treat and prevent various human ailments is widespread due to their many pharmaceutically significant properties, most of which are very specific to saponin type and structure. However, saponins are not new to medicine. Various cultures have used saponin-rich plants such as yucca, ginseng, horse chestnut, licorice, and sarsaparilla for centuries to treat poor health; over 85% of the herbs used in traditional Chinese medicine contain significant amounts of saponins (Güçlü-Üstündağ and Mazza, 2007). Saponins are able to lower cholesterol by both interacting with cholesterol in the small intestine to form insoluble complexes that prevent absorption and forming mixed micelles with bile acids that increase their excretion and necessitate their replacement by hepatic synthesis from cholesterol (Sidhu and Oakenfull, 1986). This effect has been shown to selectively lower the more harmful LDL cholesterol in the serum of gerbils, rats, and human subjects (Francis et al., 2002) without changing the concentration of the healthier HDL cholesterol (Bora and Sharma, 2011). A product made available in the United States called Cholestaid™ containing 900 mg of saponin-rich *M. sativa* extract with 100 mg citric acid claims to "neutralize" cholesterol in the stomach and facilitate its excretion (Bora and Sharma, 2011). Saponins can act as adjuvants, compounds that are used in combination with a specific antigen to increase immunity to that antigen, and immunostimulants, substances that are used to induce a more generalized resistance to pathogens or tumors (Lacaille-Dubois, 2007; Sun et al., 2009).

Certain saponins in the form of both purified compounds and crude mixtures have been shown to induce specific cytotoxic T-lymphocyte responses, increase immune cell proliferation, and enhance antibody production (Francis et al., 2002). There is particular interest in using saponins as adjuvants in orally administered vaccines due to their ability to increase the permeability of intestinal cells, facilitating the uptake of antigens (Francis et al., 2002; Güçlü-Üstündağ and Mazza, 2007). QuilA, isolated from *Q. saponaria* bark, and its derivative QS-21 are the foremost saponin adjuvant candidates and have been used in formulations of vaccines in Phase I and II human trials for HIV, cancer, and malaria and in experimental vaccines for respiratory syncytial virus, cytomegalovirus, *Toxoplasma gondii*, and visceral leishmaniasis in mice (Lacaille-Dubois, 2007; Sun et al., 2009). Saponins have been reported to be hypoglycemic by a variety of mechanisms. Oleanolic acid glycosides inhibited gastric emptying, a process that occurs faster in type I and II diabetic patients and contributes to higher glucose levels, and glucose uptake at the small intestine brush border when added to the nutrient meal diet of mice (Matsuda et al., 1999). However, a saponin-rich fenugreek seed extract fed to rats decreased blood glucose levels by increasing plasma insulin levels, presumably by stimulating β-cells (Petit et al., 1993). Particular saponins may have useful neuroprotective properties. Rg1, a *Panax ginseng* saponin, was able to reduce neuronal death through a variety of mechanisms after a hypoxic-ischemic type insult *in vitro* and could be used to protect the brain after an ischemic cerebral stroke (Zhang et al., 2008). Ginseng saponins, often called ginsenosides, induced nitric oxide production in the brain and inhibited stress-induced hypothalamo-pituitary-adrenal responses (Francis et al., 2002), while the purified ginsenoside Rb1 can minimize the toxic effects of β-amyloid peptides present in Alzheimer's disease (Lacaille-Dubois, 2007). Saponins have exhibited biological activity against important bacterial, viral, and fungal human pathogens, including *Staphylococcus aureus*, human immunodeficiency virus (HIV), herpes simplex virus (HSV), influenza virus, and *Candida albicans* (Bora and Sharma, 2011; Coleman et al., 2010; Lacaille-Dubois, 2007). Glycyrrhizin from licorice roots actively inhibits replication of severe acute respiratory syndrome (SARS) virus (Cinatl et al., 2003), and certain steroid saponins from *T. terrestis* are very active against fluconazole-resistant *Candida* species fungal pathogens (Zhang et al., 2005). In contrast to their stimulatory effect on specific immune responses, saponins have an inhibitory effect on inflammation, a non-specific immune reaction (Francis et al., 2002). Oral administration of AR-6, a triterpene saponin

isolated from *Clematis chinensis Osbeck* root, to rats with an induced rheumatoid-like arthritis resulted in a clear decrease of clinical signs of the disease (Sun et al., 2010). Saponins from various sources have been reported to prevent liver damage through a variety of mechanisms *in vitro* and *in vivo* (Lacaille-Dubois, 2007). Oleanolic acid has even been used as an oral drug in China to treat acute and chronic hepatitis, as well as other liver disorders, by decreasing the occurrence of cirrhosis and returning elevated aminotransferase activity in the serum to normal levels (Liu, 1995). Saponins in crude mixtures and as purified compounds have been shown to have a variety of potentially therapeutic effects on the cardiovascular system. *M. sativa* aerial saponins were able to prevent atherosclerosis, the buildup of plaque in arteries, and induce regression of existing athersclerosis in rats (Bora and Sharma, 2011). The ginsenosides Rb1 and Re have been shown to inhibit contraction of ventricular myocytes from adult rat hearts *in vitro* by the stimulation of nitric oxide production, a property that could be used in the treatment of patients with hypertension (Scott et al., 2001; Sung et al., 2000). A mixture of saponins from *Herniaria glabra* decreased systolic and diastolic blood pressure in rats, while certain nonhemolytic steroidal saponins from *Anemarrhena asphodeloides* strongly inhibit platelet aggregation (Lacaille-Dubois, 2007). Saponins exhibit a variety of properties beneficial to human health that have led to the current and possible future use of saponins in medicine and are specific to saponin type and structure, including hypocholesterolemic, immunostimulatory, hypoglycemic, neuroprotective, anti-inflammatory, and hepatoprotective activities as well as therapeutic effects on the cardiovascular system and activity against viral, fungal, and bacterial human pathogens.

Cytotoxicity of saponins

In addition to the previously described health benefits of saponins and most pertinent to this research, saponins from a variety of sources have been shown to have cytotoxic effects on many different cancer cell types in both *in vitro* and *in vivo* experiments. While the most obvious mechanism for these effects would seem to be membrane permeabilization, a non-specific process that would lead to inflammatory necrosis, this is fortunately not the case. Saponins without hemolytic activity have been shown to have cytotoxic activity (Mimaki et al., 1998), and a study of 63 different steroidal saponins found no correlation between hemolytic and cytotoxic activities (Wang et al., 2007). In addition, numerous studies of the cytotoxic activities of different triterpene and steroidal saponins identified a variety of

mechanisms besides membrane permeabilization (Bachran et al., 2008; Podolak et al., 2010; Weng et al., 2011). The most common and best studied effect of saponins on cancer cells is induction of apoptosis, or programmed cell death. Apoptosis can be initiated through either the extrinsic pathway, triggered at the plasma membrane by activation of specific receptors, or the intrinsic pathway, initiated within the cell in response to stress. In a recent review of the cytotoxic activity of saponins, there was only one example of the induction of the extrinsic pathway of apoptosis, by the saponin fraction from *Bupleurum kaoi*; instead, most saponins induce the intrinsic pathway of apoptosis through different mechanisms, including activation of p53 and various caspases, inhibition of the anti-apoptotic Bcl-2 protein and activation of the pro-apoptotic Bax protein, induction of cytochrome c release, and mitochondrial membrane permeabilization (Bachran et al., 2008; Podolak et al., 2010). Avicin G isolated from *Acacia victoriae* (Bentham) induced apoptosis in the Jurkat human T cell leukemia cell line through activation of caspase-3 and initiation of cytochrome c release by direct action on the mitochondrial membrane (Haridas et al., 2001). Another common effect of saponins on cancer cells is cell cycle arrest through the downregulation of cyclin and cyclin dependent kinase (CDK) proteins and upregulation of CDK inhibitors (Bachran et al., 2008). The ginsenoside Rg3 has been shown to decrease proliferation of the LNCaP human prostate carcinoma cell line through activation of the expression of cyclin kinase inhibitors p21 and p27, arresting the cells in G1 phase (Liu et al., 2000). Other less frequent cytotoxic effects of saponins include initiation of autophagic cell death, when stressed cells break down their own contents to produce energy, inhibition of angiogenesis, cytoskeleton disintegration, and inhibition of metastasis (Podolak et al., 2010).

Saikosaponins, ginsenosides, soyasaponins, julibrosides, saikosaponins, dioscins, and avicins are among the most studied cytotoxic saponin groups, while oleanolic acid, the soyasapogenols, diosgenin, and betulinic acid are among the most studied cytotoxic sapogenins (Bachran et al., 2008; Güçlü-Üstündağ and Mazza, 2007). While it seems evident that saponin structure affects cytotoxic activity, there is not a sufficient amount of research to make generalizations about specific aspects of the structure-activity relationship (Weng et al., 2011). However, interesting studies do exist that begin to elucidate the structure-activity relationship in specific situations. In a study on the effects of various soyasaponins and soyasapogenols on the proliferation of HT-29 colon cancer cells, the more lipophilic

soyasapogenols were the most bioactive (Gurfinkel and Rao, 2003). These results were confirmed in a study by Salyer et al. (2013) that found that suppression of Caco-2 colon cancer cells intensified with the increasing lipophilicity of the two soyasaponins and one soyasapogenol tested. Gauthier et al. (2009) tested the cytotoxic activities of 31 natural and synthetic triterpene oleanane and lupane-type saponins and found that for oleanane-type saponins, an oleanolic acid backbone resulted in greater activity against A549 lung carcinoma cells *in vitro* than a hederagenin backbone. This finding was confirmed in a study by Bang et al. (2005), in which saponins with an oleanolic acid backbone showed more cytotoxic activity against four different cancer cell lines *in vitro* than saponins with hederagenin backbones. However, when the same saponins were used to treat BDF1 mice with Lewis lung carcinoma *in vivo*, saponins with hederagenin backbones had more potent antitumor activity than saponins with oleanolic acid backbones (Bang et al., 2005). This complicating lack of correlation between *in vitro* and *in vivo* data has been shown for multiple saponins and in this case could be due to differences in bioavailability (Podolak et al., 2010); Weng et al., 2011). A greater understanding of the relationship between saponin structure and cytotoxic activity would be useful in the synthetic modification of saponins to maximize their effects.

Whereas the majority of saponin cytotoxicity studies have been done *in vitro*, the number of *in vivo* studies, most of which are conducted in mice, is growing. Oral administration of playcodin D, purified from *Platycodi radix,* to athymic nude mice bearing H520 human lung cancer cells resulted in decreased tumor volumes and increased tumor cell apoptosis, as well as stimulation of the immune system and amelioration of cancer-related cachexia, or weakness (Park et al., 2014). *Rhizoma paridis* saponins delivered via intragastric administration to mice injected with mouse lung adenocarcinoma cells inhibited tumor growth by inducing apoptosis, downregulating levels of MMP-2 and MMP-9, and upregulating expression of the MMP inhibitor TIMP-2 (Man et al., 2009). Oral administration of deltonin from *Dioscorea zingiberensis* Wright to mice with C26 murine colon cancer resulted in inhibition of tumor growth through apoptosis induction and inhibition of angiogenesis (Tong et al., 2011). The safety and tolerability of individual saponins has not been extensively studied, but research on their effects in rodents and dogs generally suggests they are acceptably tolerable. In a subchronic toxicity study of the ginsenoside metabolite protopanaxadiol done for 90 days in beagles, the no observed adverse effect level (NOAEL) was 6.7 mg/day when given intravenously, which corresponds with high toxicological safety (Weng et al.,

2011). As with their bioactivity, the safety of individual saponins is likely to vary based on their structure. The most obvious point of concern when considering the use of saponins as chemotherapeutic drugs is their hemolytic activity. It is important to note though that as mentioned earlier, some cytotoxic saponins do not have hemolytic properties, and even for some saponins that exhibit both properties the levels necessary for cytotoxic activity are lower than the levels necessary for hemolysis (Mimaki et al., 1998). As more is understood about the structure-activity relationship for both cytotoxicity and hemolytic activity, it may become feasible to modify highly cytotoxic saponins to abolish their hemolytic activity while keeping their cytotoxicity intact.

In addition to their possible use as individual chemotherapeutic agents, saponins could be used in combination with other cancer treatments to increase their efficacy. Various saponins have additive or even synergistic effects on the efficacy of other chemotherapeutic agents (Bachran et al., 2008). The ginsenoside Rh2 on its own had no effect on HRA human ovarian tumor growth in mice, but a combination of cisplatin and Rh2 prolonged survival of the mice significantly more than treatment of cisplatin or Rh2 alone, indicating synergistic effects (Kikuchi et al., 1991). Saponins may also play a role in the emerging field of targeted therapy, in which a toxin is attached to antibodies, cytokines, or growth factors that target the toxin to cancer cells, decreasing the negative side effects so common in non-targeted chemotherapy (Thakur et al., 2013). In a study performed using HCT116 human colon carcinoma tumor-bearing mice, treatment with a combination of the plant ribosome-inactivating protein dianthin attached to epidermal growth factor and the saponin SO-1861 resulted in a greater than 95% tumor volume reduction (von Mallinckrodt et al., 2014). The synergistic effect of saponins on other chemotherapeutics and targeted toxins is likely due to effects on cellular transport processes that could facilitate more efficient delivery of these compounds to their cellular destination (Weng et al., 2011). Saponins from various plant sources have been shown to have cytotoxic activity against many different cancer cell types *in vitro* and *in vivo* through structure-specific targeting of various cellular processes and not through non-specific membrane permeabilization. They have also been shown to improve the efficacy of targeted toxins and other cytotoxic agents.

Medicago truncatula

Legumes are grown on 180 million hectares or 12-15% of the earth's arable land and make up 27% of the world's primary crop production, supplying 33% of the dietary protein nitrogen requirements of humans (Graham and Vance, 2003). Legumes are set apart from other plants by their ability to fix nitrogen from the environment, reducing N_2 to NH_3, through a symbiotic relationship with rhizobia bacteria in the soil that takes place in root nodules (Benedito et al., 2008). As a result, legumes do not require application of nitrogen fertilizer and can be used in crop rotation to increase soil nitrogen for use by other plants, leading to a decreased use of expensive and environmentally harmful nitrogen fertilizers. The primary dietary legumes include soybeans, peas, chickpeas, peanut, and lentils (Graham and Vance, 2003). In addition to being sustainable sources of human protein, legumes are also used as forage for livestock, raw materials for industrial applications, and increasingly as natural sources of secondary metabolites (Choi et al., 2004b; Gholami et al., 2014). *Medicago sativa*, or alfalfa, is the most cultivated forage legume in the world and the fourth most economically valuable crop in North America, grown for soil improvement, animal forage, human food (as sprouts), and medicinal uses (Bora and Sharma, 2011; Gholami et al., 2014). Unfortunately, widely cultivated legumes in general, with their large, complex genomes, and alfalfa in particular, a tetraploid and obligate out-crosser, are not easily used in molecular genetic studies (Gholami et al., 2014; Tadege et al., 2005). In their place, *Medicago truncatula*, a close relative of alfalfa, has been widely studied as model legume due to its small diploid genome of about 500 Mbp, self-fertilization and prolific nature, easy genetic transformation, and rapid generation time (Huhman and Sumner, 2002). Work in *M. truncatula* can be translated to the more economically valuable legume species due to genetic similarity; *M. truncatula* genes share very high sequence identity with corresponding alfalfa genes and seem to have similar chromosome placement as other legumes (Kapusta et al., 2005b).

The barrel medic, *M. truncatula*, originated in the Mediterranean basin, where diverse native populations can still be found in the wild, and has been cultivated in Australia as winter forage and in rotation with wheat for the last century (Choi et al., 2004a). In addition to the natural traits that make it an attractive model, the scientific community has developed many tools to facilitate genetic studies of *M. truncatula*, including a fully sequenced genome, almost 270,000 expressed sequence tags (ESTs) in

Genbank, a metabolic pathway map called MedicCyc, and a collection of mutants produced by various methods, including the use of the *Tnt1* tobacco retrotransposon (Gholami et al., 2014; Tadege et al., 2005). Genetic maps of molecular markers based on anonymous sequence polymorphisms (Thoquet et al., 2002) and sequence-characterized loci have been created (Choi et al., 2004a), as well as a comprehensive gene expression atlas with information on all the major organ systems and multiple stages of seed and nodule development (Benedito et al., 2008). For all the genetic work done in *M. truncatula*, the biological activity of its saponins towards cancer cells has not previously been studied (Podolak et al., 2010). Four genetically distinct accessions of *M. truncatula* that vary in the level and types of saponins produced were used in this project, A17, ESP105, PRT178, and GRC43. All are single seed descent lines chosen from a collection of 145 geographically diverse populations sampled primarily from the western part of the Mediterranean basin and maintained by the Institut National de la recherché Agronomique (INRA) in France (Delalande et al., 2007). A17, the primary focus of previous genetic research, is derived from Jemalong, a commercial variety of *M. truncatula*, and reflects the efforts of selective breeding, while the other three accessions, ESP105, PRT178, and GRC43, were isolated from the wild in Spain, Portugal, and Greece, respectively, and reflect adaptation to the selective pressures of their unique environments. Legumes are economically important due to their ability to fix nitrogen and serve as a sustainable source of protein for humans and livestock, and *M. truncatula* is widely used as a model legume in scientific research. The availability of genetically distinct accessions with varying levels of saponin accumulation and the genetic tools that have been developed for *M. truncatula* make it an ideal organism for research on saponin biosynthesis.

Triterpene saponin biosynthesis in *M. truncatula*

More than 30 triterpene saponins have been detected in *M. truncatula* (Huhman and Sumner, 2002). As previously described, triterpene saponin synthesis originates in the MVA pathway with the conversion of acetyl-CoA into IPP and DMAPP, which then condense to form FPP. Two FPP molecules condense to form squalene, which is then epoxidized to form 2,3-oxidosqualene (Moses et al., 2014a). 3-hydroxy-3methylglutarylcoenzyme A reductase (HMGR) is the key rate-limiting enzyme in the MVA pathway. Five isoforms of HMGR have been characterized in *M. truncatula*, and HMGR gene expression is induced by methyl jasmonate in a pattern similar to other triterpene saponin synthesis genes in *M.*

truncatula (Kevei et al., 2007; Pollier et al., 2013). One squalene synthase and two squalene epoxidase gene candidates were identified in the *M. truncatula* genome by comprehensive gene expression clustering analysis (Naoumkina et al., 2010) and then functionally characterized by expression in *Escherichia coli* and complementation of the yeast *erg1* mutant, respectively (Suzuki et al., 2002). In *M. truncatula* the first committed step in triterpene saponin synthesis is the cyclization of the linear 2,3-oxidosqualene by β-amyrin synthase to form β-amyrin. All saponins in *M. truncatula* are based on the modification of β-amyrin and have pentacyclic oleanane-type sapogenin backbones with a double bond between C-12 and C-13 and an oxygen at the C-3 position (Gholami et al., 2014). The β-amyrin synthase gene has been characterized by expression in yeast by two separate research groups (Iturbe-Ormaetxe et al., 2003; Suzuki et al., 2002). An alternate fate for 2,3-oxidosqualene is synthesis of cycloartenol, the tetracyclic plant sterol precursor, by cycloartenol synthase (Gholami et al., 2014).

After formation of β-amyrin, the synthesis pathway branches to create two distinct classes of triterpene saponins, hemolytic and non-hemolytic. The non-hemolytic sapogenins (soysapogenol A, B, and E) are characterized by oxidation at the C-24 position that is compatible with modifications at the C-21 and C-22 positions, while the hemolytic sapogenins (oleanolic acid, hederagenin, bayogenin, medicagenic acid, and zanhic acid) are characterized by a carboxyl group at the C-28 position that is compatible with C-23 modification but prohibits C-24 modification (Moses et al., 2014c). In non-hemolytic saponin synthesis, CYP93E2 carries out the hydroxylation at C-24, which is followed by oxidation at the C-22 position catalyzed by CYP72A61v2 (Fukushima et al., 2013). This yields soyasapogenol B which can then be further modified to form soyasapogenol A, characterized by oxidation at the C-21 position, or soysapogenol E, characterized by further oxidation at the C-22 position. The enzymes responsible for these conversions have not yet been identified in *M. truncatula* (Gholami et al., 2014). In hemolytic saponin synthesis, CYP716A12 catalyzes three consecutive oxidations at the C-28 position of β-amyrin to yield oleanolic acid (Carelli et al., 2011; Fukushima et al., 2011). Carelli et al. (2011) showed that *lha* (lacking hemolytic activity) mutant plants with disrupted CYP716A12 genes were unable to produce hemolytic saponins, synthesizing only soysaponins. Oleanolic acid can then be converted into alternate hemolytic sapogenin backbones through a series of sequential reactions catalyzed by cytochrome P450 enzymes. CYP72A68 expressed in yeast in combination with CYP716A12 exhibited oxidation activity at

the C-23 position of oleanolic acid but was unable to oxidize β-amyrin when expressed in yeast alone

(Fukushima et al., 2013). Two CYP72A68 isoforms were identified in the *M. truncatula* genome by Biazzi

et al. (2015) and shown to preferentially transform substrates that had previously been oxidized at the C-2

position (2β-hydroxy oleanolic acid and bayogenin) over substrates lacking C-2 oxidation (hederagenin

and oleanolic acid). CYP72A67 expressed in yeast was shown to catalyze this C-2 oxidation and to do so

most efficiently when the substrate lacked C-23 and C-16 oxidation (Biazzi et al., 2015). Zanhic acid, a

sapogenin detected in *M. truncatula*, is oxidized at the C-16 position, in addition to the C-23 and C-2

position. The enzyme responsible for this oxidation has not yet been identified in *M. truncatula.* However,

CYP87D16 and CYP716Y1 have been shown to have C-16 oxidase activity in triterpene saponin

synthesis in *Maesa lanceolata* and *Bupleurum falcatum*, respectively (Moses et al., 2014b; Moses et al.,

2015). This data, especially the recent Biazzi et al. (2015) study, was collectively used to construct the

hypothetical *M. truncatula* sapogenin biosynthesis pathway depicted in Figure 2.
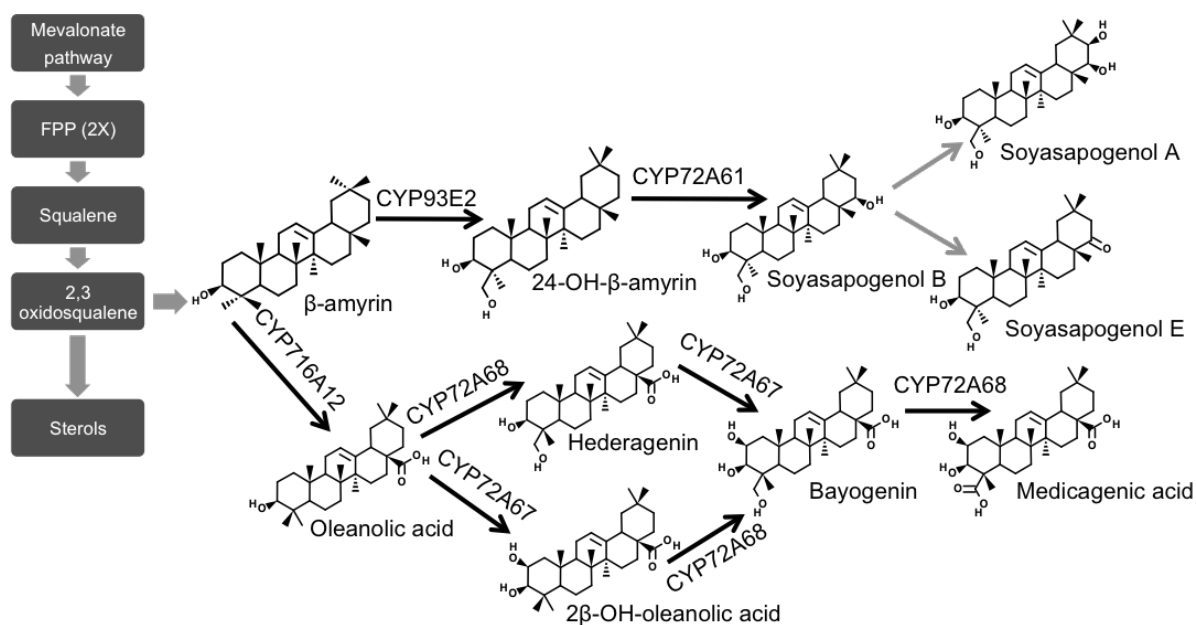
Figure 2. Saponin aglycone biosynthesis in *M. truncatula*. Hypothetical pathway shown is according to Biazzi et al. (2015).

Once oxidative modification of the sapogenin backbone is complete, it is glycosylated at different

positions with varying sugar moieties by UDP-dependent glycosyltransferases (UGTs) that are still largely

unknown. In *Medicago* species, glycosylation usually occurs at the C-3 hydroxy and C-28 carboxy groups, and the attached sugar chains have been reported to include glucose, rhamnose, xylose, glucuronic acid, galactose, and arabinose (Gholami et al., 2014). UGT73K1 and UGT71G1 have been functionally characterized as UDP-dependent glucosyltransferases active in *M. truncatula* saponin synthesis with specificity for hederagenin, and soyasapogenols B and E (UGT73K1) and medicagenic acid (UGT71G1) (Achnine et al., 2005). Plants with a UGT73F3 loss-of-function mutation were found to have decreased levels of glycosylated saponins derived from hederagenin, bayogenin, medicagenic acid, and soyasapogenol E backbones (Naoumkina et al., 2010).

The differential accumulation of varying saponin mixtures across plant tissues and in response to herbivory suggests the existence of specific regulators of saponin biosynthesis. Stress-induced saponin biosynthesis is mediated by concerted transcriptional activation of synthesis genes controlled by a complex signaling cascade involving jasmonates (JAs), oxylipin-derived phytohormones that control a variety of plant physiological processes (De Geyter et al., 2012; Gholami et al., 2014; Mertens et al., 2016). *M. truncatula* cell culture suspensions exposed to methyl jasmonate show a 50-fold induction of β-amyrin transcripts and increased synthesis of over 30 different triterpene saponins (Broeckling et al., 2005; Suzuki et al., 2004). By looking at the transcriptome of *M. truncatula* suspension-cultured cells after elicitation with JAs, Pollier et al. (2013) were able to identify the RING E3 ubiquitin ligase makibishi 1 (Mkb1) gene, which encodes a protein involved in the endoplasmic-reticulum-associated degradation (ERAD) quality control system, as being transcriptionally upregulated before similar upregulation of saponin synthesis genes. MKB1 manages saponin synthesis by targeting HMGR, the rate-limiting enzyme in the MVA pathway, for degradation by the 26S proteasome, and prevents unrestrained bioactive saponin accumulation that could threaten plant development and integrity (Pollier et al., 2013). Until very recently nothing was known about transcription factors that might mediate the effect that jasmonate signaling has on regulation of saponin biosynthesis. Two JA-inducible transcription factors from the basic helix-loop-helix family, TSAR1 and TSAR2 have recently been shown to transactivate the genes encoding HMGR1 and MKB1, as well as other downstream saponin biosynthesis genes. Overexpression of *TSAR1* and *TSAR2* in *M. truncatula* hairy roots results in increased transcript levels of saponin biosynthesis genes and saponin accumulation, with TSAR1 overexpression specifically

increasing nonhemolytic saponin biosynthesis and TSAR2 overexpression specifically increasing

hemolytic saponin biosynthesis (Mertens et al., 2016). Saponin synthesis originates in the mevalonate

pathway, and production of β-amyrin is the first commited step. Cytochrome P450 enzymes catalyze

various oxidative modifications of β-amyrin to form several chemically distinct sapogenins belonging to

two groups, non-hemolytic soyasaponins and hemolytic saponins. Saccharide sugar chains are then

attached to sapogenins by UDP-dependent glycosyltransferases. While many of the genes encoding

enzymes in the early stages of saponin synthesis have been identified, there is still much to be

discovered, including the genes encoding most of the UDP-dependent glycosyltransferases and the

mechanisms of saponin biosynthesis regulation.

Caco-2 cell culture

The Caco-2 cell culture line was originally derived from colorectal adenocarcinoma cells obtained

from a primary tumor in the colon of a 72-year-old Caucasian male (Lindl and Steubing, 2013). Upon

reaching confluence in culture, Caco-2 cells undergo a spontaneous differentiation pathway and begin to

express many characteristics of small intestine enterocytes (Sambuy et al., 2005). As a result, Caco-2

monolayers are frequently used in research as an *in vitro* model for intestinal drug absorption (Artursson

and Karlsson, 1991). Caco-2 cells form moderately well differentiated adenocarcinoma tumors that

resemble primary colonic tumors when injected in nude mice and can be used in cell culture to test the

cytotoxicity of certain compounds against colon cancer cells if they are not permitted to grow to

confluence and differentiate (Lindl and Steubing, 2013). Saponins from *Gaultheria trichophylla,*

Pleiocarpa pycnantha, and *Glycine max* have all been shown to inhibit Caco-2 cell proliferation (Alam et

al., 2015; Omoyeni et al., 2015; Salyer et al., 2013). Colorectal cancer is the third most common cancer

and the third leading cause of cancer death for both men and women in the United States and among the

most common cancers worldwide (Omoyeni et al., 2015). Colorectal cancer risk increases with age, with

90% of new cases and 94% of deaths occurring in people over 50 years of age (von Mallinckrodt et al.,

2014). Surgical removal is the preferred treatment for localized disease, but many colorectal cancer

patients have metastasis by the time they are diagnosed. Treatment in this situation is mainly based on 5-

fluorouracil and folinate chemotherapy in combination with radiotherapy. However, standard treatment

protocols result in only a moderate decline in mortality, and risk of disease recurrence remains high (Tong

et al., 2011). Caco-2 cells grown in culture will be used to study the effects *M. truncatula* saponins on cell proliferation and apoptosis induction in human colon cancer cells.

Concluding statements

When attacked by pathogens, plants defend themselves by specialized metabolism of a broad range of compounds with varied biological activities that can be exploited by humans to solve many of the problems our world faces today. Saponins are a chemically diverse family of specialized metabolites whose amphipathic character and complex structure imparts them with many properties that have led to their current and possible future use in industry, agriculture, and the improvement of human health. Among these properties is the cytotoxic activity saponins from various plant sources have been shown to exercise against different cancer cell types *in vitro* and *in vivo* through structure-specific targeting of various cellular processes. More complete knowledge of the genetic basis for saponin biosynthesis could lead to more extensive utilization of these beneficial compounds, and investigation of the effects of previously untested saponins on cancer cells could add to our understanding of their cytotoxic activity and lead to the development of novel chemotherapeutic drugs. The natural variation in saponin accumulation among genetically distinct accessions of the model legume *M. truncatula* and the genetic tools associated with this species make it an ideal organism for research on the regulation of saponin biosynthesis in plants.

<center>**Objectives**</center>

Hypothesis

Crosses between *M. truncatula* accessions with divergent levels of saponin accumulation will produce populations with large variation in saponin content that can be used to identify the genetic components controlling saponin synthesis, and saponin extracts from *M. truncatula* accessions with diverse saponin profiles will demonstrate varying levels of cytotoxic activity against Caco-2 cells grown in culture.

Objective 1

Compare the CYP72A68 cDNA sequences of *M. truncatula* accessions to examine accession differences in protein functionality and identify single nucleotide polymorphisms that can be used to develop accession-specific molecular markers for confirmation of putative crosses.

Objective 2

Quantify accumulation of hemolytic saponins and expression of genes encoding hemolytic saponin synthesis enzymes in foliar tissue of a *M. truncatula* population to determine the pattern of inheritance of saponin accumulation traits and investigate the mechanisms regulating saponin synthesis.

Objective 3

Analyze crude saponin extracts from foliar tissue of *M. truncatula* accessions by HPLC mass spectrometry to quantify total saponins for treatment of cell cultures with accurate and equal concentrations of saponins across the accessions and identify individual saponins that may contribute to cytotoxic activity.

Objective 4

Evaluate Caco-2 cell proliferation and apoptosis induction after treatment with *M. truncatula* saponin extracts to determine the differential effects of accessions extracts on colon cancer cell growth and the mechanism through which these effects are mediated.

Chapter 2

Genetic basis for triterpene saponin biosynthesis in *Medicago truncatula*

Introduction

Saponins are a family of specialized metabolites produced by a variety of plants that are believed to play a role in plant defense due to their biological activity against many plant pathogens and the ability of jasmonates to elicit their production (Gholami et al., 2014). Saponins are composed of a hydrophobic isoprene-derived aglycone, often called a sapogenin, and a polar sugar moiety. The presence of both components makes saponins amphiphilic, which allows them to create stable foams in aqueous solution, the trait for which they were named; saponin comes *sapo*, the Latin word for soap. There are two types of saponins, steroidal saponins, which have a 27-carbon aglycone, and triterpene saponins, which have a 30-carbon aglycone (Weng et al., 2011). The saponin polar sugar moiety is made up of saccharide side chains containing varying numbers and types of sugar constituents, and saponins can be further classified as mono-, bi-, or tri-desmoside depending on the number of saccharide side chains that are attached to the aglycone (Francis et al., 2002). Saponins are currently used in industry as additives in food and cosmetics due to their ability to act as emulsifying, foaming, and flavoring agents and are utilized in the removal of cholesterol from dairy products. In agriculture, there is growing interest in replacing synthetic chemicals used in crop protection and antibiotics used as livestock feed additives with saponins. Many saponins have been shown to have insecticidal, nematicidal, allelopathic, fungicidal, antibacterial, and antiviral properties that allow them to protect the plants that produce them from pathogen attack (Augustin et al., 2011). Finally, saponins exhibit a variety of properties that could be useful in the improvement of human health, including hypocholesterolemic, hypoglycemic, hepatoprotective, immunostimulatory, anti-inflammatory, and anticancer properties (Lacaille-Dubois, 2007). Unfortunately, the large-scale commercial and medicinal use of saponins is limited by the relatively low levels of saponins produced in nature and the expense associated with chemical synthesis of such complex molecules (Misawa, 2011). Engineering of the saponin synthesis pathway to increase production in plants that endogenously produce saponins and plants that lack saponin production could result in higher yields of commercially and medicinally important saponins and allow for additional pest protection

of economically important crops. However, pathway engineering for increased production of saponins is not yet possible due to gaps in our knowledge of the genetic basis for saponin biosynthesis (Gholami et al., 2014; Misawa, 2011).

The annual legume *M. truncatula* is a close relative of alfalfa grown as winter forage in Australia and is often used in research to study root nodulation and specialized metabolism due to natural characteristics that make it easy to work with and several tools developed by the scientific community that facilitate genetic studies (Choi et al., 2004a; Huhman and Sumner, 2002). Saponin synthesis in *M. truncatula* is restricted to triterpene saponins, of which two types are produced, non-hemolytic soyasaponins and hemolytic saponins. Synthesis originates in the mevalonate pathway with the production of IPP, which combines with DMAPP to form FPP. Two FPP molecules condense to form squalene, which is epoxidized to form 2,3-oxidosqualene that is cyclized to produce β-amyrin. This serves as the branchpoint between the two types of saponin synthesis present in *M. truncatula* as β-amyrin can be funneled into the production of soyasapogenol B or oleanolic acid. Cytochrome P450 enzymes catalyze the oxidative modification of soyasapogenol B and oleanolic acid to produce the full range of non-hemolytic and hemolytic sapogenins, respectively. Glycosyltransferases are responsible for attaching saccharide sugar chains to the sapogenins to yield the completed saponin product (Gholami et al., 2014). Many of the genes encoding enzymes responsible for the various steps of saponin synthesis in *M. truncatula* have been identified and functionally characterized by heterologous expression in microbial systems. Concerning the early steps of saponin synthesis, five isoforms of HMGR, the rate-limiting enzyme in the mevalonate pathway, a squalene synthase gene, and two isoforms of squalene epoxidase have been characterized (Kevei et al., 2007; Suzuki et al., 2002). The enzyme β-amyrin synthase is responsible for cyclization of 2,3-oxidosqualene into β-amyrin, the first committed step in triterpene saponin synthesis (Iturbe-Ormaetxe et al., 2003; Suzuki et al., 2002). In the non-hemolytic soyasapogenol synthesis pathway, CYP93E2 and CYP72A61v2 have been characterized as the enzymes responsible for conversion of β-amyrin into soyasapogenol B by catalyzing oxidation at the C-24 and C-22 positions, respectively (Fukushima et al., 2013). The enzymes responsible for further oxidation of soyasapogenol B at the C-22 position to form soyasapogenol E and oxidation of soyasapogenol B at the C-21 position to form soyasapogenol A have not been identified. In the hemolytic sapogenin synthesis pathway,

CYP716A12 has been characterized as the enzyme catalyzing the conversion of β-amyrin into oleanolic acid via three consecutive oxidations at the C-28 position (Carelli et al., 2011; Fukushima et al., 2011). Oleanolic acid is then converted into alternate sapogenins through a series of sequential reactions catalyzed by CYP72A67, a C-2 oxidase, and CYP72A68, a C-23 oxidase (Biazzi et al., 2015; Fukushima et al., 2013). The enzyme responsible for the final reaction in the series, the oxidation of medicagenic acid at the C-16 position to form zanhic acid, has not yet been identified. Of the many glycosyltransferases likely to be active in *M. truncatula*, only three have been functionally characterized, UGT73K1, UGT71G1, and UGT73F3 (Achnine et al., 2005; Naoumkina et al., 2010). The factors regulating saponin synthesis in *M. truncatula* are still largely unknown. A RING E3 ubiquitin ligase Mkb1 involved in posttranscriptional regulation of HMGR and two transcription factors TSAR1 and TSAR2 controlling expression of genes encoding enzymes along the entire saponin synthsis pathway have been identified in *M. truncatula* hairy roots (Mertens et al., 2016; Pollier et al., 2013).

The *M. truncatula* species is made up of many genetically distinct accessions, some of which have been isolated from the wild while others originate from selective breeding programs. Three single-seed descent accessions that vary in the levels and types of saponins they accumulate were chosen for this research. A17, derived from a commercial variety of *M. truncatula*, and PRT178, isolated from the wild in Portugal, produce relatively high levels of saponins in their foliar tissue compared to ESP105, an accession isolated from the wild in Spain (Lloyd Sumner, unpublished data). Separate crosses were made between the low saponin producing ESP105 accession and each of the high saponin producing accessions, A17 and PRT178, to produce populations that were used to study the pattern of inheritance of foliar saponin accumulation. Expression of genes encoding enzymes involved in late steps of hemolytic saponin synthesis was also analyzed in the foliar tissue of cross progeny. The CYP72A68 cDNA from the three accessions was sequenced and compared to generate SNPs for confirmation of crosses and to assess whether amino acid substitutions among the accessions could explain differential saponin accumulation.

Materials and Methods

Plant maintenance

For plants used in crossing, *M. truncatula* seeds from the A17, ESP105, and PRT178 accessions were removed from seedpods and scarified for 5 min in a small volume (<1 ml) of 18 M sulfuric acid. Scarified seeds were then vernalized to induce early flowering by germinating on 0.8% water agar in Petri plates wrapped in foil in a 4°C cooler for 10 days. Successfully germinated seedlings were planted in Sunshine LC1 mix potting medium (SunGro Horticulture Distribution Inc., Bellevue, WA) and grown in a growth chamber with a 25°C daytime and a 21°C nighttime temperature and 60% humidity with a 16:8 (light:dark) hour photoperiod as recommended in the *Medicago truncatula* Handbook (Barker et al., 2007). Plants were watered when the top layer of the potting medium had dried, approximately every three days. Plants were fertilized weekly using Miracle-Gro© water-soluble tomato plant food (N-P-K ratio 18-18-21) (Scotts Miracle-Gro Products, Inc., Marysville, OH) at a rate of 15 g per 3.78 liters (1 gallon), and treated biweekly with a biological larvicide (Gnatrol© WDG, Valent BioSciences Corporation, Libertyville, IL) at a rate of 1.7 g per 3.78 liters. Transplantation to larger pots occurred as needed. Seed pods resulting from self-fertilization were removed regularly to maintain full-flowering potential.

Seed of F_1 and F_2 generation was scarified and germinated between two layers of moistened filter paper in a closed container. The container was placed in a dark cabinet for four days to allow for germination. Resulting plants were maintained in a growth chamber at 22.5°C and 60% humidity with a 16:8 (light:dark) hour photoperiod. Once presence of male parent DNA was confirmed in putative progeny, clonal cuttings were made from F1 plants to increase seed production. Using a sharp razor blade, a cut was made at a 45° angle below the youngest node of a healthy branch. The tip was immediately placed in a beaker of water for approximately 3 s to prevent xylem cavitation and then dipped in Hormex Rooting Powder No. 3 (Brooker Chemical, Chatworth, CA) before planting in Sunshine LC1 mix potting medium (SunGro Horticulture Distribution Inc., Bellevue, WA) in a tray covered with a dome to maintain high humidity. When seed pods were near maturity, original plants and cuttings were moved to a greenhouse, staked up, and wrapped in perforated plastic to allow for the collection of seedpods as they naturally detached from the plant.

Crosses

Crosses in *M. truncatula* were performed as described by Veerappan et al. (2014). On the female parent plant, an unopened flower with immature pollen was cut along the outer edge of the keel petal using a sharp scalpel. The standard and keel petal were then pushed back and forceps were used to remove the immature anthers, taking care not to disrupt anthers and stimulate pollen release. An unopened flower with mature pollen was then removed from the male parent plant. Using forceps, all petals were removed to expose the pollen-loaded anthers and the pollen was deposited on the stigma of the first flower. When possible, three flowers from the male parent plant were used to ensure an excess of pollen was deposited. The standard and keel were then repositioned and rubbed closed on the emasculated flower, and the flower was wrapped in plastic cling wrap for 2 days to create a humid environment that would discourage desiccation of internal organs. Seedpods from successful crosses were wrapped in perforated plastic wrap and collected when they naturally detached from the plant.

RNA extraction

Extraction of RNA for sequencing of CYP72A68 cDNA was performed according to the manufacturer's instruction using TRI Reagent® (Molecular Research Center, Inc. Cincinnati, OH). Roots from A17, ESP105, and PRT178 plants grown in a hydroponic system were collected into foil packets and immediately submerged in liquid nitrogen. Approximately 100 mg of root tissue were ground to a fine powder in liquid nitrogen using a mortar and pestle and transferred to a 2 ml tube with 1 ml of TRI Reagent®. After 5 min incubation at room temperature, 100 µl of bromochloropropane were added. Samples were mixed vigorously, incubated for 10 min at room temperature, and centrifuged at 12,000 g and 4°C for 15 min. The aqueous phase of each sample was then transferred to a fresh 1.5 ml tube and mixed by tube inversion with 250 µl isopropanol and 250 µl of a high salt solution to precipitate RNA. Samples were incubated at room temperature for 7 min and centrifuged at 12,000 g and 4°C for 7 min. The resulting pellets were washed with 1 ml 75% ethanol, centrifuged at 7,500 g and 4°C for 5 min, and air-dried. The pellet was dissolved in diethylpyrocarbonate-treated water and quantified using the BioSpec-nano™ spectrophotometer (Shimadzu Scientific Instruments, Kyoto, Japan).

RNA used in qPCR was extracted using the Qiagen RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany). Trifoliates were wounded using a hemostat to damage each leaflet 3.5 hours before tissue collection. Three trifoliates were collected per replicate into a foil packet and immediately submerged in liquid nitrogen. Three replicates were collected per experimental condition (A17, ESP105, and A17xESP105 F_2 plants with high, medium, or low hemolysis) with each replicate consisting of an individual plant. Trifoliates were ground to a fine powder in liquid nitrogen using a mortar and pestle and transferred to a 1.5 ml tube of lysis buffer with β-mercaptoethanol. The lysate was then centrifuged through a QIAshredder homogenizing column. Ethanol was added to the supernatant of the flow-through, and the sample was loaded onto an RNeasy mini spin column. Once the RNA had bound to the silica-based column membrane, contaminants were removed in three separate washes. The RNA was then eluted using RNase-free water and quantified using the BioSpec-nano™ spectrophotometer (Shimadzu Scientific Instruments, Kyoto, Japan).

All cDNA was synthesized using the iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Inc.). RNA template (1 µg) was combined with reaction buffer, reverse transcriptase, and nuclease-free water, and the reaction was run on a PTC-100™ Programmable Thermal Controller (MJ Research, Inc.) following a program of 25°C for 5 min, 42°C for 30 min, and 85°C for 5 min.

Sequencing

The CYP72A68 root cDNA was amplified through PCR using primers based on the known sequence of A17 CYP72A68 cDNA. PCR with 25 µl reaction volumes were carried out using 10X Taq buffer (GenScript), Taq DNA polymerase (5 U/µl, GenScript), and 1 µl cDNA template diluted 1:5 in water. The final concentrations of the dNTPs were 0.2 mM, and the final concentrations of the forward and reverse primers were 1.2 µM. PCR product was purified using the GeneJET™ PCR Purification Kit (Fermentas Life Sciences) and quantified using the BioSpec-nano™ spectrophotometer (Shimadzu Scientific Instruments, Kyoto, Japan). cDNA template (40 ng) was combined with 3.4 pmol primer and brought up to 13 µl total volume with water. Samples were then sent to the University of Arkansas DNA Resource Center for sequencing. Obtained sequences were aligned using the CLUSTAL W multiple sequence alignment tool (Higgins et al., 1992; Thompson et al., 1994) and translated using the

SIXFRAME tool made available through the San Diego Supercomputer Center (SDSC) Biology Workbench. Translated CYP72A68 sequences and CYP72A protein sequences from other plant species (GenBank accession numbers in Table 1) were aligned using CLUSTAL W and a rooted phylogenetic tree was created with PHYLIP (Felsenstein, 1989). *M. truncatula* CYP72A68 and legume CYP72A protein sequences were aligned and highlighted using the Multiple Align Show tool made available on the Sequence Manipulation Suite online (Stothard, 2000).

Table 1. GenBank accession numbers for plant CYP72A proteins

Species	Annotation	Accession No.
Arabidopsis thaliana	CYP72A15	OAD04430.1
Arachis duranensis	CYP72A15-like	XP_015932338.1
Cajanus cajan	CYP72A1-like	KYP48914.1
Catharanthus roseus	CYP72A1	Q05047.1
Cicer arietinum	CYP72A219-like	XP_004488665.1
Glycine max	CYP72A219-like	XP_003546751.1
Glycyrrhiza uralensis	CYP72A154	H1A988.1
Medicago truncatula (A17)	CYP72A68	ABC59077.1
Morus notabilus	CYP72A1-like	XP_010092742.1
Oryza sativa	CYP72A15-like	XP_015615867.1
Panax ginseng	CYP72A219	H2DH21.1
Vitus vinifera	CYP72A219-like	XP_003634773.1
Zea mays	CYP72A16	AAM77716.1

DNA extraction

DNA was extracted using the Qiagen DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany). Young trifoliates were collected on ice and used immediately for extraction. Approximately 100 mg tissue, or 3-4 trifoliates, was disrupted in lysis buffer containing RNase A stock solution using a small plastic pestle powered by a drill. After incubation at 65°C, proteins and polysaccharides were precipitated from the lysate using a salt solution. The precipitates and additional cell debris were then removed from the lysate by filtration through a QIAshredder Mini spin column. Binding buffer and ethanol were added to the lysate and the sample was applied to a DNeasy Mini spin column. After binding to the column membrane, the

DNA was washed twice to remove remaining contaminants and then eluted with 50 µl of a low-salt elution

buffer. DNA was quantified using the BioSpec-nano™ spectrophotometer (Shimadzu Scientific

Instruments, Kyoto, Japan).

Confirmation of crosses

PCR reactions with 25 µl reaction volumes were carried out using 10X Taq buffer (GenScript),

Taq DNA polymerase (5 U/µl, GenScript), and 1 µl undiluted DNA template (8-24 ng). The final

concentrations of the dNTPs were 0.2 mM, and the final concentrations of the forward and reverse

primers were 1.2 µM. Amplification of actin was performed simultaneously to confirm DNA was present

and able to be amplified. A Hybaid PCR Express thermal cycler was used to run the reactions. Different

primer sets required specific annealing temperatures to ensure stringent amplification (Table 2), so the

reactions were run using 1 cycle of 95°C for 2 min; 30 cycles of 95°C for 30 s, varying temperatures for

30 s, and 72°C for 30 s; and 1 cycle of 72°C for 5 min. Actin amplification was performed using an

annealing temperature of 55°C. Electrophoresis of PCR products was performed on 1% agarose gels

using GelGreen (Biotium, Hayward, CA). DNA was visualized with a FluorChem™ 8800 gel

documentation system (Alpha Innotech Corporation).

Table 2. Annealing temperatures for primers used in cross confirmation

Primer Name	Cross Tested	Annealing Temperature (°C)	Sequence
AxE_For	A17 x ESP105	58	5'-TGCTCCTTACATTCATCAC
AxE_Rev	A17 x ESP105	58	5'-AAGTCATTCTCAGTGGCA
ExP_For	ESP105 x PRT178	63.4	5'-GGGGTTTCAAAACTGGACAGTG
ExP_Rev	ESP105 x PRT178	63.4	5'-GCCGGAACCTATGCTGA
PxE_For	PRT178 x ESP105	66.6	5'-GGCAATGGCATTGATTTTAC
PxE_Rev	PRT178 x ESP105	66.6	5'-GATGACATACGTTTCTTCTCTGA

Saponin hemolysis assay

Plants used for the screening of A17 x ESP105 F_2 progeny were 4- to 7-weeks old. A handheld

plant sap press (Spectrum Technologies, Inc., Aurora, IL) was used to extract juices from five trifoliates

per plant. Leaf juices were kept on ice throughout the extraction process. Wells of approximately 3 mm diameter were made in blood agar plates (5% sheep blood in tryptic soy agar, BioMerieux Inc., Durham, SC for A17 x ESP105 F1 and Hardy Diagnostics, Santa Maria, CA for all others) using TipOne micropipette filter tips (1-200 µl, USA Scientific, Inc., Ocala, FL) that had been cut at the second line from the bottom of the tip and sterilized. Leaf extract (15 µl) was added to each well. Each plate contained 6 wells loaded with parent accession leaf juices, 3 per parent, as a control between plates and a standard for analyzing sample hemolysis. Plates were placed in a closed plastic container with a damp paper towel and left to incubate in a dark place at room temperature for approximately 4.5 hours or overnight.

To quantify the hemolysis observed for the 141 A17 x ESP105 F_2 plants, Microsoft Powerpoint was used to measure the diameter of clearing zones, including the well diameter. For each plate, the diameters of the three ESP105 wells were averaged. Then, the diameters for the A17 and F_2 clearing zones on the same plate were divided by this number to account for differences in picture size among plates. This ratio was called d/ESP. The d/ESP ratios for all the A17 clearing zones across all plates (n=24) were averaged and the standard deviation was calculated. The A17 d/ESP average minus the A17 d/ESP standard deviation served as the lower limit for the classification of the high hemolysis phenotype. Any F_2 plant with a d/ESP ratio lower than the high classification limit that had visible hemolysis was classified as medium. F_2 plants with no visible hemolysis were classified as low.

Gene expression studies using qPCR

Real-time PCR was performed using the Applied Biosystems StepOnePlus™ real time PCR systems. RNA extracted using the Qiagen RNeasy® Plant Mini Kit was used to synthesize cDNA using iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Inc.) as described previously. The cDNA was diluted 1:5 in water and 1 µl was added to 10 µl of *Power* SYBR® Green PCR master mix (Applied Biosystems) mixed with 0.27 µl each forward and reverse primers (15 µM) and 8.47 µl water, for a total reaction volume of 20 µl. Three technical and biological replicates were run for each sample using the following cycling protocol: 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 56°C for 1 min.

Equation 1: $\Delta Ct = Ct_{target} - Ct_{reference}$

target = gene of interest; reference = reference housekeeping gene

$$\text{Equation 2: } \Delta\Delta Ct = \Delta Ct_{experimental} - \Delta Ct_{control}$$

$$\text{experimental} = \text{A17 or } F_2 \text{ phenotype; control} = \text{ESP105}$$

$$\text{Equation 3: Fold Change} = 2^{-\Delta\Delta Ct}$$

Gene expression ratios for each gene were calculated using the $2^{-\Delta\Delta Ct}$ method described by Livak and Schmittgen (2001). The average Ct was calculated for each biological replication (e.g. A17 1, A17 2, A17 3, ESP105 1, etc.). Actin was used as the reference housekeeping gene. The average Ct values for the target and reference gene of the same biological replication were used in equation 1 to generate a ΔCt value for each biological replication. An average ΔCt was calculated for the ESP105 biological replications, arbitrarily chosen as the control due to its generally low levels of saponin synthesis gene expression. The ΔCt for each biological replication was used as the experimental value and, the average ESP105 ΔCt was used as the control value in equation 2 to generate a $\Delta\Delta Ct$ value for each biological replication. This value was then used in equation 3 to generate a fold change for each biological replication. The average and standard deviation of the fold change values for the three biological replications per biological group were then calculated.

Statistical analysis

GraphPad Prism (GraphPad Software Inc., La Jolla, CA) was used to analyze qPCR gene expression data. Biological group averages were compared using the Student's *t* test. Significance for all tests was defined as $P<0.05$, with the exception of Figure 8 for which significance was defined as $P<0.1$.

Results

Sequence of CYP72A68 cDNA is highly similar across *M. truncatula* accessions

Cytochrome P450 enzymes (CYPs) are members of a large superfamily of proteins encoded in a wide array of organisms. In plants, CYPs play important roles in detoxification and biosynthesis of specialized metabolites (Bak et al., 2011). In *M. truncatula*, CYP72A68 encodes the enzyme responsible for oxidation of triterpene hemolytic sapogenins at the C-23 position (Fukushima et al., 2013) and the conversion of oleanolic acid, 2β-OH-oleanolic acid, and bayogenin into hederagenin, bayogenin, and medicagenic acid, respectively (Biazzi et al., 2015)(Figure 2). Gene expression of CYP72A68 is induced by herbivory in A17 and PRT178 plants, but not in ESP105 plants, and accumulation of medicagenic acid, a product of CYP72A68 oxidation, in foliar tissue is approximately 18 times lower in ESP105 than it is in A17 (Rogers, 2015). The cDNA of CYP72A68 synthesized from RNA extracted from the roots of A17, ESP105, and PRT178 *M. truncatula* plants was sequenced to identify single nucleotide polymorphisms (SNPs) among the accessions that could be used as molecular markers for confirming the presence of male parent DNA in putative cross plants. In addition, SNPs in the open reading frame could identify potentially important differences in amino acid residue sequences between accessions.

```
PRT178      ----CGAGGAAAACATTTCACACTTAGTTTAGTTTTACTATTACAAACAAAGAAATGGAA
ESP105      ---TCGAGGAAAACATTTCACACTTAGTTTAGTTTTACTATTACAAACAAAGAAATGGAA
A17         ----CGAGGAAAACATTTCACACTTAGTTTGGTTTTACTATTACAAACAAAGAAATGGAA

PRT178      TTATCTTGGGAAGCAAAATCAGCCATAATTCTAATCACTGTGACATTTGGTTTGGTATAC
ESP105      TTATCTTGGGAAGCAAAATCAGCCATAATTCTAATCACTGTGACATTTGGTTTGGTATAC
A17         TTATCTTGGGAAACAAAATCAGCCATAATTCTCATCACTGTGACATTTGGTTTGGTATAC

PRT178      GCATGGAAGGTATTGAATTGGATGTGGCTGAAGCCAAAGAAGATAGAGAAGCTTTTAAGA
ESP105      GCATGGAGGGTATTGAATTGGATGTGGCTGAAGCCAAAGAAGATAGAGAAGCTTTTAAGA
A17         GCATGGAGGGTATTGAATTGGATGTGGCTGAAGCCAAAGAAGATAGAGAAGCTTTTAAGA

PRT178      GAACAAGGCCTTCAAGGGAATCCTTATAGACTTTTGCTTGGAGATGCAAAGGATTATTTT
ESP105      GAACAAGGCCTTCAAGGGAATCCTTATAGACTTTTGCTTGGAGATGCAAAGGATTATTTT
A17         GAACAAGGCCTTCAAGGGAACCCTTATAGACTTTTGCTTGGAGATGCAAAGGATTATTTT

PRT178      GTAATGCAAAAGAAAGTTCAATCCAAACCCATGAATCTATCTGATGATATTGCACCACGT
ESP105      GTAATGCAAAAGAAAGTTCAATCCAAACCCATGAATCTATCTGATGATATTGCACCACGT
A17         GTGATGCAAAAGAAAGTTCAATCCAAACCCATGAATCTATCTGATGATATTGCGCCACGT

PRT178      GTTGCTCCTTACATTCATCACGCTGTTCAAACTCATGGGAAAAAGTCTTTTATTTGGTTT
ESP105      GTTGCTCCTTACATTCATCACGCTGTTCAAACTCATGGGAAAAAGTCTTTTATTTGGTTT
A17         GTCGCTCCTTACATTCATCATGCTGTTCAAACTCATGGGAAAAAGTCTTTTATTTGGTTT
```

```
PRT178    GGAATGAAACCATGGGTGATTCTCAATGAACCTGAACAAATAAGAGAAGTATTCAACAAG
ESP105    GGAATGAAACCATGGGTGATTCTCAATGAACCTGAACAAATAAGAGAAGTATTCAACAAG
A17       GGAATGAAACCATGGGTGATTCTCAATGAACCTGAACAAATAAGAGAAGTATTCAACAAG

PRT178    ATGTCTGAGTTCCCAAAGGTTCAATATAAGTTTATGAAGTTAATAACTCGCGGTCTTGTT
ESP105    ATGTCTGAGTTCCCAAAGGTTCAATATAAGTTTATGAAGTTAATAACTCGCGGTCTTGTT
A17       ATGTCTGAGTTCCCAAAGGTTCAATATAAGTTTATGAAGTTAATAACTCGCGGTCTTGTT

PRT178    AAACTAGAAGGAGAAAAGTGGAGCAAGCATAGAAGAATAATCAACCCTGCGTTTCACATG
ESP105    AAACTAGAAGGAGAAAAGTGGAGCAAGCATAGAAGAATAATCAACCCTGCGTTTCACATG
A17       AAACTAGAAGGAGAAAAGTGGAGCAAGCATAGAAGAATAATCAACCCTGCGTTTCACATG

PRT178    GAAAAGTTGAAGATTATGACACCAACATTCTTGAAAAGCTGCAATGATTTGATTAGCAAT
ESP105    GAAAAATTGAAGATTATGACACCAACATTCTTGAAAAGCTGCAATGATTTGATTAGCAAT
A17       GAAAAATTGAAGATTATGACACCAACATTCTTGAAAAGCTGCAATGATTTGATTAGCAAT

PRT178    TGGGAAGAAAAGTTGTCTTCAAATGGATCATGTGAAATGGACATATGGCCTTCCCTTCAG
ESP105    TGGGAAGAAAAGTTGTCTTCAAATGGATCATGTGAAATGGACATATGGCCTTCCCTTCAG
A17       TGGGAAAAAATGTTGTCTTCAAATGGATCATGTGAAATGGACGTATGGCCTTCCCTTCAG

PRT178    AGCTTGACAAGTGATGTTATCGCTCGTTCGTCATTTGGAAGTAGTTATGAAGAAGGAAGA
ESP105    AGCTTGACAAGTGATGTTATCGCTCGTTCGTCATTTGGAAGTAGTTATGAAGAAGGAAGA
A17       AGCTTGACAAGTGATGTTATCGCTCGTTCGTCATTTGGAAGTAGTTATGAAGAAGGAAGA

PRT178    AAAGTATTTCAACTTCAAATAGAGCAAGGTGAACTTATAATGAAAAATCTAATGAAATCT
ESP105    AAAGTATTTCAACTTCAAATAGAGCAAGGTGAACTTATAATGAAAAATCTAATGAAATCT
A17       AAAGTATTTCAACTTCAAATAGAGCAAGGTGAACTTATAATGAAAAATCTAATGAAATCT

PRT178    TTAATCCCTTTATGGAGGTTTTTACCTACCGCTGATCATAGAAAGATAAATGAAAATGAA
ESP105    TTAATCCCTTTATGGAGGTTTTTACCTACCGCTGATCATAGAAAGATAAATGAAAATGAA
A17       TTAATCCCTTTATGGAGGTTTTTACCTACCGCTGATCATAGAAAGATAAATGAAAATGAA

PRT178    AAACAAATAGAAACTACTCTTAAGAATATAATTAACAAGAGGGAAAAAGCAATTAAGGCA
ESP105    AAACAAATAGAAACTACTCTTAAGAATATAATTAACAAGAGGGAAAAAGCAATTAAGGCA
A17       AAACAAATAGAAACTACTCTTAAGAATATAATTAACAAGAGGGAAAAAGCAATTAAGGCA

PRT178    GGTGATGCCACTGAGAATGACTTATTAGGTCTCCTCCTAGAGTCGAACCACAGAGAAATT
ESP105    GGTGATGCCACTGAGAATGACTTATTAGGTCTCCTCCTAGAGTCGAACCACAGAGAAATT
A17       GGTGAAGCCACTGAGAATGACTTATTAGGTCTCCTCCTAGAGTCGAACCACAGAGAAATT

PRT178    AAAGAACATGGAAACGTCAAGAATATGGGATTGAGTCTTGAAGAAGTAGTCGGGGAATGC
ESP105    AAAGAACATGGAAACGTCAAGAATATGGGATTGAGTCTTGAAGAAGTAGTCGGGGAATGC
A17       AAAGAACATGGAAACGTCAAGAATATGGGATTGAGTCTTGAAGAAGTAGTCGGGGAATGC

PRT178    AGGTTATTCCATGTTGCAGGGCAAGAGTCTACTTCAGATTTGCTTGTTTGGACGATGGTG
ESP105    AGGTTATTCCATGTTGCAGGGCAAGAGACTACTTCAGATTTGCTTGTTTGGACGATGGTG
A17       AGGTTATTCCATGTTGCAGGGCAAGAGACTACTTCAGATTTGCTTGTTTGGACGATGGTG

PRT178    TTGTTGAGTAGGTACCCTGATTGGCAAGAACGTGCAAGGAAGGAAGTATTAGAGATATTT
ESP105    TTGTTGAGTAGGTACCCTGATTGGCAAGAACGTGCAAGGAAGGAAGTATTAGAGATATTT
A17       TTGTTGAGTAGGTACCCTGATTGGCAAGAACGTGCAAGGAAGGAAGTATTAGAGATATTT

PRT178    GGCAATGAAAAACCCGACTTTGATGGACTAAATAAACTTAAGATTATGGCCATGATTTTG
ESP105    GGCAATGAAAAACCCGACTTTGATGGACTAAATAAACTTAAGATTATGGCCATGATTTTG
A17       GGCAATGAAAAACCCGACTTTGATGGACTAAATAAACTTAAGATTATGGCCATGATTTTG

PRT178    TATGAGGTTTTGAGGTTGTACCCTCCTGTAACCGGCGTTGCTCGAAAAGTTGAGAATGAT
ESP105    TATGAGGTTTTGAGGTTGTACCCTCCTGTAACCGGCGTTGCTCGAAAAGTTGAGAATGAT
```

```
A17             TATGAGGTTTTGAGGTTGTACCCTCCTGTAACCGGCGTTGCTCGAAAAGTTGAGAATGAT

PRT178          ATAAAACTTGGAGACTTGACATTATATGCTGGAATGGAGGTTTACATGCCAATTGTTTTG
ESP105          ATAAAACTTGGAGACTTGACATTATATGCTGGAATGGAGGTTTACATGCCAATTGTTTTG
A17             ATAAAACTTGGAGACTTGACATTATATGCTGGAATGGAGGTTTACATGCCAATTGTTTTG

PRT178          ATTCACCATGATTGTGAACTATGGGGTGATGATGCTAAGATTTTCAATCCTGAGAGATTT
ESP105          ATTCACCATGATTGTGAACTATGGGGTGATGATGCTAAGATTTTCAATCCTGAGAGATTT
A17             ATTCACCATGATTGTGAACTATGGGGTGATGATGCTAAGATTTTCAATCCTGAGAGATTT

PRT178          TCTGGTGGAATTTCCAAAGCAACAAACGGTAGATTTTCATATTTTCCGTTTGGAGCGGGT
ESP105          TCTGGTGGAATTTCCAAAGCAACAAACGGTAGATTTTCATATTTTCCGTTTGGAGCGGGT
A17             TCTGGTGGAATTTCCAAAGCAACAAACGGTAGATTTTCATATTTTCCGTTTGGAGCGGGT

PRT178          CCTAGAATCTGCATTGGACAAAACTTTTCCCTGTTGGAAGCAAAGATGGCAATGGCATTG
ESP105          CCTAGAATCTGCATTGGACAAAACTTTTCCCTGTTGGAAGCAAAGATGGCAATGGCATTG
A17             CCTAGAATCTGCATTGGACAAAACTTTTCCCTGTTGGAAGCAAAGATGGCAATGGCATTG

PRT178          ATTTTAAAGAATTTTTCATTTGAACTTTCTCAAACATATGCTCATGCTCCATCTGTGGTG
ESP105          ATTTTACAGAATTTTTCATTTGAACTTTCTCAAACATATGCTCATGCTCCATCTGTGGTG
A17             ATTTTAAAGAATTTTTCATTTGAACTTTCTCAAACATATGCTCATGCTCCATCTGTGGTG

PRT178          CTTTCTGTTCAGCCACAACATGGTGCTCATGTTATTCTACGCAAAATCAAAACATAAAAA
ESP105          CTTTCTGTTCAGCCACAACATGGTGCTCATGTTATTCTACGCAAAATCAAAACATAAAAA
A17             CTTTCTGTTCAGCCACAACATGGTGCTCATGTTATTCTACGCAAAATCAAAACATAAAAA

PRT178          CAACATCAGAGAAGAAACGTATGTCA------------------------------------
ESP105          CAACATCAGAGAAGAAACGTATGT--------------------------------------
A17             CAACATCAGAGAAGAAACGTATGTC-------------------------------------
```

Figure 3. Alignment of CYP72A68 cDNA sequences in A17, ESP105, and PRT178 *M. truncatula* plants. Sequences were 99% similar with 16 SNPs, shaded light gray. Start and stop codons are shaded dark gray.

Alignment of accession CYP72A68 cDNA sequences revealed high sequence similarity (Figure 3). There were 13 SNPs and 99.2% similarity between the A17 and ESP105 sequences, 15 SNPs and 99.1% similarity between the A17 and PRT178 sequences, and four SNPs and 99.8% similarity between the ESP105 and PRT178 sequences. There are two apparent isoforms for CYP72A68 in the *M. truncatula* genome, annotated as Medtr2g055430 and Medtr2g055470. The A17 CYP72A68 sequence showed 100% identity with Medtr2g05570 and 96.7% identity with Medtr2g055430. The predicted translation of the sequences was generated and the 520 amino acid sequences were aligned. There were six amino acid substitutions and 98.8% similarity between A17 and ESP105 sequences, seven substitutions and 98.7% similarity between A17 and PRT178 sequences, and three substitutions and 99.4% similarity between ESP105 and PRT178 sequences. For seven out of the eight total substitutions among the accessions, where the ESP105 sequence disagreed with either the A17 or PRT178 sequence,

it agreed with the sequence of the remaining accession, either PRT178 or A17. However, at residue 485 the ESP105 sequence differs from both of the high saponin producing accessions, containing a glutamine, while the A17 and PRT178 sequences have a lysine at this position.
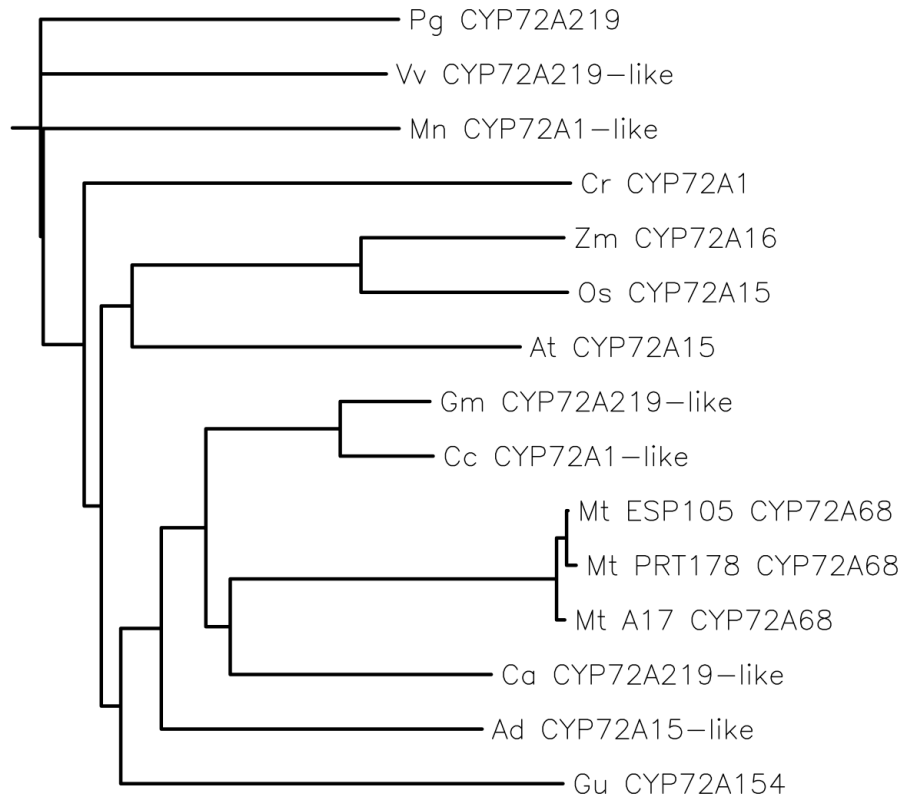
Figure 4. Phylogenetic relationships among CYP72A proteins in various plant species. A phylogenetic tree was generated based on the alignment of amino acid sequences of CYP72A proteins from *Glycyrrhiza uralensis* (Gu), *Cajanus cajan* (Cc), *Cicer arietinum* (Ca), the three *Medicago truncatula* (Mt) accessions, *Arachis duranensis* (Ad), *Arabidopsis thaliana* (At), *Zea mays* (Zm), *Orzya sativa* Japonica (Os), *Catharanthus roseus* (Cr), *Morus notabilis* (Mn), *Panax ginseng* (Pg), *Vitus vinifera* (Vv) and *Glycine max* (Gm).

A BLASTp search of the NCBI non-redundant protein database using the A17 CYP72A68 amino acid sequence yielded similar sequences for CYP72A proteins from other plant species: CYP72A15 from *Arabidopsis thaliana*, a CYP72A15-like protein from *Arachis duranensis*, a CYP72A15-like protein from *Oryza sativa* Japonica, CYP72A154 from *Glycyrrhiza uralensis*, CYP72A1 from *Catharanthus roseus*, a CYP72A1-like protein from *Morus notabilus*, a CYP72A1-like protein from *Cajanus cajan*, CYP72A16 from *Zea mays*, CYP72A219 from *Panax ginseng*, a CYP72A219-like protein from *Vitus vinifera*, a

CYP72A219-like protein from *Glycine max*, and a CYP72A219-like protein from *Cicer arietinum*. The

protein CYP72A1 from *C. roseus* (Madagascar periwinkle) has been functionally characterized as the

ring-opening enzyme that converts loganin into secologanin, an early step in indole alkaloid synthesis

(Irmler et al., 2000). The enzyme CYP72A154 from *G. uralensis* (Chinese licorice) has been functionally

characterized as a C-30 oxidase in the aglycone modification phase of biosynthesis of the triterpenoid

saponin glycyrrhizin. The *M. truncatula* enzyme CYP72A63 was shown to possess the same activity, but

no C-30 oxidated aglycones have yet been detected in profiled saponin extracts (Seki et al., 2011). The

phylogenetic tree in Figure 4 shows that proteins from legumes were more closely related to CYP72A68

in *M. truncatula* than proteins from non-legume plants.

The amino acid sequences for *M. truncatula* A17, ESP105, and PRT178 CYP72A68 and for the

proteins in Figure 4 from the five legume species were aligned to investigate whether the amino acid

substitution at residue 485 of the ESP105 sequence resulted in a chemically distinct amino acid in a

position that was otherwise highly conserved across all the legume species (Figure 5). The protein

sequences for all five legume species agree with the *M. truncatula* ESP105 CYP72A68 protein sequence

and contain a glutamine in the same position.

```
ESP105  - - - - M E L S W E A K S A I I L I T V T F G L V - - Y A W R V L N W M W L K P  34
PRT178  - - - - M E L S W E A K S A I I L I T V T F G L V - - Y A W K V L N W M W L K P  34
A17     - - - - M E L S W E T K S A I I L I T V T F G L V - - Y A W R V L N W M W L K P  34
Ca      M N K D V E V N W T T T S S I I L I I V T I A I L - - W A W K V L N L L W L K P  38
Gm      - - - - M E A A W A T I L T L I L I - - - L A L I - - C A W K V L I W L W L R P  31
Cc      - - - - M E A A W T I I L T L I L I V - - L V L I - - C T W R V L N W L W L R P  32
Ad      - - - - M E A A W P T - I V I A A A T V V V L L S - - W G W R I L N W L W L R P  33
Gu      - - - - M D A S S T P G A I W V V L T V I L A A I P I W V C H M V N T L W L R P  36

ESP105  K K I E K L L R E Q G L Q G N P Y R L L L G D A K D - - Y F V M Q K K V Q S K P  72
PRT178  K K I E K L L R E Q G L Q G N P Y R L L L G D A K D - - Y F V M Q K K V Q S K P  72
A17     K K I E K L L R E Q G L Q G N P Y R L L L G D A K D - - Y F V M Q K K V Q S K P  72
Ca      K K L E K L L R E Q G L Q G N P Y R F L V G D L K D - - L F K M E N E A K S K S  76
Gm      K R L E K L L R E Q G L Q G N P Y T L F V G D S K E - - F L K M R K E A L S K P  69
Cc      K R L E R L L R E Q G L Q G N P Y R L F F G D S K D - - V L N M R K E A L S K P  70
Ad      K K L E R L L R D Q G L R G T P Y K V L V G D S K D - - F M K M Q R E A R T K P  71
Gu      K R L E R H L R A Q G L H G D P Y K L S L D N S K Q T Y M L K L Q Q E A Q S K S  76
```

```
ESP105  M N L S - D D I A P R V A P Y I H H A V Q T H G K K S F I W F G M K P W V I L N   111
PRT178  M N L S - D D I A P R V A P Y I H H A V Q T H G K K S F I W F G M K P W V I L N   111
A17       M N L S - D D I A P R V A P Y I H H A V Q T H G K K S F I W F G M K P W V I L N   111
Ca        M N L T - D D I V P R V F P Y T Q S V K I H G K N S F I W F G T K P R L T L T   115
Gm        M N L S - D D I I P R V S S Y E Q H S V N K H G K N S F I W L G P I P R V T I L   108
Cc        M N L S - D D I V P R V S S Y D Q H S V N K H G K N S F I W L G P T P R V T L L   109
Ad        M N T S D D D I A P R T Q P F V L Q C F N K Y G K N S F F W S G G I P I V I V T   111
Gu        I G L S K D D A A P R I F S L A H Q T V H K Y G K N S F A W E G T A P K V I I T   116

ESP105  E P E Q I R E V F N K M S E F P K V Q Y K - F M K L I T R G L V K L E G E K W S   150
PRT178  E P E Q I R E V F N K M S E F P K V Q Y K - F M K L I T R G L V K L E G E K W S   150
A17       E P E Q I R E V F N K M S E F P K V Q Y K - F M K L I T R G L V K L E G E K W S   150
Ca        E P E Q I K D V L N K I S D F P K T N Y K - I F K L L A S G L A S Y H G E K W S   154
Gm        D P E Q I K D V F N K I Y D F P K P N M N P L V K L L A T G L A G Y E G E K W S   148
Cc        D P E Q I K D V F N K I N D F P K P D S N P L V K L L A T G V A G Y E G E K W S   149
Ad        D P E L I K D V F N K N Y D F Q K P N T N P L I K L L V N G L G G H E G D K W R   151
Gu        D P E Q I K E V F N K I Q D F P K P K L N P I A K Y I S I G L V Q Y E G D K W A   156

ESP105  K H R R I I N P A F H M E K L K I M T P T F L K S C N D L I S N W E E K L S S N   190
PRT178  K H R R I I N P A F H M E K L K I M T P T F L K S C N D L I S N W E E K L S S N   190
A17       K H R R I I N P A F H M E K L K I M T P T F L K S C N D L I S N W E K M L S S N   190
Ca        K H R K L I N S A F H L E K L K I M T P S F F T S C N D L I S E W E E I L S S N   194
Gm        K H R R I I N P A F N L E K L K I M L P L F F K S C N D L V S K W E G M L S S E   188
Cc        K H R R I I N P A F N L E K L K I M L P L F F K S C N D L V S E W E G M L S S N   189
Ad        K H R R I I N P A F H S E K L K I M I P I F F K S C N D I I C K W E K M I S S N   191
Gu        K H R K I I N P A F H L E K L K G M L P A F S H S C H E M I S K W K G L L S S D   196

ESP105  G S C E M D I W P S L Q S L T S D V I A R S S F G S S Y E E G R K V F Q L Q I E   230
PRT178  G S C E M D I W P S L Q S L T S D V I A R S S F G S S Y E E G R K V F Q L Q I E   230
A17       G S C E M D V W P S L Q S L T S D V I A R S S F G S S Y E E G R K V F Q L Q I E   230
Ca        G S C E I D I W P S L Q N M A S D A I S R T A F G S S Y E E G K R I F Q L Q R E   234
Gm        G S C E M D A W P F L Q N L A S D V I A R S A F G S S Y E E G R R I F Q L Q R E   228
Cc        G S C E M D A W P F L Q N L A S D V I A R S A F G S S Y E E G R R I F Q L L R E   229
Ad        E S C E I D V W P S F Q N L A G D V I S R A A F G S S Y E E G I R I F E L Q R E   231
Gu        G T C E V D V W P F L Q N L T C D V I S R T A F G S S Y A E G A K I F E L L K R   236

ESP105  Q G E L I M K N L M K S L I P L W R F L P T A D H R K I N E N E K Q I E T T L K   270
PRT178  Q G E L I M K N L M K S L I P L W R F L P T A D H R K I N E N E K Q I E T T L K   270
A17       Q G E L I M K N L M K S L I P L W R F L P T A D H R K I N E N E K Q I E T T L K   270
Ca        Q A E L I T K V V M K S F I P L W R F V P T I V H R R M N E V D K D I K S S L K   274
Gm        Q A E L L I K V L L K I Q I P G W R F L P T N T H R R M K E I D R D I K A S L K   268
Cc        Q A E L L M K V L L K I Q I P G W R F L P T N T H R R M K E I D K D I K A S L K   269
Ad        Q L K L T M E V V M K V Y I P G S R F L P T R N N R R M K E I D R D I K T S L K   271
Gu        Q G - Y A L M T A R Y A R I P L W W L L P S T T K R R M K E I E R G I R D S L E   275
```

39

```
ESP105  N I INKREKAIKAG D ATEN DLLGLLLES NHREIKEHG--NV  308
PRT178  N I INKREKAIKAG D ATEN DLLGLLLES NHREIKEHG--NV  308
A17     N I INKREKAIKAG E ATEN DLLGLLLES NHREIKEHG--NV  308
Ca      D M INKREKTLKAG E ANKN DLLGILLES NHKEIKEHG--NN  312
Gm      D M INKREKALKAG E ATKN DLLGILLES NHKEIQEHGNRNN  308
Cc      D M INKREKALKAG E ATKN DLLGILLES NHKEIEEHG--NN  307
Ad      K I INKKEKALKAG E TAKD DLLGILLES NQKEIEEYG--KK  309
Gu      G I IRKREKALKSG K STDD DLLGILLQS NHIENKGDE--NS  313

ESP105  KNMGLSLEEVVGE C RLF HVAGQE T TSDLLVWTMVLLSRYP  348
PRT178  KNMGLSLEEVVGE C RLF HVAGQE S TSDLLVWTMVLLSRYP  348
A17     KNMGLSLEEVVGE C RLF HVAGQE T TSDLLVWTMVLLSRYP  348
Ca      KNAGMNIEDVIEE C KLF YFAGQE T TSVLLVWTMILLSRYP  352
Gm      KNVGMSLEEVIEE C KLF YFAGQE T TSVLLVWTMVLLSRYP  348
Cc      KNVGMSLKEVIEE C KLF YFAGQE T TSVLLVWTMVLLSRYP  347
Ad      KNNGMSVEDVIDE C KLF YFAGQE T TSVLLVWSMVLLSKHS  349
Gu      KSAGMTTQEVMEE C KLF YLAGQE T TAALLAWTMVLLGKHP  353

ESP105  DWQERARKEVLEIFGNEKPDFDGLNKLKIMAMILYEVLRL  388
PRT178  DWQERARKEVLEIFGNEKPDFDGLNKLKIMAMILYEVLRL  388
A17     DWQERARKEVLEIFGNEKPDFDGLNKLKIMAMILYEVLRL  388
Ca      DWQARAREEVFHVFGNKKPDFDGLSNLKIVTMVLYEVLRL  392
Gm      DWQARAREEVFQVFGYQKPDFDGLSRLKIVTMILYEVLRL  388
Cc      DWQTRAREEVLQVFGNQKPDFDGLSRLKIVTMILYEVLRL  387
Ad      YWQARAREEVLQVFGSQKPHLDGLNRLKIVTMILYEVLRL  389
Gu      EWQARARQEVLQVFGNQNPNFEGLGRLKIVTMILYEVLRL  393

ESP105  YPPVTGVARKVENDIKLGDLTLYAGMEVYMPIVLIHHDCE  428
PRT178  YPPVTGVARKVENDIKLGDLTLYAGMEVYMPIVLIHHDCE  428
A17     YPPVTGVARKVENDIKLGDLTLYAGMEVYMPIVLIHHDCE  428
Ca      YPPVIGLARKIDKDMKLGNLTLPAGVEVFLPIILLHHDTQ  432
Gm      YPPVAGMTRSIEKDVKLGTLTLPAGVHVLLPTILIHHDRK  428
Cc      YSPVLGLTRSVKKDIKLGNLTLPAGVHVFLPTMLIHHDCE  427
Ad      YPPVPGLTRTTAKDIKLGDLNLPQGVQIFFPIVLIHHDFR  429
Gu      YPPGIYLTRALRKDLKLGNLLLPAGVQVSVPILLIHHDEG  433

ESP105  LWGDDAKIFNPERFSGGISKATNGRFSYFPFGAGPRICIG  468
PRT178  LWGDDAKIFNPERFSGGISKATNGRFSYFPFGAGPRICIG  468
A17     LWGDDAKIFNPERFSGGISKATNGRFSYFPFGAGPRICIG  468
Ca      LWGDDATMFNPERFSGGVSKATNGRVSFFPFGWGPRICVG  472
Gm      FWGEDAKQFNPERFSEGVLKATNGRVSFFPFGWGPRICIG  468
Cc      LWGEDAKQFNPERFSEGILKATNGRVSFFPFGWGPRICIG  467
Ad      LWGNDAKNFNPERFSEGVFKATNGKASFLPFGWGPRICIG  469
Gu      IWGNDAKEFNPERFAEGIAKATKGQVCYFPFGWGPRICVG  473
```

```
ESP105 QNFSLLEAKMAMALILQNFSFELSQTYAHAPSVVLSVQPQ 508
PRT178 QNFSLLEAKMAMALILKNFSFELSQTYAHAPSVVLSVQPQ 508
A17    QNFSLLEAKMAMALILKNFSFELSQTYAHAPSVVLSVQPQ 508
Ca     QNFSLLEAKMAMAMILQHFSFELSPTYSHAPTTVITLRPQ 512
Gm     QNFSLLEAKMALSMILQHFSFELSPAYAHAPTALITIQPQ 508
Cc     QNFSLLEAKMALSMILQNFSFELSPAYAHAPTMEITLHPQ 507
Ad     QNFSLLEAKMALSLILQHFSFELSPAYTHAPMTMALLHPQ 509
Gu     QNFALLEAKIVLSLLLQNFSFELSPTYAHVPTTVLTLQPK 513

ESP105 HGAHVILRKIKT 520
PRT178 HGAHVILRKIKT 520
A17    HGAHVILRKIKT 520
Ca     HGAHIILRKVET 524
Gm     YGAHIILRKVTI 520
Cc     YGAHIILRKVNK 519
Ad     HGAHIILHKVKI 521
Gu     HGAPIILHKL-- 523
```

Figure 5. Alignment of amino acid sequences for CYP72A proteins in various legumes. Amino acid sequences for CYP72A68 in *M. truncatula* accessions A17, PRT178, and ESP105 were aligned with amino acid sequences for a CYP72A219-like protein in *G. max*, a CYP72A15-like protein in *A. duranensis*, CYP72A153 in *G. uralensis*, a CYP72A1-like protein in *C. cajan*, and a CYP72A219-like protein in *C. areitinum*. Amino acids that are identical across the sequences are shaded black, while amino acids that are chemically similar across the sequences are shaded dark gray. No shading indicates no consensus, and red lettering indicates the position of substitutions among the *M. truncatula* accessions.

Confirmation of the presence of male parent DNA in putative cross progeny

Crosses were made between accessions with high levels of triterpene saponin accumulation, A17 or PRT178, and an accession with low levels of triterpene saponin accumulation, ESP105. The resulting population could potentially be used to investigate and identify the genetic components controlling saponin synthesis in *M. truncatula*. Two separate crosses of A17 and ESP105 *M. truncatula* plants in which A17 acted as the female parent and ESP105 acted as the male parent were performed, as well two separate reciprocal crosses of ESP105 and PRT178 plants in which each accession acted as the female parent in one cross and the male parent in the other cross. Successful fertilization resulted in the formation of a seedpod containing three to ten seeds of the F_1 generation. Due to the possibility of unwanted self-fertilization, the presence of the desired male parent DNA was confirmed in all putative F_1 plants. Primers were designed around SNPs in the A17, ESP105, and PRT178 CYP72A68 cDNA

sequences specific to the male parent accession and used to amplify DNA from parent accessions and putative F_1 plants. Six A17 x ESP105, two ESP105 (female) x PRT178 (male) and one PRT178 (female) x ESP105 (male) F_1 plants were tested and all were confirmed to be crosses by the amplification of male parent accession markers (Figure 6). Seed resulting from self-fertilization of confirmed F_1 plants and their clonal cuttings was collected.
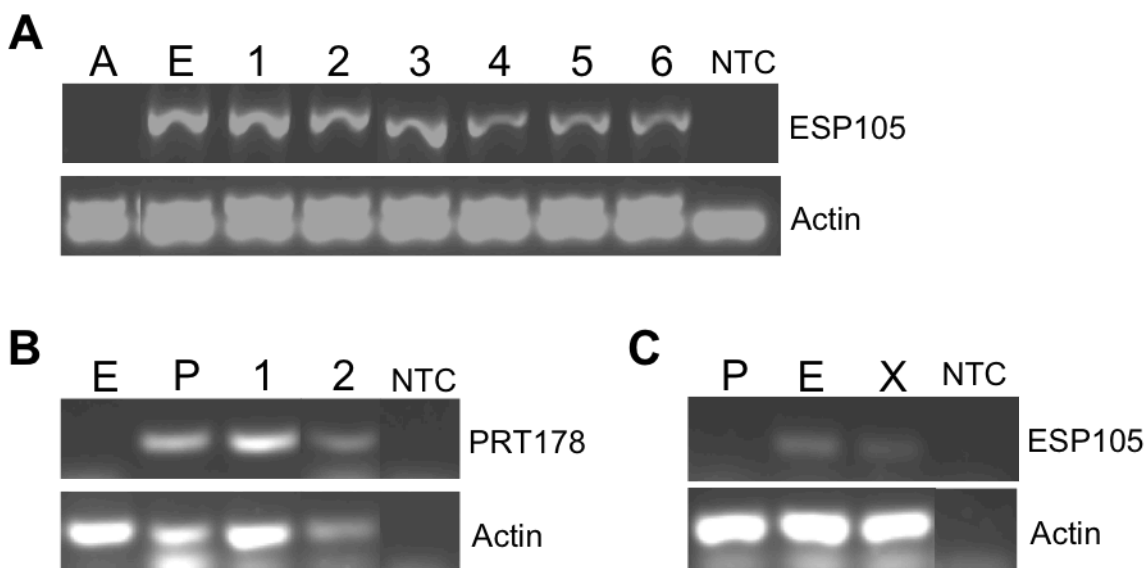
Figure 6. Confirmation of presence of male parent DNA in putative cross progeny. Primers based on SNPs in the A17 (A), ESP105 (E), and PRT178 (P) CYP72A68 cDNA sequences were used. DNA was amplified from the male parent accessions and from putative F_1 plants (1-6 and X), but not in the female parent accessions or the no-template control (NTC) reactions. A17 x ESP105 (panel A), ESP105 x PRT178 (panel B), and PRT178 x ESP105 (panel C) F_1 plants were tested.

Cross progeny and parent accession leaf extracts display varying levels of hemolytic activity

Because saponin accumulation in leaves varies widely in the parents, F_2 progeny from the crosses described above were predicted to have varying levels of hemolytic saponins segregating in the population. To screen the A17 x ESP105 F_2 generation for inheritance of triterpene saponin accumulation quickly and cheaply, the hemolytic activity of leaf extracts from F_2 and parent accession plants was assessed by application of extracts to blood agar Petri plates and observation of hemolytic clearing zones (Figure 7). Leaf extracts from the A17 parent accession produced high levels of hemolysis, while extracts from the ESP105 parent accession did not produce any visible hemolysis. Hemolytic activity of extracts

from 141 F_2 plants resulting from self-fertilization of all six confirmed F_1 plants (Figure 6A) was tested. Hemolysis caused by A17 and F_2 leaf extracts was expressed as d/ESP by measuring the clearing zone diameter (d) and dividing it by the average diameter of the three ESP105 wells on the same Petri plate. Three distinct hemolysis phenotypes were observed; extracts from 28 F_2 plants caused no visible hemolysis, resembling the ESP105 parent accession, extracts from another 28 F_2 plants caused high levels of hemolysis resembling those caused by the A17 parent accession, and 85 F_2 plants caused intermediate levels of hemolysis between the high and low levels caused by the parent accessions. The ratio of high:medium:low hemolysis phenotype plants was therefore approximately 1:3:1. The presence of the intermediate phenotype and 1:3:1 F_2 phenotypic ratio suggests that inheritance of hemolytic saponin accumulation could depend on incomplete dominance involving more than a single gene with recessive and dominant alleles. Alternatively, there could be epistatic effects of genes on the regulators of saponin accumulation.
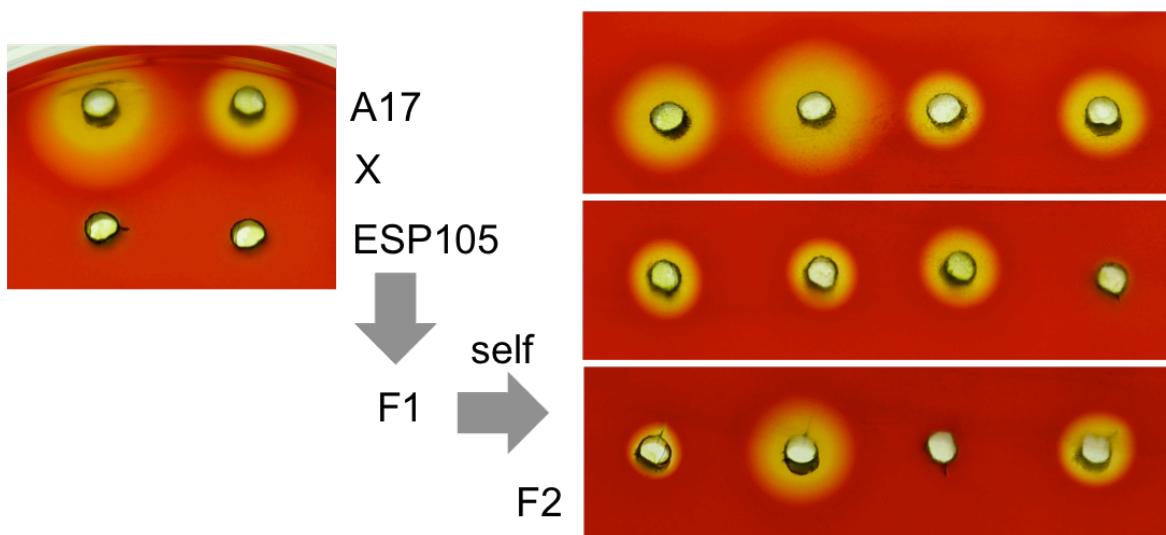
Figure 7. Segregation of hemolytic saponin levels in A17 x ESP105 F_2 plants as demonstrated on blood-agar Petri plates. F_2 progeny demonstrate high, medium or low levels of hemolytic activity. The high:medium:low segregation ratio was approximately 1:3:1, suggesting that several genes contribute to regulation of saponin accumulation.

Expression of saponin synthesis genes correlates with A17 x ESP105 F_2 hemolysis phenotypes

Quantitative reverse transcriptase PCR (qRT-PCR) was used to examine the relationship between the three hemolysis phenotypes observed in the A17 x ESP105 F_2 generation and expression of

saponin synthesis genes. RNA was extracted from three F_2 plants each with high, medium, and low

hemolysis phenotypes, as well as three plants each of the A17 and ESP105 parent accessions. Leaves

were mechanically wounded 3.5 hours prior to extraction to induce expression of saponin synthesis

genes. Gene expression was evaluated using the $2^{-\Delta\Delta Ct}$ method which utilizes a reference control sample

and reference gene target to evaluate expression of the gene of interest. The parent accession ESP105

was chosen as the reference control sample, and the sequence encoding actin, an established

housekeeping gene whose expression is similar in the varying genotypes of this experiment, was used as

the reference gene.

Figure 8. Relative gene expression of CYP716A12 in *M. truncatula* leaves subjected to mechanical damage, taken from A17, ESP105, and from A17xESP105 F2 plants with high, medium, and low hemolysis phenotypes. Expression ratios were plotted relative to ESP105. Error bars indicate SEM (n=3). Statistical significance calculated by Student's *t* test (P < 0.1). For correlation between CYP716A12 gene expression and hemoysis levels R^2=0.748.

CYP716A12 gene expression in A17 and ESP105 parent accessions and F_2 progeny correlated

with the observed hemolysis phenotypes (Figure 8). CYP716A12 encodes the enzyme responsible for the

conversion of β-amyrin into oleanolic acid through oxidation at the C-28 position, the first committed step

in hemolytic saponin synthesis in *M. truncatula*. The A17 parent accession, which accumulates high

levels of hemolytic saponins and caused high levels of hemolysis in the blood agar assay, had much

higher levels of CYP716A12 expression than the ESP105 parent accession that accumulates low levels

of hemolytic saponins and did not cause any visible hemolysis. CYP716A12 expression in A17 x ESP105 F_2 plants that caused high levels of hemolysis was even higher than that of A17, though the difference was not statistically significant. Expression of CYP716A12 in F_2 plants with low levels of hemolysis was similar to ESP105 expression levels, while expression in medium hemolysis phenotype F_2 plants was higher than, but not significantly different from, ESP105 expression. The correlation between CYP716A12 gene expression and d/ESP values was positive and had an R^2 value of 0.748.
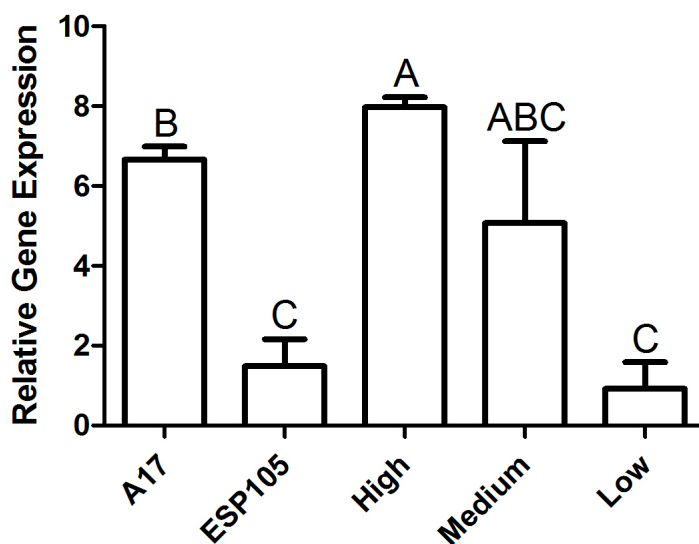
Figure 9. Relative gene expression of CYP72A67 in *M. truncatula* leaves subjected to mechanical damage, taken from A17, ESP105, and from A17xESP105 F2 plants with high, medium, and low hemolysis phenotypes. Expression ratios were plotted relative to ESP105. Error bars indicate SEM (n=3). Statistical significance calculated by Student's *t* test (P < 0.05). For correlation between CYP72A67 gene expression and hemoysis levels R^2=0.719.

Oleanolic acid formed by the CYP716A12 enzyme can be further modified by oxidation at the C-2 position catalyzed by the CYP72A67 enzyme to form additional sapogenins. The pattern of CYP72A67 gene expression in A17 x ESP105 F_2 plants was comparable to that of CYP716A12 expression and hemolytic activity (Figure 9). Accumulation of CYP72A67 transcripts was approximately six times higher in A17 parent accession plants than in ESP105 plants, in agreement with the higher levels of saponin accumulation and hemolytic activity observed in A17 plants. In F_2 plants from the A17 x ESP105 cross with high levels of hemolysis in blood agar assays, CYP72A67 mRNA expression levels were similar to A17 expression levels in that they were significantly higher than ESP105 expressions levels (P=0.0008).

45

Interestingly, they were also significantly higher than A17 expression levels (P=0.033). Expression of CYP72A67 transcripts in low-hemolysis phenotype F_2 plants was similar to expression in ESP105 plants, and also significantly lower than expression in A17 parents and the high-hemolysis F_2 plants (P<0.01). Expression of CYP72A67 mRNA in medium-hemolysis F_2 plants was higher than expression in low-hemolysis plants and lower than expression in high-hemolysis plants, but not significantly different from either due to high variability in gene expression values. The correlation between CYP72A67 gene expression and d/ESP values was positive and had an R^2 value of 0.719.
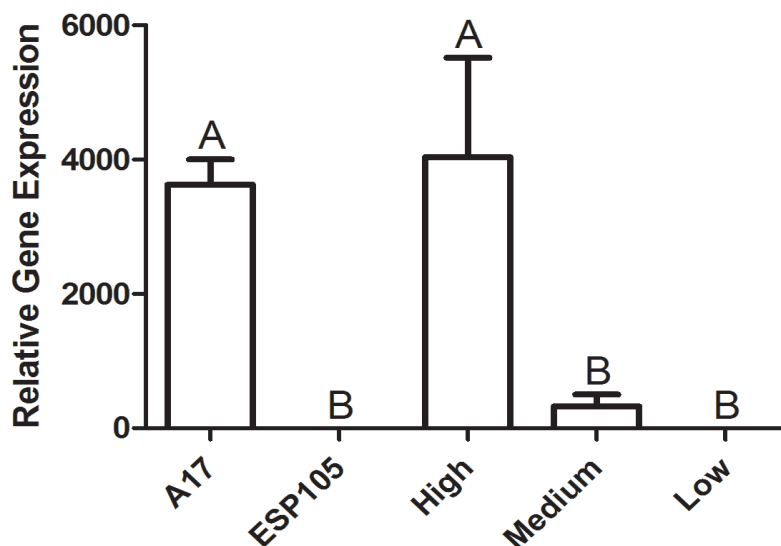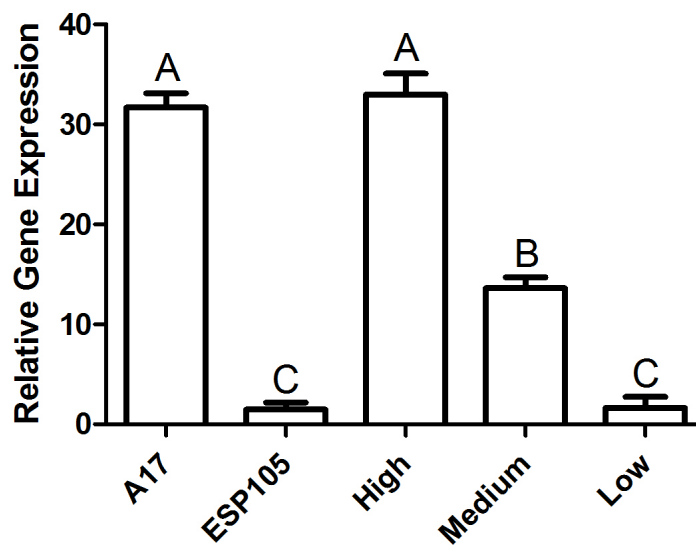

Figure 10. Relative gene expression of CYP72A68 in *M. truncatula* leaves subjected to mechanical damage, taken from A17, ESP105, and from A17 x ESP105 F2 plants with high, medium, and low hemolysis phenotypes. Expression ratios were plotted relative to ESP105. Error bars indicate SEM (n=3). Statistical significance calculated by Student's *t* test (P < 0.05). For correlation between CYP72A68 gene expression and hemoysis levels R^2=0.959.

Oleanolic acid can also be modified by oxidation at the C-23 position, in a reaction catalyzed by the CYP72A68 enzyme. Accumulation of CYP72A68-encoding transcripts in A17 x ESP105 F_2 plants was strongly correlated with the three observed hemolysis phenotypes (Figure 10). Similar to the expression pattern seen for CYP716A12 and CYP72A67, CYP72A68 expression in the A17 parent accession was approximately 32-times higher and significantly different from expression in ESP105 (p<0.0001). High-hemolysis F_2 plants demonstrated very similar expression levels of CYP72A68 that were also significantly higher than expression levels in ESP105 (P=0.0001). CYP72A68 expression in low-hemolysis F_2 plants

was similar to that in ESP105 and significantly lower than expression in A17 and high-hemolysis F_2 plants (P<0.001), while CYP72A68 expression in medium hemolysis F_2 plants was significantly higher than that of ESP105 and low-hemolysis F_2 plants (P<0.01) and significantly lower than that of A17 and high-hemolysis F_2 plants (P<0.001). The correlation between CYP72A68 gene expression and d/ESP values was positive and had an R^2 value of 0.959.

Discussion

Saponins are a family of specialized metabolites produced by a variety of plants as a means of defense. They are used in many industrial applications, including as additives in food and cosmetic products, due their emulsifying, foaming, detergent, and surfactant properties. There has been interest in utilizing saponins in agriculture to protect crops from a wide range of pests and as feed additives to improve livestock performance. Saponins also possess many properties that could be exploited to improve human health, including hypocholesterolemic, immunostimulatory, hypoglycemic, neuroprotective, antibacterial, antiviral, and cytotoxic activities. In spite of their usefulness, the genetic basis for the saponin biosynthetic pathway has not been completely elucidated. The annual legume *M. truncatula* accumulates triterpene saponins constitutively and in response to herbivory in its roots, stems, leaves, seedpods, and seeds. While the enzymes involved in many of the steps of saponin biosynthesis in *M. truncatula* have been identified and functionally characterized, little is known about the genes responsible for regulation of those enzymes. The *M. truncatula* species encompasses genetically distinct accessions that vary in the levels and types of saponins that they produce. We hypothesized that populations generated by crosses between an accession that accumulates low levels of saponins in foliar tissue, ESP105, and an accession that accumulates high levels of saponin in foliar tissue, A17 or PRT178, could be used to investigate the genetic components controlling saponin synthesis. Three types of crosses were performed, A17 x ESP105, PRT178 x ESP105, and ESP105 x PRT178, and confirmed using accession molecular markers based on SNPs found in the cDNA sequence of CYP72A68, an enzyme involved in later steps of hemolytic triterpene saponin biosynthesis. Hemolytic saponin accumulation inheritance in A17 x ESP105 F_2 progeny was evaluated by measuring hemolytic activity of leaf extracts, and three distinct hemolysis phenotypes occurring at a ratio of 1:3:1 were observed. By investigating gene expression in F_2 progeny exhibiting the three hemolysis phenotypes, we found that hemolytic saponin accumulation correlated with expression of three genes involved in hemolytic saponin biosynthesis, CYP716A12, CYP72A67, and CYP72A68.

A CYP72A68 cDNA from *M. truncatula* A17, ESP105, and PRT178 plants was sequenced to identify SNPs among the accessions that could be used as molecular markers to confirm the presence of male parent DNA in putative crosses. The cDNA and protein sequences obtained were very similar with

only 16 SNPs and 8 amino acid substitutions across the accessions. The ESP105 and PRT178

sequences were more similar to each than they were to the A17 sequence, with 99.8% similarity at the

nucleotide level and 99.4% similarity at the amino acid level. The CYP72A68 enzyme is a C-23 oxidase

required for the production of the triterpene backbones hederagenin, bayogenin, and medicagenic acid,

listed in order of increasing oxidation at the C-23 position (Biazzi et al., 2015; Fukushima et al., 2013).

Two isoforms of the CYP72A68 gene have been identified in the *M. truncatula* genome, Medtr2g055430

and Medtr2g055470, which show 97% identity at the nucleotide level and 96% identity at the amino acid

level. The Medtr2g055470 version demonstrated a higher efficiency than the Medtr2g055430 version

when each was expressed in a yeast system and supplied with substrates. Both isoforms more effectively

transformed substrates carrying a C-2 hydroxl group than those lacking C-2 oxidation, suggesting that

CYP72A68-mediated oxidation at the C-23 position of triterpene aglycones preferentially occurs after

CYP72A67-mediated oxidation at the C-2 position (Biazzi et al., 2015). The obtained A17 CYP72A68

cDNA sequence shows 100% identity with the Medtr2g055470 cDNA. Expression of CYP72A68 in foliar

tissue is induced by herbivory in A17 and PRT178 plants but not in ESP105 (Rogers, 2015), and A17

plants constitutively accumulate higher levels of hemolytic triterpene saponins, the products of CYP72A68

activity, than ESP105 in foliar tissue (Figure 7). This suggests that CYP72A68 gene expression is

differentially regulated between the high saponin producing accessions A17 and PRT178 and the low

saponin producing accession ESP105 and/or that the protein encoded by the ESP105 version of the

CYP72A68 gene is non-functioning. Although the amino acid sequences of the three accessions are very

similar, even a single substitution can have an effect on protein function and processing. For seven of the

eight amino acid substitutions found among the accessions, where the ESP105 sequence differs from the

A17 sequence, it is the same as the PRT178 sequence, and vice versa. This would suggest that these

substitutions, present in both high and low saponin producing accessions, do not lead to a non-

functioning CYP72A68 protein. However, there is one substitution at residue 485 for which ESP105 has

an amino acid, glutamine, different from the lysine present in the A17 and PRT178 sequences. The

sequences of CYP72A proteins from other plant species were used to investigate the importance of this

substitution.

The cytochrome P450 superfamily of proteins is made up of hemothiolate enzymes with highly divergent sequences that catalyze a wide range of reactions and are present in all organisms from humans to bacteria. Plant cytochrome P450s control the synthesis of phytohormones and signaling molecules, such as gibberellins and jasmonates, metabolize and detoxify xenobiotics like pesticides and pollutants, and are involved in the biosynthesis of pigments, antioxidants, volatiles, allelochemicals, and defense compounds. The superfamily is divided into families and subfamilies with members of the same family generally sharing at least 40% identity and members of the same subfamily sharing at least 55% identity (Bak et al., 2011). The *M. truncatula* CYP72A68 protein that was sequenced is part of the A subfamily of the 72 family of cytrochrome P450 superfamily. Twelve CYP72A and CYP72A-like proteins from twelve different plant species were chosen based on similarity to the *M. truncatula* A17 CYP72A68 sequence and relevance of the species and protein function and aligned with the three *M. truncatula* CYP72A68 sequences to generate a phylogenetic tree (Figure 4). Five of the proteins were from legumes, including *Arachis duranensis*, an herb used in peanut research, *Glycyrrhiza uralensis* (Chinese licorice), the important crop *Glycine max* (soybean), *Cajanus cajan* (pigeon pea), and *Cicer arietinum* (chickpea), and seven of the proteins were from non-legumes, including the model organism *Arabidopsis thaliana*, *Catharanthus roseus* (Madagascar periwinkle), the medicinal plant *Panax ginseng*, *Zea mays* (corn), *Oryza sativa* (rice), *Vitus vinifera* (common grape vine), and *Morus notabilus* (mulberry tree). The *C. roseus* protein CYP72A1 has been functionally characterized as the ring-opening enzyme responsible for the conversion of loganin into secologanin in the early steps of indole alkaloid biosynthesis (Irmler et al., 2000), and the *G. uralensis* CYP72A154 protein has been functionally characterized as a C-30 oxidase responsible for aglycone modification in the biosynthesis of the oleanane-type triterpene saponin glycyrrhizin (Seki et al., 2011). The remaining proteins have been annotated as CYP72A proteins or as similar to CYP72A, and their function is unknown. The uncharacterized CYP72A proteins from legumes clustered around the *M. truncatula* CYP72A68 and *G. uralensis* CYP72A154 proteins, while the uncharacterized CYP72A proteins from non-legumes clustered around the *C. roseus* CYP72A1 protein. This could indicate that the two groups of proteins may function differently; with the uncharacterized CYP72A proteins from legumes catalyzing reactions similar to those catalyzed by CYP72A68 and

CYP72A154, and uncharacterized CYP72A proteins from non-legumes catalyzing reactions similar to the reaction catalyzed by CYP72A1.

The three *M. truncatula* CYP72A68 sequences were aligned with the CYP72A protein sequences from the five legume species to investigate amino acid conservation at residue 485. The sequences for the five proteins, including that of the characterized *G. uralensis* CYP72A154 enzyme, agreed with the ESP105 CYP72A68 sequence and contained a glutamine at this position. At every position that the CYP72A68 sequence from the low saponin producing accession ESP105 differed from one or both of the sequences from the high saponin producing accessions A17 and PRT178, it agreed with the sequences of either the other high-saponin producing accession or CYP72A sequences from other legumes. This suggests that the low levels of hemolytic saponin accumulation present in ESP105 plants are not due to allelic variation in synthesis genes resulting in non-functioning enzymes, but are instead due to differential regulation of those synthesis genes and/or enzymes resulting in lower levels of enzyme activity.

Two crosses between A17 and ESP105 plants in which A17 acted as the female parent and two crosses between ESP105 and PRT178 plants in which each accession had a turn acting as the female parent were made and successfully confirmed using primers based on the SNPs found in the CYP72A68 cDNA sequence. Seed resulting from self-fertilization of the F_1 generation and its clonal cuttings was collected for all cross types. The A17 x ESP105 population has already begun to be utilized in this thesis research for the study of saponin accumulation inheritance. All the populations should prove useful for the further study of the genetic basis of saponin biosynthesis and the study of the relationship between saponin production and various aspects of *M. truncatula* biology. A cross between a genotype of the crucifer *Barbarea vulgaris* that was resistant to the flea beetle larvae *Phyllotreta nemorum* and a susceptible genotype was used to identify bioactive compounds involved in defense by assaying flea beetle larvae susceptibility and metabolite levels of the F_2 offspring and identifying metabolites that negatively correlated with larvae survival (Kuzina et al., 2009). Populations of *M. truncatula* resulting from crosses between accessions demonstrating varying levels of resistance to the spotted alfalfa aphid *Therioaphis trifolii* allowed for the characterization of *T. trifolii* resistance as phloem-mediated and involving tolerance and antibiosis and the identification of three quantitative trait loci involved in distinct modes of aphid defense (Kamphuis et al., 2013).

The hemolytic activity of leaf extracts from A17 x ESP105 F_2 plants was quantified using sheep's blood agar Petri plates to screen for inheritance of hemolytic saponin accumulation traits. Three distinct hemolysis phenotypes were observed, a high hemolysis phenotype that resembled the level of hemolysis caused by the A17 parent accession, a low hemolysis phenotype that caused no visible hemolysis similar to the ESP105 parent accession, and a medium hemolysis phenotype that caused an intermediate level of hemolysis between that caused by the A17 and ESP105 parent accessions. The phenotypic ratio of high:medium:low hemolysis was approximately 1:3:1. The presence of the medium hemolysis phenotype and the 1:3:1 phenotypic ratio suggests that inheritance of hemolytic saponin accumulation could depend on incomplete dominance involving more than a single gene with recessive and dominant alleles. Investigation of gene expression in A17 and ESP105 parent accessions and F_2 plants exhibiting high, medium, or low hemolysis phenotypes showed that expression of three genes involved in late steps of hemolytic saponin biosynthesis, CYP716A12, CYP72A67, and CYP72A68, correlated with the observed hemolysis phenotypes. This indicates that accumulation of hemolytic saponins is correlated with expression of genes encoding saponin synthesis enzymes and suggests that the differential regulation of saponin synthesis enzymes that results in high levels of hemolytic saponins in A17 foliar tissue and low levels of saponins in ESP105 foliar tissue occurs at the transcription level.

The association between saponin accumulation and saponin synthesis gene expression has been previously explored by several papers that address methyl jasmonate elicitation in *M. truncatula.* Suzuki et al. (2002) identified and functionally characterized squalene synthase, squalene epoxidase, and β-amyrin synthase enzymes and then developed a methyl jasmonate-induced root cell suspension culture to show that transcript levels of the newly identified enzymes and saponin accumulation levels were both increased after methyl jasmonate treatment. This methyl jasmonate-inducible cell culture system was further optimized by Suzuki et al. (2004) and confirmed by Broeckling et al. (2005) to produce increased levels of triterpene saponins. Achnine et al. (2005) and Naoumkina et al. (2010) were able to identify and functionally characterize three UDP-dependent glycosyltransferase enzymes by using this system to focus on candidate genes with methyl jasmonate-induced expression. Naoumkina et al. (2010) also found that expression of cytochrome P450 gene family members was increased after methyl jasmonate elicitation. Carelli et al. (2011) addressed the association between saponin accumulation and saponin

synthesis gene expression by showing a relationship between expression of CYP716A12 and hemolytic saponin accumulation in leaves, roots, stems, flowers, and pods of *M. truncatula* plants at different developmental stages.

Research on regulation of saponin biosynthesis in *M. truncatula* has been limited. Pollier et al. (2013) discovered an RING membrane anchor-like E3 ubiquitin ligase, makibishi 1 (Mkb1) involved in endoplasmic reticulum-associated degradation that was transcriptionally upregulated in response to methyl jasmonate elicitation concurrently with saponin genes encoding HMGR, squalene synthase, squalene epoxidase, β-amyrin synthase, and CYP93E2. Transgenic *M. truncatula* hairy roots with knocked down Mkb1 accumulated higher amounts of monoglycosylated saponins and lower amounts of higher glycosylated saponins than control roots and had severe morphological deficiencies caused in part by the overaccumulation of bioactive monoglycosylated saponins. The Mkb1 ubiquitin ligase was shown to control the activity of the endoplasmic-reticulum-localized HMGR enzyme that serves as the rate-limiting enzyme in the MVA pathway from which saponin synthesis originates. In control roots, methyl jasmonate application resulted in increased HMGR transcript levels, but HMGR protein levels remained stable, while in Mkb1 knockdown roots HMGR transcript and proteins levels both increased. Pollier et al. (2013) propose that Mkb1 may also manage other endoplasmic-reticulum-associated enzymes involved in saponin synthesis, including cytochrome P450s. While it is possible that Mkb1 or a similar protein could play a role in cytochrome P450 post-transcriptional regulation, this does not explain the vast differences in foliar hemolytic saponin accumulation that occur between A17 and ESP105 plants. The differences in hemolytic saponin accumulation between A17 and ESP105 plants are mirrored by differences in accumulation of transcripts encoding cytochrome P450s involved in hemolytic saponin synthesis, indicating that the regulation responsible for the different levels of hemolytic saponins in the foliar tissue A17 and ESP105 plants occurs at the transcription level.

Carelli et al. (2015) used a population derived from the crossing of two *Medicago* species, *M. sativa* and *M. arborea*, that accumulate differing levels of hemolytic and non-hemolytic triterpene saponins to analyze the variation in saponin accumulation across different progeny and plant organs, specifically leaves and roots, and the mechanisms underlying this variation. Saponin and sapogenin amounts and gene expression of β-amyrin synthase, CYP716A12, CYP72A67, CYP72A68, and

CYP93E2 were evaluated in the leaves and roots of cross progeny. They proposed that sapogenin biosynthesis in the leaves and roots may be controlled by different mechanisms. Hemolytic saponin content was higher and two-times less variable in the roots than in the leaves, suggesting a more stringent control mechanism in the roots. In addition, accumulation of the major hemolytic sapogenins, medicagenic acid, hederagenin, and bayogenin, correlated with accumulation of the non-hemolytic soyasapogenol B, and expression of hemolytic saponin synthesis genes, CYP716A12, CYP72A67, and CYP72A68 correlated with expression of the non-hemolytic saponin synthesis gene CYP93E2 in the roots but not in leaves, suggesting that the hemolytic and non-hemolytic branches of saponin synthesis are more coordinated in the roots than they are in the leaves. This is likely also true in *M. truncatula*, where roots accumulate higher levels of saponins than the aerial organs (Huhman et al., 2005). Moreover, while the ESP105 accession accumulates relatively low levels of saponins in foliar tissue compared to the A17, PRT178, and GRC43 accessions, it accumulates relatively very high levels of saponins in the roots compared to the other accessions (Table 4). Due to this accession variation in saponin accumulation in the roots, the cross populations generated in this thesis research, especially between ESP105 with very high root saponins and PRT178 with intermediate root saponins, could also be used to investigate the mechanisms controlling saponin synthesis in the roots. Carelli et al. (2015) also found that expression levels of the three hemolytic saponin synthesis genes were positively and significantly correlated in both a synchronic analysis of gene expression in 60 cross progeny plants and a diachronic analysis of gene expression in two cross families with variable sapogenin content at three different time points, indicating consistent genetic control of these genes. A comparison of the contents of various aglycones in the foliar tissue of F_1 mother plants and F_2 progeny found a significant correlation between mother and progeny plant aglycone levels for medicagenic acid only. The F_1 versus F_2 relationship was not significant for zanhic acid accumulation, but it was significant when the medicagenic acid:zanhic acid ratio was considered. This indicates that genetic control of the sapogenin pathway in foliar tissue is primarily focused on medicagenic acid, and that conversion of medicagenic acid into zanhic acid may be one method of controlling medicagenic acid levels. We also found that expression patterns of CYP716A12, CYP72A67, and CYP72A68 were similar to each other, all correlating with hemolytic activity of leaf extracts. In addition, A17 foliar accumulation of the non-hemolytic sapogenins, hederagenin, and

bayogenin is approximately three-, nine-, and six-times higher than ESP105 foliar accumulation, respectively, while A17 foliar accumulation of medicagenic and zanhic acid is 22- and 31-times higher than ESP105 foliar accumulation, respectively (Lloyd Sumner, unpublished data). This suggests that the genetic control of the sapogenin pathway responsible for the saponin accumulation differences between the A17 and ESP105 accessions may also be focused on medicagenic acid. As has already been noted, this regulation likely occurs at the transcriptional level. If accumulation of medicagenic acid were the primary focus of saponin synthesis regulation, CYP72A68, the enzyme responsible for medicagenic acid synthesis would be the primary target of possible transcription factors.

Mertens et al. (2016) identified two basic helix-loop-helix family transcription factors, triterpene saponin biosynthesis activating regulator 1 (TSAR1) and TSAR2 that act as specific regulators of triterpene saponin biosynthesis in *M. truncatula*. Overexpression of TSAR1 and TSAR2 in transgenic *M. truncatula* hairy roots resulted in increased expression of genes in the mevalonate pathway, β-amyrin synthase, aglycone-modifying cytochrome P450s, and glycosyltransferases but did not result in increased expression of sterol synthesis genes, demonstrating that TSAR1 and TSAR2 regulation encompasses and is specific to triterpene saponin synthesis. TSAR1 overexpression primarily increased expression of non-hemolytic saponin synthesis genes and production of soyasaponins, while TSAR2 overexpression specifically increased expression of hemolytic saponin synthesis genes and production of hemolytic saponins. Because this research was exclusively performed in hairy roots and mechanisms of saponin biosynthesis regulation may differ between root and leaf tissue (Carelli et al. 2015), it is possible that TSAR1 and TSAR2 may not function in *M. truncatula* foliar tissue in the same way that they were shown to function in hairy roots. Moreover, the 1:3:1 hemolysis phenotype ratio observed among A17 x ESP105 F_2 progeny does not match the standard 3:1 or 1:2:1 phenotypic ratios that indicate that a trait is controlled by one gene with a dominant and recessive allele or with codominant alleles, indicating that hemolytic saponin synthesis in *M. truncatula* foliar tissue is regulated by more than one gene. If TSAR2 does regulate hemolytic saponin synthesis in foliar as well as root tissue, it is likely that there are other contributing transcription factors that have not yet been identified.

Using genetically distinct *M. truncatula* accessions that accumulate varying levels of foliar saponins as a reflection of the different locations and circumstances from which they originate, we were

able to investigate the genetic basis for control of saponin accumulation in leaves. The high conservation of the amino acid sequence for CYP72A68, a cytochrome P450 enzyme involved in modification of hemolytic sapogenins, across the accessions and the correlation between hemolytic saponin levels and expression of hemolytic saponin synthesis genes in the foliar tissue of A17 x ESP105 F_2 progeny suggest that variation in hemolytic saponin accumulation among the accessions is not a result of allelic variation in genes encoding saponin synthesis enzymes, but a result of differential regulation of the expression of the genes encoding saponin synthesis enzymes. Three distinct phenotypes of hemolytic activity were observed among A17 x ESP105 F_2 progeny at a ratio of 1:3:1 indicating that the proposed regulation responsible for the accession differences in foliar saponin accumulation is not controlled by a single gene with dominant and recessive alleles. While our understanding of the regulation of triterpene saponin synthesis in *M. truncatula* has been growing, especially with the recent discovery of transcription factors controlling saponin synthesis gene expression, our research indicates that there additional regulators that have not yet been identified. In the future, the CYP72A68 amino acid sequences from each accession could be expressed in yeast and supplied with substrate to confirm that the proteins function with similar efficiency. Saponin extracts from A17 x ESP105 progeny foliar tissue could be analyzed by HPLC to provide more specific information on differences in accumulation of various sapogenins and individual saponins, and expression of additional saponin synthesis genes, including those involved in soyasaponin synthesis and in earlier and later portions of the pathway, in the same foliar tissue could be analyzed by quantitative PCR to provide additional information on the regulation of saponin synthesis gene expression. Hemolytic saponin accumulation in the foliar tissue of progeny of ESP105 x PRT178 reciprocal crosses could be assayed to confirm what was found in A17 x ESP105 progeny, and saponin accumulation in root tissue of progeny of all cross types could be studied to identify the mechanisms responsible for differential root saponin accumulation in accessions. Cross progeny could be characterized using molecular markers located at regular distances along the chromosomes to identify quantitative trait loci (QTL) for saponin accumulation that could then be sequenced to discover candidate genes responsible for saponin synthesis regulation. Saponins possess many useful properties that make them attractive for applications in industry, agriculture, and human health and are valuable to the plants that produce them due to their activity against a wide variety of plant pests. Engineering of increased or

novel saponin synthesis in various plant species could be used to generate higher yields of commercially or medicinally important saponins or increase production of anti-pest saponins in economically important crops. The genes involved in saponin biosynthesis and the way those genes are regulated must be thoroughly understood before this engineering can be done. The economically important soybean plant produces soyasaponins and has orthologous genes to those involved in *M. truncatula* hemolytic saponin synthesis, but does not produce hemolytic saponins. The genes responsible for positive control of hemolytic saponin synthesis in *M. truncatula* would likely be key components necessary to engineering soybean plants to produce hemolytic saponins that could provide them with an extra layer of pest protection.

Chapter 3

Cytotoxic activity of *Medicago truncatula* triterpene saponins

Introduction

When a pathogen attacks a plant, a complex cascade of signals is initiated that leads to the induction of the specialized metabolism of defense compounds that are specifically tailored to the type of pathogen involved (Pichersky and Lewinsohn, 2011). One family of specialized metabolites is the saponins, which are composed of a hydrophobic aglycone backbone and a polar sugar moiety. Saponins have been shown to have insecticidal, allelopathic, antiviral, antibacterial, nematicidal, and fungicidal properties that provide protection to the plants that produce them. The ability of saponins to permeabilize membranes imparts them with hemolytic activity and may also be partly responsible for saponin anti-pest properties (Augustin et al., 2011). Saponin-membrane interaction involves the spontaneous incorporation of saponins into the membrane outer layer driven by the attraction of the lipophilic aglycone to the lipid interior of the membrane (Lorent et al., 2014). Saponins then form 1:1 complexes with membrane sterols that aggregate to form plaques whose sterical nature results in membrane curvature, leading to the formation of pores that increase membrane permeability (Armah et al., 1999). In the case of erythrocytes, this pore formation can become severe enough to cause lysis of the cell. Various aspects of saponin structure and the type and concentration of cholesterol present in the membrane both affect saponin-membrane interaction (Augustin et al., 2011; Moses et al., 2014a). Some saponins, including the triterpene soyasaponins in *M. truncatula*, do not demonstrate hemolytic activity as a result of structural characteristics (Yoshiki et al., 1998).

Plants containing significant amounts of saponins have been used in the traditional medicines of many cultures for centuries, and in modern times research has shown that specific purified saponin compounds and crude saponin mixtures exhibit a variety of properties useful for improving human health (Güçlü-Üstündağ and Mazza, 2007). There has been substantial interest in the use of saponins as vaccine adjuvants, which are compounds used in combination with the antigens present in vaccines to increase immunity to those antigens. Saponins could be used in either traditional injection-administered vaccines or in orally administered vaccines as they have the ability to increase intestinal cell permeability

to facilitate the uptake of antigens. Orally administered vaccines are cheaper, easier, and less hazardous than injection-administered vaccines, and the intestinal infections that these vaccines could prevent are among the most common causes of mortality and morbidity in humans worldwide Francis et al., 2002). Quil A, isolated from *Q. saponaria* bark, and its derivative QS-21 have been used in experimental vaccines and vaccines in Phase I and II human trials (Sun et al., 2009). Saponins can prevent liver damage through a variety of mechanisms and are already being used to treat patients with hepatitis. The tritepene saponin glycyrrhizin, which is also the primary sweet-tasting component in licorice, is a primary component of Stronger Neo-Minophagen C (SNMC), a drug used in Japan to treat patients with hepatitis. The long term administration of SNMC was shown to reduce incidence of liver carcinogenesis in susceptible patients with chronic hepatitis C by suppressing inflammation (Arase et al., 1997). Oleanolic acid has been shown to decrease cirrhosis occurance in patients with liver diseases and return elevated aminotransferase activity in the serum to normal levels. It is currently used in China as an oral drug to treat acute and chronic hepatitis, as well as other liver disorders (Liu, 1995). Saponins have also been shown to have hypocholesterolemic, hypoglycemic, neuroprotective, and anti-inflammatory properties as well as activity against fungal, viral, and bacterial human pathogens and a therapeutic effect on the cardiovascular system (Francis et al., 2002; Lacaille-Dubois, 2007).

However, most pertinent to this research is the ability of a variety of purfied saponins and crude saponin mixtures from multiple plant species to demonstrate cytotoxic activity against different cancer cell types *in vitro* and *in vivo*. The most obvious mechanism for this activity would seem to be pore formation resulting in necrosis through the same saponin-membrane interaction that causes hemolysis. Anticancer agents that cause necrosis are undesirable due to the resulting inflammation and adverse side effects (Bachran et al., 2008). Fortunately, the cytotoxic activity of saponins appears to be mediated through specific targeting of various cellular processes that is highly dependent on saponin structure and cancer cell type. Saponins have been reported to kill cancer cells and inhibit tumor growth through apoptosis induction, cell cycle arrest, autophagic cell death induction, angiogenesis inhibition, cytoskeleton disintegration, and metastasis inhbition (Podolak et al., 2010). For example, ardisiacrispin (A+B), a mixture of two tritepene saponins isolated from *Ardisia crenata*, was shown to decrease proliferation of Bel-7402 human hepatoma cells by induction of the intrinsic pathway of apoptosis through perturbation of

the mitochondrial membrane and by cytoskeleton disintegration through microtuble disassembly (Li et al., 2008). Diosgenin, a steroid saponin found in fenugreek, inhibited Akt signaling, a pathway upregulated in many aggressive cancers, by downregulating pAkt gene expression and caused G1 cell cycle arrest by downregulating expression of specific cyclin and cyclin-dependent kinase genes in both estrogen receptor positive and negative breast cancer cells without significant toxicity to normal breast epithelial cells. These effects were verified *in vivo*; diosgenin significantly inhibited tumor growth in MCF-7 and MDA-231 breast cancer xenograft nude mice (Srinivasan et al., 2009). Limited research exists on the safety of using saponins as chemotherapeutic drugs. A small portion of studies on the effects of saponins on cancer cells *in vitro* simultaneously test the effects of saponins on corresponding non-cancer cell lines, and the number of *in vivo* studies on saponin cytotoxicity is growing. The safety and tolerability of certain saponins has been studied in rodents and dogs, and saponins generally exhibit acceptable tolerability, especially at therapeutic concentrations (Weng et al., 2001). The most obvious point of concern when considering saponins as chemotherapeutic drugs is their hemolytic activity. This obstacle could be overcome by orally administering chemotherapeutic saponin drugs, as is done in many *in vivo* saponin cytotoxicity studies (Podolak et al., 2010). However, it is also important to consider that some cytotoxic saponins do not have hemolytic activity, and for the saponins that possess both properties, the saponin levels necessary for cytotoxic activity are lower than the levels necessary for hemolytic activity (Mimaki et al., 1998).

Colorectal cancer is the third most common cancer and third leading cause of cancer death for both men and women in the United States (Omoyeni et al., 2015). While colorectal cancer is treatable, especially by surgical removal when diagnosed early, the current standard treatment used when metastasis has occurred, chemotherapy involving 5-fluorouracil and folinate in combination with radiotherapy, results in only a moderate decline in mortality, and risk of disease recurrance remains high (Tong et al., 2011). The Caco-2 cell culture line was originally derived from colorectal adenocarcinoma cells taken from a primary tumor in the colon of a 72-year-old Caucasian male and have been used to test the cytotoxicity of certain compounds against colon cancer cells (Lindl and Steubing, 2013). Saponins from *G. trichophylla, P. pycnantha,* and *G. max* have previously been shown to inhibit Caco-2 cell proliferation *in vitro* (Alam et al., 2015; Omoyeni et al., 2015; Salyer et al., 2013). The annual legume *M.*

truncatula is a close relative of alfalfa grown as winter forage in Australia and often used as a model legume in scientific research (Choi et al., 2004a). Commonly known as the barrel medic, *M. truncatula* produces two types of triterpene saponins, non-hemolytic soyasaponins derived from some of the same sapogenins produced by *G. max* and hemolytic saponins (Moses et al., 2014c). To the author's knowledge, the effects of saponins derived from *M. truncatula* on cancer cells have not previously been studied. The *M. truncatula* species encompasses both commercial varieties that reflect the efforts of selective breeding and diverse native populations growing in the Mediterranean basin that reflect adaptation to the selective pressures of their unique environments. Four genetically distinct accessions that vary in the levels and types of saponins that they produce were used in this study, A17, derived from the commercial variety Jemalong, and ESP105, PRT178, and GRC43 isolated from the wild in Spain, Portugal, and Greece, respectively. Total saponins were extracted from the foliar tissue of the four accessions and analyzed by HPLC-MS to understand accession differences in saponin content that could explain differences in cytotoxic activity. The accession extracts were screened for cytotoxic activity against Caco-2 human colon cancer cells by measuring cell proliferation after treatment with extracts at various timepoints. To begin to understand the mechanism through which *M. truncatula* saponins induce cytotoxicity, levels of caspase-3 activation were measured in treated cells as an indication of apoptosis induction.

Materials and Methods

Plant maintenance

A17, ESP105, PRT178, and GRC43 seeds were scarified using the previously described procedure. The seeds were then thoroughly rinsed with distilled water and placed between two layers of moistened filter paper in a closed container. The container was placed in a dark cabinet for four days to allow for germination. Successfully germinated seedlings were then planted in Sunshine LC1 mix potting medium (SunGro Horticulture Distribution Inc., Bellevue, WA) and grown in a growth chamber at 22.5°C and 60% humidity with a 16:8 (light:dark) hour photoperiod. At 6- to 8-weeks of age, plants were moved to a greenhouse. Plants were watered, fertilized, and treated with a biological larvicide using the previously described procedures. Transplantation to larger pots occurred as needed.

Saponin extraction

A17, ESP105, PRT178, and GRC43 plants (11- to 13-weeks old) were subjected to mechanical damage three times across two days, at 4:00 p.m. on the first day and at 11:00 a.m. and 2:00 p.m. on the second day. Leaves were harvested the third day and stored at -80°C. After being freeze-dried for 48 hours, leaves were pureed in 85% methanol at a ratio of 25 ml 85% methanol per 1 g leaf dry weight. Samples were mixed on a platform shaker for 3.5 hours at room temperature and then filtered through four layers of cheesecloth. The filtered extracts were centrifuged at 5000 RPM for 15 minutes to pellet out remaining leaf particles. Methanol was removed from the extracts through rotary evaporation. For crude purification, a Waters Sep-pak Vac 35cc C18 column was prewashed with two 7.5 ml volumes of 35% methanol and then loaded with 10 ml of extract diluted to a final concentration of 65% methanol. The column was washed with two 7.5 ml volumes each of water and 15% methanol, respectively, before eluting with four 7.5 ml volumes of 100% methanol. Extracts were dried using a stream of nitrogen gas, weighed, resuspended in 100% methanol to a concentration of 5 mg/ml, and stored at -20°C.

High performance liquid chromatography/mass spectrometry

A System Gold high performance liquid chromatograph (Beckman-Coulter, Fullerton, CA, USA) with autosampler (model 508), dual pump (model 126), photodiode array detector (model 168) with

Beckman-Coulter System 32 Karat software (version 8,2006) was used to analyze the saponin samples. The saponins were separated using a Phenomenex (Torrance, CA) Aqua 5μm C18 (250 × 4.6 mm) column and a binary gradient of 0.1% acetic acid for mobile phase A in water and 0.1% acetic acid in acetonitrile for mobile phase B at a flow rate of 0.8 mL/min. The gradient began isocratically from 0-10 min at 1% B, then from 1 to 55% B from 10 to 110 min, from 55 to 100% B from 110 to 140 min, and from 100 to 1% B at 141 min. Characteristic ions were used for peak assignment using HPLC/MS analysis. The HPLC/MS apparatus was interfaced to a Bruker Esquire LC/MS ion trap mass spectrometer. Mass spectral data were collected with the Bruker software, which also controlled the instrument and collected the signal at 200 to 400 nm. Conditions for mass spectral analysis in negative ion electrospray mode for flavonols included a capillary voltage of 4000 V, a nebulising pressure of 30.0 psi, a drying gas flow of 9.0 ml min^{-1} and a temperature of 300°C. Data were collected in full scan mode over a mass range of m/z 200–2000 at 1.0 s per cycle. Peak area was converted into mg hederagenin equivalents using an equation generated by running known amounts of a hederagenin standard.

Preparation of saponin extracts for cell treatment

Extracts were placed under a thin stream of nitrogen gas to remove methanol. HPLC-MS quantification found varying mg hederagenin equivalents per g dried extract concentrations for the four accessions (Table 3). Because of the imprecision of using hederagenin as a standard for a mixture of saponins, the GRC43 saponin extract was quantified as having more than 1 g of saponins in 1 g of dried extract. To correct for this, all of the mg hederagenin equivalents per g dried extract concentrations were divided by the concentration of the GRC43 extract. Enough dried extract was then resuspended in dimethyl sulfoxide (DMSO) using the corrected mg hederagenin equivalents per g dried extract concentrations to ensure a saponin concentration of 83.33 mg/ml for all 4 accession treatments.

Table 3. Corrected saponin concentrations in crude extracts

Accession	mg Hederagenin Equivalents/g Dried Extract	Corrected mg Hederagenin Equivalents/g Dried Extract
GRC43	1103.44	1
PRT178	993.07	899.98
A17	768.46	696.42
ESP105	630.92	571.78

Cell maintenance

Caco-2 human colon cancer cells were obtained from the American Type Culture Collection (Manassas, Virginia) and maintained in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum, 2% non-essential amino acids, and 1% antibiotic-antimycotic in a 75 cm^3 flask at 37°C in a humidified 5% CO_2 atmosphere. All media components were obtained from Gibco® through Life Technologies (Carlsbad, CA). Media in the flasks was changed every three to four days, and cells were passaged into two new flasks at 80% confluence, one to two times per week.

Cell proliferation assay

Cells at passage numbers 21-33 were used for proliferation assays. Caco-2 cells in 100 µl of 10% fetal bovine serum media were seeded into the wells of four 96-well plates at a density of 2 x 10^3 cells per well and incubated at 37°C in a humidified 5% CO_2 atmosphere. A serial dilution of cells (0-3.2 x 10^4) was also seeded in triplicate. After stable attachment (48 hrs), media was removed and replaced with 200 µl test media containing saponins extracted from A17, ESP105, PRT178, or GRC43 *M. truncatula* plants at a concentration of 100 µg/ml or 250 µg/ml, with a final DMSO concentration of 0.3%. Six wells were used per treatment type. For the negative control, three wells of cells were treated with 200 µl test media containing DMSO at a concentration of 0.3%. Three cell-free wells were also filled with 200 µl test media containing saponins or DMSO for each treatment type. Wells with serial dilution of cells were filled with 200 µl of fresh media.

Cell proliferation was measured using the CellTiter 96® AQeuous One Solution Cell Proliferation Assay (Promega Corp. Madison, WI). Proliferation was measured at 2, 24, 48, and 72 hours post-treatment. Assay reagent (40 µl) was directly added to all wells and incubated for 2 hours at 37°C in a

humidified 5% CO_2 atmosphere before measuring absorbance at 490 nm. Absorbance of wells containing

cells was corrected by subtracting the background absorbance of cell-free wells with the corresponding

test media. Absorbance was then converted into cell numbers using a standard curve equation obtained

by measuring the absorbance of wells containing the serial dilution of cells. The standard curve data for

one of the three independent experiments is depicted in Figure 11.
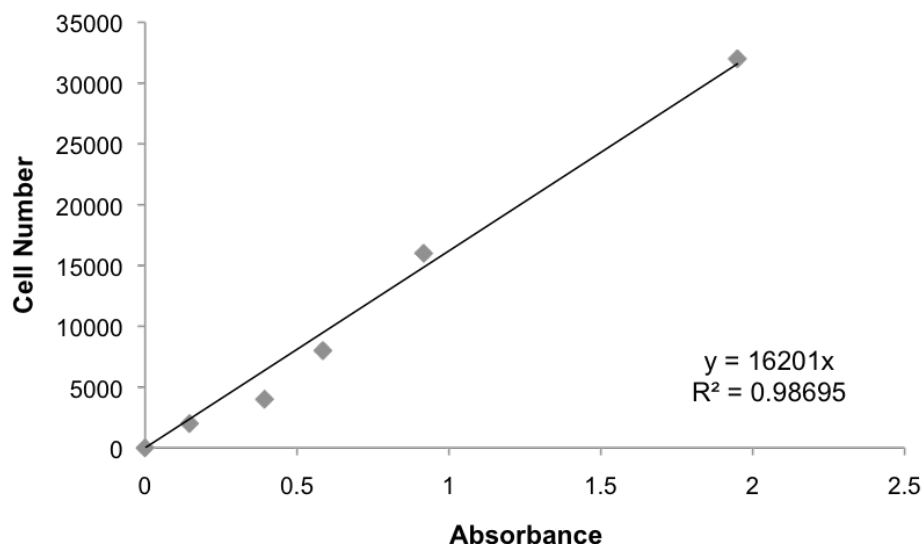
Figure 11. Standard curve for calculation of cell number. Data is from one of the three independent cell proliferation assays performed. Equation was used to convert absorbance readings of treatment wells into number of live cells present. Absorbance readings were strongly correlated with cell number, R^2=0.987.

Caspase-3 assay

Cells at passage numbers 35-38 were used for caspase-3 assays. Caspase-3 activity was

measured using the Caspase-3 Fluorescence Assay Kit (Cayman Chemical Company, Ann Arbor, MI).

Caco-2 cells in 100 µl of 10% fetal bovine serum media were seeded into the wells of a 96-well plate at a

density of 3 x 10^4 cells per well and incubated at 37°C in a humidified 5% CO_2 atmosphere. After stable

attachment (48 hrs), media was removed and replaced with 200 µl test media containing saponins

extracted from A17, ESP105, PRT178, or GRC43 *M. truncatula* plants at a concentration of 100 µg/ml,

with a final DMSO concentration of 0.12%. For the negative control, cells were treated with test media

containing DMSO at a concentration of 0.12%. Seven wells were used per treatment type.

Caspase-3 activity was measured 48 hours post-treatment. The 96-well plate was centrifuged for

5 min at 1000 rpm. Test media was removed and replaced with 200 µl of assay buffer, and the plate was

centrifuged again for 5 min at 1000 rpm. The assay buffer was then removed and replaced with 100 µl

lysis buffer. The plate was mixed on an orbital shaker at room temperature with gentle shaking for 30 min

before being centrifuged for 10 min at 1000 rpm. The supernatant (90 µl) was then transferred to the

corresponding wells of a black 96-well plate. Assay buffer (10 µl) was added to 6 wells per treatment type,

while 10 µl of Caspase-3 inhibitor solution was added to the remaining well of each treatment type to

verify substrate specificity. Active caspase-3 standard was serially diluted and 100 µl of each dilution was

added to the plate. Caspase-3 substrate solution (100 µl) was then added to each well, and the plate was

incubated for 30 min at 37°C in a humidified 5% CO_2 atmosphere. The fluorescent intensity of each well

was measured at 485 nm excitation and 528 nm emission. Fluorescence measurements were converted

into caspase-3 U/well using an equation obtained from measuring the fluorescence of the serially diluted

active caspase-3 standard (Figure 12).
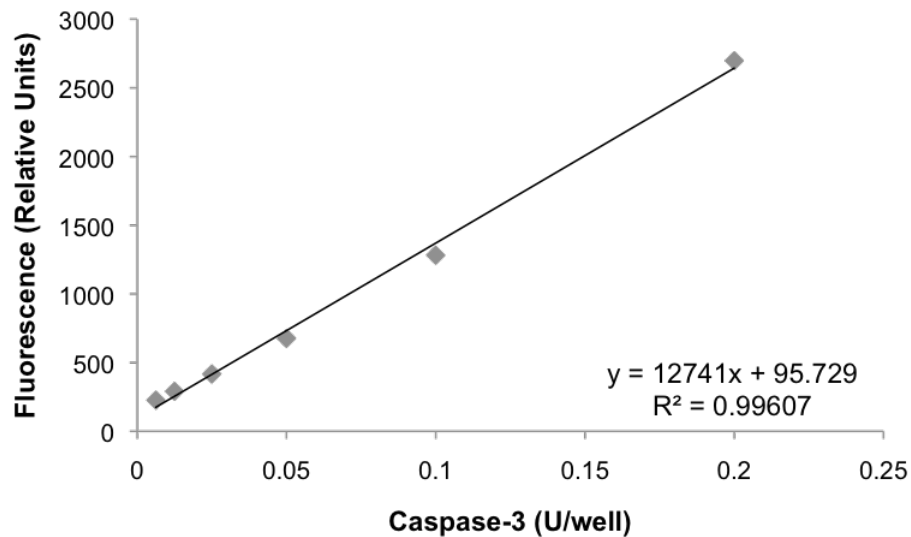
Figure 12. Standard curve for calculation of caspase-3 U/well. Data is from one of the three independent caspase-3 activation assays performed. Equation was used to convert fluorescence readings of treatment wells into caspase-3 U/well. Fluorescence readings were strongly correlated with caspase-3 U/wll, R^2=0.996.

Statistical analysis

GraphPad Prism (GraphPad Software Inc., La Jolla, CA) was used to analyze cell proliferation and caspase-3 activation data. For cell proliferation data, means were compared to each other by one-way ANOVA followed by Tukey's multiple comparison tests between every possible pair of treatments and then compared to 100% by Student's *t* test to assess significant difference from the solvent-treated control. Significance for all tests was defined as $P<0.05$. For caspase-3 activation data, treatment means were compared to each other and a solvent-treated control by Student's *t* test ($P<0.05$).

M. truncatula accessions accumulate varying levels and types of saponins in leaf tissue

The four accessions of *M. truncatula* used in this study were chosen due to their natural variation in saponin accumulation profiles. Triterpene saponin levels in the roots and foliar tissue of many *M. truncatula* ecotypes, including the four accessions of interest (Table 4), have been described by metabolic profiling performed at the Samuel Roberts Noble Foundation (Lloyd Sumner, unpublished data).

Table 4. Reported relative basal levels of triterpene saponins among accessions of *M. truncatula*

	Tissue	
Accession	Foliar	Root
A17	High	High
ESP105	Low	Very High
PRT178	High	Intermediate
GRC43	Very High	Low

Table 4. Reported relative basal levels of triterpene saponins among accessions of *M. truncatula.* Values are based on unpublished metabolic profiles (Lloyd Sumner and David Huhman, Samuel Roberts Noble Foundation, personal communication).

Saponin profiles of foliar tissue from the four accessions of interest were investigated by HPLC mass spectrometry to allow for accurate saponin concentration values for cell culture treatments and information that could explain differences in cytotoxicity among accession treatments. Known amounts of a hederagenin standard were also analyzed to create a standard curve for converting area under the curve for peaks from the crude extract analysis into mg hederagenin equivalents. The amounts of saponins in the crude extracts are reported as mg hederagenin equivalents per g dried extract.

Extracts from A17, ESP105, PRT178, and GRC43 foliar tissue contained varying levels of total saponins. Because saponins were not extracted with an internal standard, some variation may be a result of error in the extraction process and not a true reflection of saponin accumulation levels in foliar tissue. However, total saponin levels roughly agreed with the relative levels previously reported (Table 4); GRC43 extracts had the highest concentration of saponins, while A17 and PRT178 extracts had lower levels of saponins but maintained higher saponin concentrations than ESP105, which had the lowest

concentration of saponins. Because hederagenin is not a perfect standard for the quantification of a mixture of different saponins, GRC43 extracts were characterized as having more than 1 g of saponins in 1 g of dried extract. This was accounted for in preparation of cell culture treatments, as described in Materials and Methods.
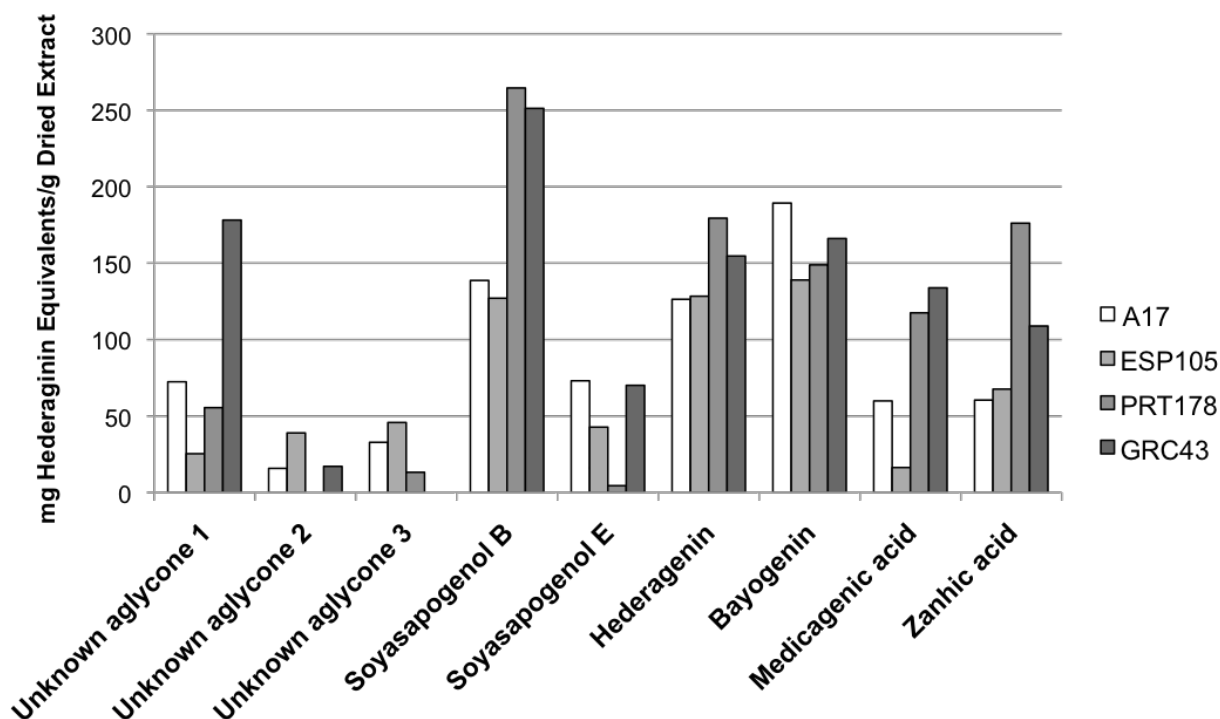
Figure 13. Saponin profiles in *M. truncatula* accessions. Sapogenin accumulation levels in extracts from mechanically damaged *M. truncatula* A17, ESP105, PRT178, and GRC43 leaf tissue were quantified by HPLC-MS using a hederagenin standard.

Saponin amounts for the most abundant sapogenins were pooled and compared in Figure 13 and Table 5. The most abundant sapogenin in both GRC43 and PRT178 extracts was soyasapogenol B, while the most abundant sapogenin in A17 and ESP105 extracts was bayogenin. The second most abundant sapogenin in the GRC43 extract was an unknown sapogenin labeled as aglycone 1, followed by fairly similar amounts of the hemolytic sapogenins, hederagenin, bayogenin, and medicagenic and zanhic acid, and then low levels of soyasapogenol E and unknown aglycone 2. The second most abundant sapogenin in the PRT178 extract was hederagenin, closely followed by zanhic acid and then bayogenin and medicagenic acid. Unknown aglycone 1, soysasapogenol E, and unknown aglycone 3

were detected at low levels. The second most abundant sapogenin in the A17 extract was soyasapogenol B, followed by hederagenin, and then roughly equivalent amounts of unknown aglycone 1, soyasapogenol E, and medicagenic and zanhic acid. Low levels of unknown aglycone 2 and 3 were detected. The second most abundant sapogenin in the ESP105 extract was hederagenin, followed by an almost equal amount of soyasapogenol B, intermediate amounts of zanhic acid, unknown aglycone 3, soyasapogenol E, and unknown aglycone 2, and low levels of unknown aglycone 1 and medicagenic acid. In general, the accession extracts contained a higher proportion of hemolytic sapogenins than soyasapogenins and high levels of soyasapogenol B, hedergenin, and bayogenin, intermediate levels of medicagenic and zanhic acid, and low levels of soyasapogenol E and unknown aglycones. Exceptions include the high proportion of unknown aglycone 1 in GRC43 and zanhic acid in PRT178 and the low proportion of medicagenic acid in ESP105.

Table 5. Sapogenin amounts in *M. truncatula* accessions as a percent of total saponins

Sapogenin	% of total saponins			
	A17	ESP105	PRT178	GRC43
Unknown aglycone 1	9.42	4.01	5.58	16.14
Unknown aglycone 2	2.05	6.17	n.d.	1.55
Unknown aglycone 3	4.27	7.25	1.33	n.d.
Soyasapogenol B	18.04	20.14	26.64	22.77
Soyasapogenol E	9.50	6.77	0.45	6.35
Hederagenin	16.44	20.34	18.06	14.01
Bayogenin	24.63	22.02	14.99	15.05
Medicagenic acid	7.79	2.58	11.83	12.13
Zanhic acid	7.86	10.71	17.73	9.87

Table 5. Sapogenin amounts in *M. truncatula* accessions as a percent of total saponins. Crude saponin extracts from A17, ESP105, PRT178, and GRC43 foliar tissue were analyzed using HPLC-MS. Saponin amounts were combined by sapogenin type and compared to the total saponin amount to generate a percentage.

Individual saponin components of the extracts were identified and quantified by HPLC-MS (Table 6). Most of the glycosidically attached sugar constituents were characterized as unspecified hexose sugars. However, pentose, galactose, glucose, galacturonic acid, rhamnose, arabinose, apiofuranose,

and xylose attachments were more specifically identified as modifying multiple sapogenin backbones.

Many saponins were uniquely identified in only one accession extract. The A17 extract was the most

diverse with 39 different saponins identified and eight of those unique to A17, while the ESP105 extract

was the least diverse with only 21 different saponins. 3-Glc-medicagenic acid and dHex-hex-HexA-

soyasapogenol B were the only saponins uniquely identified in the ESP105 extract. The PRT178 extract

contained 30 different saponins, including six unique to PRT178, and the GRC43 extract contained 32

saponins, including eight unique to GRC43. Only seven saponins were present in all four accessions:

dHex-Hex-Hex-HexA-soyasapogenol B, 3-Rha-Gal-GlcA-soyasapogenol B, 3-Rha-Xyl-GlcA-

soyasapogenol B, Malonyl-Hex-Hex-HexA-soyasapogenol E, Hex-bayogenin, 3-Glc-Glc-28-Ara-Rha-Xyl-

Api-zanhic acid, and 3-Glc-28-Ara-Rha-Xyl-zanhic acid. One of those saponins, dHex-Hex-Hex-HexA-

soyasapogenol B, was also the most abundant saponin detected in the A17, PRT178, and GRC43

extracts. The most abundant saponin in the ESP105 extracts was Hex-dHex-Hex-HexA-hederagenin.

Table 6. Mass spectral identification of saponins

Saponin	mg hederagenin equivalents/g dried extract			
	A17	ESP105	PRT178	GRC43
Hex-Hex-HexA-unknown aglycone 1	15.27	n.d.	54.62	81.09
Hex-HexA-unknown aglycone 1	36.56	n.d.	0.83	22.29
Malonyl-Hex-Hex-HexA-unknown aglycone 1	7.57	n.d.	n.d.	36.98
Hex-HexA-Malonyl-unknown aglycone 1	n.d.	25.33	n.d.	37.78
Malonyl-Hex-Hex-Hex-HexA-unknown aglycone 1	12.98	n.d.	n.d.	n.d.
dHex-Hex-HexA-unknown aglycone 2	15.75	38.93	n.d.	17.06
Hex-Hex-unknown aglycone 3	32.79	19.23	n.d.	n.d.
Hex-unknown aglycone 3	n.d.	26.53	13.20	n.d.
Pen-Hex-Hex-unknown aglycone 4	n.d.	n.d.	33.70	n.d.
Hex-Hex-unknown aglycone 5	n.d.	n.d.	n.d.	6.99
dHex-Hex-Hex-HexA-soyasapogenol B	65.26	34.22	158.58	118.02
3-Rha-Gal-GlcA-soyasapogenol B	23.36	18.74	73.80	73.26
3-Rha-Xyl-GlcA-soyasapogenol B	25.30	17.01	2.55	30.22
Pen-Hex-HexA soyasapogenol B	10.79	n.d.	29.61	n.d.
Malonyl-dHex-Hex-HexA-soyasapogenol B	13.95	n.d.	n.d.	4.11
dHex-hex-HexA-soyasapogenol B	n.d.	57.08	n.d.	n.d.
Malonyl-Hex-Hex-HexA-soyasapogenol B	n.d.	n.d.	n.d.	25.58
dHex-Hex-HexA-soyasapogenol A	n.d.	n.d.	n.d.	16.58
Malonyl-Hex-Hex-HexA-soyasapogenol E	18.99	18.86	3.98	19.62
3-Rha-Gal-GlcA-soyasapogenol E	n.d.	23.87	n.d.	50.42
Rha-Hex-Hex-soyasapogenol E	12.09	n.d.	n.d.	n.d.
Hex-Hex-HexA-soyasapogenol E	24.98	n.d.	n.d.	n.d.
Hex-HexA-soyasapogenol E	16.94	n.d.	n.d.	n.d.
Hex-dHex-Hex-HexA-soyasapogenol E	n.d.	n.d.	0.44	n.d.
dHex-Hex-Hex-hederagenin	20.96	20.06	63.04	n.d.

Table 6. (Cont.)

Saponin	mg hederagenin equivalents/g dried extract			
	A17	ESP105	PRT178	GRC43
Hex-Hex-HexA-hederagenin	13.39	43.31	n.d.	65.90
Malonyl-Hex-Hex-HexA-hederagenin	39.90	n.d.	24.75	51.36
dHex-Hex-HexA-hederagenin	30.25	n.d.	16.71	n.d.
dHex-Hex-Hex-Hex-hederagenin	21.85	n.d.	36.26	n.d.
Hex-dHex-Hex-HexA-hederagenin	n.d.	64.97	18.28	n.d.
Malonyl-Hex-HexA-hederagenin	n.d.	n.d.	20.33	n.d.
Ara-Glc-Ara-hederagenin	n.d.	n.d.	n.d.	15.50
Malonyl-Hex-Malonyl-Hex-hederagenin	n.d.	n.d.	n.d.	21.87
Hex-bayogenin	26.42	25.11	31.63	19.33
dHex-Hex-HexA-bayogenin	13.35	55.14	66.49	n.d.
Hex-Hex-bayogenin	56.30	25.95	n.d.	62.63
Hex-dHex-Hex-Hex-bayogenin	17.51	32.74	n.d.	33.10
Hex-HexA-bayogenin	16.98	n.d.	23.11	n.d.
Hex-Hex-Rha-bayogenin	18.41	n.d.	27.59	n.d.
Malonyl-Hex-HexA-bayogenin	6.97	n.d.	n.d.	26.26
Malonyl-Hex-bayogenin	11.55	n.d.	n.d.	n.d.
Malonyl-Hex-Hex-HexA-bayogenin	21.79	n.d.	n.d.	n.d.
Hex-Hex-HexA-bayogenin	n.d.	n.d.	n.d.	24.76
3-Glc-Glc-28-Ara-Rha-Xyl-medicagenic acid	7.67	n.d.	18.52	60.85
3-Glc-Glc-Rha-28-Glc-medicagenic acid	8.67	n.d.	15.41	n.d.
3-GlcA-28-Ara-Rha-Xyl-medicagenic acid	35.05	n.d.	n.d.	6.70
3-Glc-28-Ara-Rha-Xyl-medicagenic acid	8.45	n.d.	n.d.	15.45
3-Glc-medicagenic acid	n.d.	16.29	n.d.	n.d.
Malonyl-Hex-Malonyl-Hex-medicagenic acid	n.d.	n.d.	56.77	n.d.
3-GlcA-28-Glc-medicagenic acid	n.d.	n.d.	26.75	n.d.
3-Glc-Glc-medicagenic acid	n.d.	n.d.	n.d.	18.76
Malonyl-Hex-Hex-medicagenic acid	n.d.	n.d.	n.d.	32.07
3-Glc-Glc-28-Ara-Rha-Xyl-Api-zanhic acid	8.24	32.84	54.54	25.93
3-Glc-28-Ara-Rha-Xyl-zanhic acid	7.54	16.57	33.11	33.72
3-Glc-Glc-28-Ara-Rha-Xyl-Ara-zanhic acid	32.69	n.d.	35.25	24.60
3-Glc-Glc-28-Ara-Rha-Api-zanhic acid	7.68	n.d.	20.09	24.62
3-Glc-Glc-28-Ara-Rha-Ara-zanhic acid	n.d.	18.16	22.07	n.d.
3-Glc-Glc-Glc-23-Ara-28-Ara-Rha-Xyl- zanhic acid	4.28	n.d.	n.d.	n.d.
3-Glc-Glc-28-Ara-Rha-Xyl-zanhic acid	n.d.	n.d.	11.05	n.d.
Total	**768.46**	**630.92**	**993.07**	**1103.44**

Table 6. Mass spectral identification of saponins. Crude saponin extracts from A17, ESP105, PRT178, and GRC43 foliar tissue were analyzed by HPLC-MS. Concentrations were calculated using a standard curve obtained by running known amounts of a hederagenin standard. Abbreviations: Api, apiofuranose; Ara, arabinose; dHex, 6-deoxyhexose such as rhamnose or furanose; Gal, galactose; Glc, glucose; GlcA, galacturonic acid; Hex, hexose such as glucose or galactose; HexA, uronic acid such as glucuronic acid or galacturonic acid; n.d., not detected; Pen, pentose; Rha, rhamnose; Xyl, xylose.

Treatment with *M. truncatula* saponin extracts results in decreased Caco-2 cell proliferation

Caco-2 cell cultures were treated with crude saponin extracts from the four *M. truncatula* accessions at a concentration of 100 or 250 µg/ml final concentration. Percent viability was calculated by measuring the number of live cells present at 2, 24, 48, and 72 hours post-treatment and comparing to the solvent-treated control at the corresponding time point. The mean percent viability for three independent experiments with six replicates each (n=18) was calculated for each treatment.
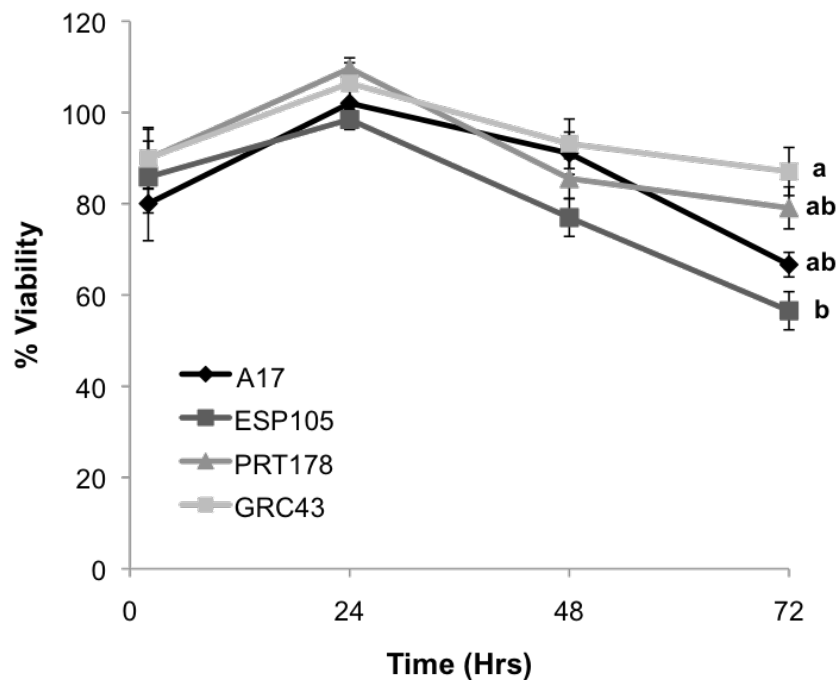

Figure 14. Percent cell viability for Caco-2 cells treated with A17, ESP105, PRT178, and GRC43 saponin extracts at 100 µg/ml 10% FBS media. Values are means ± SEM (n=18), as compared to solvent-treated cultures. Means were compared to each other by one-way ANOVA followed by Tukey's multiple comparison tests. Means marked with different letters at 72 hr are significantly different (P<0.05). Means were also compared to 100% by Student's t test. By 72 hr, all extracts caused significant reduction in cell viability (P<0.05).

Treatment with saponin extracts from all four *M. truncatula* accessions at 100 µg/ml concentration resulted in significantly decreased Caco-2 cell proliferation compared to the solvent-treated control by 72 hr (P<0.05)(Figure 14). The earliest time point tested was 2 hr, which reflects the minimum timespan necessary for treatment of the cultures and conducting the assay. At 2 hr post-treatment, only A17 extracts significantly decreased cell proliferation as compared to the control (80.0%)(P=0.025). Apparent

differences at this time point are more likely a result of inconsistencies in cell seeding number than an actual effect of treatments, due to the difficulty of achieving uniformity when dealing with live cells. After a slight increase in cell viability at 24 hr, treatment with ESP105 and PRT178 extracts resulted in significantly decreased cell proliferation as compared to the solvent-treated control at 48 hr ($P<0.01$), 77.0 and 85.5% viability, respectively, and by 72 hr all extracts had significantly reduced cell viability as compared to the solvent-treated control, A17 to 66.7%, ESP105 to 56.6%, PRT178 to 79.1%, and GRC43 to 87.1% ($P<0.05$). ESP105 extract had a significantly stronger effect than GRC43 on Caco-2 cell proliferation by 72 hr ($P<0.01$).
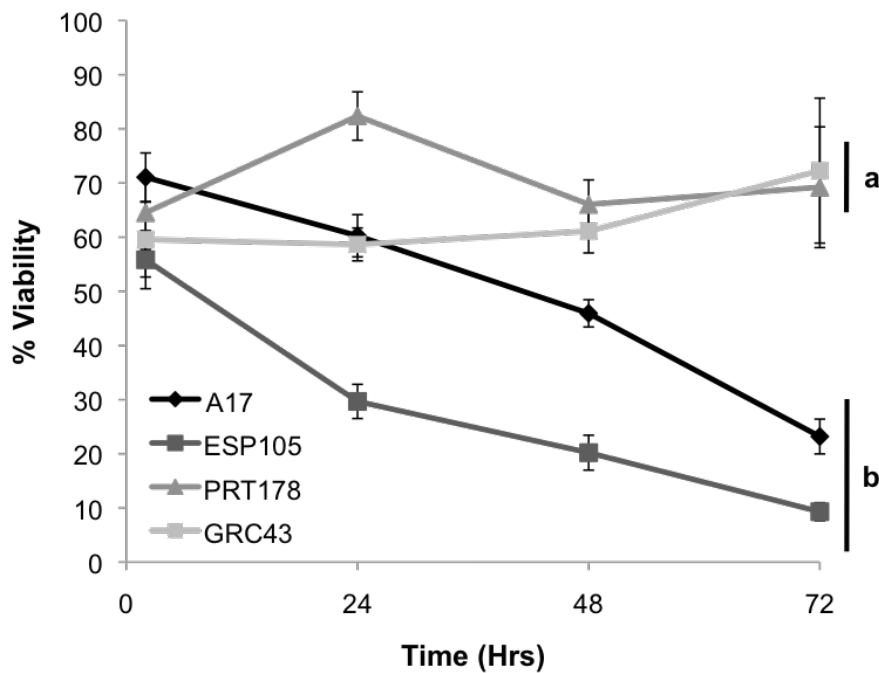
Figure 15. Percent cell viability for Caco-2 cells treated with A17, ESP105, PRT178, and GRC43 saponin extracts at 250 μg/ml 10% FBS media. Values are means ± SEM (n=18), as compared to solvent-treated cultures. Means were compared to each other by one-way ANOVA followed by Tukey's multiple comparison tests. Means marked with different letters at 72 hr are significantly different ($P<0.05$). Means were also compared to 100% by Student's t test. All extracts at all time points, except GRC43 at 72 hr, caused significant reduction in cell viability ($P<0.05$).

Treatment with saponin extracts from all four *M. truncatula* accessions at 250 μg/ml resulted in significantly decreased levels of Caco-2 cell proliferation compared to the solvent-treated control for all

time points and treatments except GRC43 treatment at 72 hr (P<0.05) (Figure 15). At 2 hr, all extracts significantly decreased percent cell viability, A17 to 71.1%, ESP105 to 55.8%, PRT178 to 64.5%, and GRC43 to 59.6% (P<0.0001). By 24 hr, ESP105 extracts had a significantly stronger effect (29.7% viability) on Caco-2 cell proliferation than A17 (60.3%) and PRT178 (82.4%) extracts (P<0.05), but not GRC43 (58.7%) extracts. However, by 48 hr, cell viability under GRC43 treatment rose to 61.1%, continued to rise to 72.3% 72 hr post-treatment, and was no longer significantly different from the 100% cell viability of the solvent-treated control. Cell viability under PRT178 treatment moderately declined to 66.1% at 48 hr and then slightly increased to 69.2% at 72 hr. Treatment with A17 and ESP105 extracts resulted in a decrease in cell viability at 48 hr, 45.9 and 20.1% respectively, with ESP105 extracts having a significantly stronger effect than PRT178 and GRC43 extracts (P<0.001). This trend continued at 72 hr, with A17 and ESP105 extracts resulting in a decrease in cell viability to 23.2 and 9.3% respectively. At 72 hr, both A17 and ESP105 extracts had a significantly stronger impact on Caco-2 cell proliferation than PRT178 and GRC43 extracts (P<0.001).

Treatment with A17, ESP105, PRT178, and GRC43 saponin extracts decreased Caco-2 cell proliferation in a time and dose-dependent manner. The 250 µg/ml treatment had a greater effect on cell proliferation than the 100 µg/ml treatment of extracts from all four accessions. Under 100 µg/ml treatment, cell viability increased until 24 hr, and the treatments did not result in decreased cell proliferation until 48 hr. Under 250 µg/ml treatment, the four accession treatments all initially decreased cell proliferation at 2 hr, but treatment with A17 and ESP105 extracts resulted in a continuing trend of decreasing percent cell viability until 72 hr, while cell viability under treatment with PRT178 and GRC43 extracts was more erratic and seemed to eventually overcome treatment, increasing slightly from 48 to 72 hr. At both concentrations, the ESP105 extract had the strongest effect on Caco-2 cell proliferation, followed by the A17, then PRT178, and finally GRC43 extracts.

Treatment with *M. truncatula* saponin extracts does not result in increased activation of caspase-3

Many saponins that have been shown to decrease cancer cell proliferation do so through the induction of apoptosis (Bachran et al., 2008; Podolak et al., 2010). Caspase-3 is an enzyme essential to the execution phase of apoptosis that can be activated by either the extrinsic or intrinsic pathway and has

been shown to be mutated in multiple cancers (Ghavami et al., 2009). Caco-2 cell cultures seeded in a

96-well plate were treated with crude saponin extracts from the four *M. truncatula* accessions at a

concentration of 100 µg/ml. After 48 hours, cells were lysed and the amount of active caspase-3 present

was measured. The mean caspase-3 U/well for three independent experiments with three or six

replicates each (n=15) was calculated for each treatment.
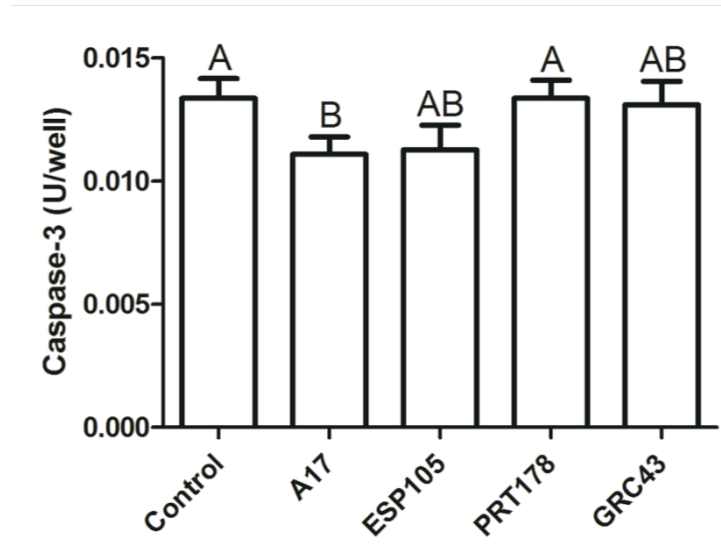
Figure 16. Active caspase-3 (U/well) in Caco-2 cells treated with A17, ESP105, PRT178, and GRC43 saponin extracts at 100 µg/ml in 10% FBS media for 48 hr. Control was treated with solvent. Values are means ± SEM (n=15). Bars marked with different letters are significantly different as tested by Student's *t* test (P<0.05).

Treatment with saponin extracts from all four *M. truncatula* accessions at 100 µg/ml did not result

in significantly increased levels of active caspase-3 compared to the solvent-treated control (P<0.05)

(Figure 16). The mean amount of active caspase-3 ranged from 0.011 to 0.013 U/well. The amount of

active caspase-3 in wells treated with ESP105, PRT178, and GRC43 extracts was not significantly

different from the solvent-treated control. A17 treatment resulted in a significantly different, but

decreased, amount of active caspase-3 when compared to the solvent-treated control (P=0.0443).

Discussion

Saponins have been shown to possess a variety of properties that could benefit human health, including neuroprotective, hypoglycemic, hepatoprotective, cytotoxic, hypocholesterolemic, and immunostimulatory activities, as well as therapeutic effects on the cardiovascular system and activity against viral, fungal, and bacterial human pathogens (Francis et al., 2002; Lacaille-Dubois, 2007). Many of these beneficial effects of saponins can be obtained through consumption of saponin-rich foods or oral administration of saponin supplements. A supplement called Cholestaid™ containing 900 mg of saponin-rich *M. sativa* extract and 100 mg citric acid that claims to "neutralize" cholesterol in the stomach and facilitate its excretion has been made available in the United States (Bora and Sharma, 2011). Mice who were fed a high-fat diet supplemented with a saponin-rich petroleum ether *Y. schidigera* extract demonstrated improved hormone levels, lipid profile, and biochemical parameters compared to mice that were fed an unsupplemented high-fat diet, indicating that *Y. schidigera* extracts could be used as preventative medicine in nutritional disorders such as obesity (Kucukkurt et al., 2016). Commercial saponin products, including a soybean concentrate and a Korean ginseng extract called saponia, have been marketed as functional food products and nutraceuticals, and a method has been developed for fortifying food products, such as olive oil, with oleanolic acid due to its presumed health benefits (Güçlü-Üstündağ and Mazza, 2007). Frequent consumption of soybeans and soybean-based products is correlated with reduced cancer incidence, especially for colon cancer, and this may be due to their relatively high saponin content (Li et al., 2015). Soybean-derived saponins have been shown to have cytotoxic activity in multiple studies (Ellington et al., 2005; Gurfinkel and Rao, 2003; Kim et al., 2004; Oh and Sung, 2001; Salyer et al., 2013).

Saponins from a variety of plant sources in the form of purified compounds and crude mixtures have been reported as having cytotoxic activity against a variety of cancer cell types. This effect is mediated through a variety of complex mechanisms dependent on both saponin structure and cancer cell type (Bachran et al., 2008; Podolak et al., 2010). *M. truncatula*, an annual legume, accumulates saponins in foliar tissue constitutively and in response to herbivory as a means of plant defense (Huhman and Sumner, 2002; Gholami et al., 2014). The *M. truncatula* accessions A17, ESP105, PRT178, and GRC43 have originated from different locations and circumstances and accumulate different levels and types of

saponins as a result of selective breeding or adaptation to the unique challenges of their environments (Delalande et al., 2007; Rogers, 2015). We hypothesized that crude mixtures of saponins extracted from the four *M. truncatula* accessions would differentially exhibit cytotoxicity activity against Caco-2 human colon cancer cells grown in culture. *M. truncatula* crude saponin extracts significantly decreased Caco-2 cell proliferation when administered at a final concentration of both 100 and 250 μg/ml. Extract from ESP105 foliar tissue exhibited the strongest effect on cell proliferation, followed by A17, PRT178, and GRC43 extracts, respectively.

Crude saponin extracts from A17, ESP105, PRT178, and GRC43 foliar tissue were analyzed by HPLC-MS. The concentration of pure saponins in the extracts was quantified to allow for treatment of cell cultures with an accurately measured amount of saponins that was equal across the four accession treatments. Individual saponins within the mixture were putatively identified to provide information on saponins that may primarily contribute to cytotoxicity of the extracts and could be candidates for further purification and study. Glycosides of 12 different sapogenin backbones were detected. Soyasapogenol B and E, hederagenin, bayogenin, and medicagenic and zanhic acid have been reported as being present in *M. truncatula* foliar tissue (Huhman and Sumner, 2002; Kapusta et al., 2005a), while the unknown sapogenins labeled as aglycone 1-2 and 4-5 and soyasapogenol A have only previously been detected in *M. truncatula* hairy roots (Pollier et al., 2011a). Aglycones 1 and 4 (referred to as Aglycones A and B by Pollier et al.) have been tentatively identified as 2β,3β-dihydroxy-23-oxo-olean-12-en-28-oic acid and gypsogenin (2011a). The backbone 2β,3β-dihydroxy-23-oxo-olean-12-en-28-oic acid is proposed to be a biosynthetic intermediate resulting from the oxidation of bayogenin at the C-23 position, which is then itself oxidized at the same position to form medicagenic acid (Tava et al., 2011). Glycosides of 2β,3β-dihydroxy-23-oxo-olean-12-en-28-oic acid have been isolated from roots of *M. hybrida* and leaves of *M. arborea* and *M. arabica* (Bialy et al., 2006; Tava et al., 2005; Tava et al., 2009). Gypsogenin has been proposed as a biosynthetic intermediate resulting from oxidation of hederagenin at the C-23 position, which is then itself oxidized at the C-2 position to form 2β,3β-dihydroxy-23-oxo-olean-12-en-28-oic acid (Gholami et al., 2014). A glycoside of gypsogenin was detected in the root tips and border cells of *M. truncatula* A17 plants (Watson et al., 2015). The structures of aglycones 2 and 5 (referred to as

Aglycones D and C by Pollier et al.) could not be determined, but their calculated molecular formulas are

$C_{30}H_{44}O_6$ and $C_{30}H_{50}O_5$, respectively (2011a). Unknown aglycone 3 has not previously been identified.

The most abundant sapogenin in A17 and ESP105 plants was bayogenin, while the most abundant sapogenin in PRT178 and GRC43 plants was soyasapogenol B (Table 5). In general, soyasapogenol B, hederagenin, and bayogenin were the major components of the extracts and medicagenic and zanhic acid were present at slightly smaller but still sizable proportions. Unknown aglycones 1-5 and soyasapogenol A and E were generally minor components of the extracts. Notable exceptions include the relatively high proportion of zanhic acid and unknown aglycone 1 in PRT178 and GRC43 extracts, respectively, and the relatively low proportion of medicagenic acid in ESP105 extracts. This disagrees with the existing literature, which reports that glycosides of medicagenic acid are the primary group of saponins accumulating in *M. truncatula* A17 foliar tissue and that glycosides of hederagenin and bayogenin are only present as trace amounts (.3-2.9%)(Huhman et al., 2005; Kapusta et al., 2005a; Tava and Pecetti, 2012). However, on the amounts of other sapogenins in A17 foliar tissue, there is disagreement within the existing literature. Tava and Pecetti (2012) and Huhman et al. (2005) found that soyasapogenol B glycosides accounted for 17 and 25.8% of total saponins, while Kapusta et al. (2005a) reported soyasapogenol B glycosides as present in only trace amounts (1.3%). The three sources also vary widely on the amount of zanhic acid glycosides present in saponin mixtures; Huhman et al. (2005) report .6%, while Tava and Pecetti (2012) report 13.5% and Kapusta et al. (2005a) report 42.1%. These discrepancies could stem from the use of different standards and different numbers of standards to quantify saponins. This study used a single hederagenin standard to quantify a chemically complex mixture of saponins, while Kapusta et al. (2005a) used 18 saponins with one of four different backbone structures as standards. The age and conditions of the plants used in the four studies also differed with some being grown in the field and others being grown in a greenhouse, as in this study. A disproportionate amount of certain sapogenins could also reflect an affinity bias of the column used in the extraction process.

Glycosylation of *M. truncatula* triterpene saponins occurs primarily at the C-3 and C-28 positions and rarely at the C-23 position and can involve apifuranose, arabinose, galactose, glucose, galacturonic acid, pentose, rhamnose, and xylose constituents (Gholami et al., 2014; Huhman et al., 2005). Non-

hemolytic soyasaponins are monodesmosidic, while hemolytic saponins can be mono-, bi- or tri-desmosidic (Tava et al., 2011). Attached sugar chains can be modified by the presence of malonyl groups (Huhman et al., 2005; Pollier et al., 2011a). Individual saponin components were identified and quantified by mass spectrometry (Table 6). Soysaponins glycosylated at the C-3 position and hemolytic saponins glycosylated at the C-3, C-28, and/or C-23 were detected. The majority of sugar chain constituents were reported as hexose because exact stereochemistry of the sugar could not be determined. All chemically characterized sugar constituents detected have been observed attached to sapogenin backbones. Malonyl modification of sugar chains was also detected. A17 extracts were the most diverse, with 39 different saponins, followed by GRC43 with 32, PRT178 with 30, and ESP105 with 21. 52 of the 59 saponins detected have been previously reported in extracts from various *M. truncatula* organs (Huhman et al., 2005; Huhman and Sumner, 2002; Kapusta et al., 2005a; Pollier et al., 2011a). Malonyl-Hex-Hex-Hex-HexA-unknown aglycone 1, Hex-Hex-unknown aglycone 3, Hex-unknown aglycone 3, dHex-hex-HexA-soyasapogenol B, Malonyl-Hex-Hex-HexA-soyasapogenol B, dHex-Hex-Hex-Hex-hederagenin, and 3-Glc-Glc-28-Ara-Rha-Xyl-Api-zanhic acid have not previously been reported in *M. truncatula* saponin extracts to the author's knowledge. Of the seven novel saponins, three were only detected here in extracts of *M. truncatula* accessions that have not yet been chemically investigated in the literature, Hex-unknown aglycone 3 in ESP105 and PRT178, dHex-hex-HexA-soyasapogenol B in ESP105, and Malonyl-Hex-Hex-HexA-soyasapogenol B in GRC43, and could be uniquely produced in those accessions.

Extracts from A17, ESP105, PRT178 and GRC43 foliar tissue at the final concentrations of 100 and 250 µg/ml significantly decreased Caco-2 cell proliferation in a time-dependent manner (Figure 14 and 15). To the author's knowledge, the effects of saponin extracts from *M. truncatula* on cancer cells have not previously been studied. Saponins purified from *M. sativa* inhibited the growth of the K562 human leukemia cell line *in vitro* (Tava and Odoardi, 1996). Soyasaponin I isolated from *Glycine max* decreased Caco-2 cell proliferation and reduced activity of the proliferation enzyme protein kinase C (Salyer et al., 2013). Soyasponin I, referred to in Table 6 as 3-Rha-Gal-GlcA-soyasapogenol B, is also produced by *M. truncatula* and was detected in extracts from all four accessions. At both treatment concentrations, ESP105 extracts had the strongest effect on Caco-2 cell proliferation, followed by A17,

PRT178, and GRC43 extracts, respectively. The differences in extract effectiveness could be due to differences in the concentration of total saponins present in crude extracts. Total saponins were quantified using HPLC-MS and reported as mg hederagenin equivalents per g dried crude extract. An amount of dried crude extract necessary to obtain the desired amount of total saponins, equal across the accessions, was then dissolved in DMSO for cell culture treatments. Because GRC43 total saponins were almost twice as concentrated as ESP105 total saponins, approximately half as much GRC43 dried crude extract was needed compared to ESP105 dried crude extract to obtain an equal amount of total saponins. The order of the accessions organized by cytotoxicity effect from strong to weak (ESP105, A17, PRT178, GRC43) is the same as the order of the accessions organized by concentration of total saponins in crude extracts from lowest to highest. Compounds in the non-saponin portion of the dried crude extracts were not identified. It is possible that the greater proportion of unidentified compounds present in ESP105 and A17 treatments could have contributed to their greater effect on Caco-2 cell proliferation.

However, differences in the composition of accession extracts could also explain the differences in cytotoxic effect. The most abundant aglycone backbone in the more cytotoxic A17 and ESP105 extracts was bayogenin, while the most abundant aglycone backbone in the less cytotoxic PRT178 and GRC43 extracts was soyasapogenol B. The most abundant saponin in the ESP105 extract was Hex-dHex-Hex-HexA-hederagenin, while the most abundant saponin in the other accession extracts was dHex-Hex-Hex-HexA-soyasapogenol B. Two saponins were uniquely detected in the ESP105 extracts, 3-Glc-medicagenic acid and dHex-hex-HexA-soyasapogenol B. While the amount of 3-Glc-medicagenic acid was relatively low in the ESP105 extract, dHex-hex-HexA-soyasapogenol B was the second most abundant saponin. The saponins Hex-dHex-Hex-HexA-hederagenin and dHex-hex-HexA-soyasapogenol B may be important contributors to the comparatively higher cytotoxic effect of the ESP105 extract. Oleanane-type saponins with a free carboxylic group at the C-28 position have been shown to have a stronger cytotoxic effect on four human solid tumor cell lines when compared to oleanane-type saponins with sugars attached at the C-28 position (Bang et al., 2005). The proportion of total saponins with a free carboxylic group at the C-28 position (all glycosides of unknown aglycone 1 and 4, bayogenin, and hederagenin and four glycosides of medicagenic acid) was approximately equal in A17, ESP105, and GRC43 extracts (48.96-50.49%) but slightly lower in PRT178 extracts (38.57%). More generally, certain

monodesmosidic saponins have been shown to have a stronger cytotoxic effect than bidesmosidic saponins. Monodesmosidic saponins showed higher cytotoxic activity than bidesmosides in a study of the cytotoxic effects of 16 triterpene saponins extracted from *Hedera helix, Calendula arvensis, Calendula officinalis,* and *Sapindus mukurossi* on HeLa cells (Quetin-Leclercq et al., 1992). ESP105 extracts have the highest proportion of monodesmosidic saponins with 89.32%, followed by A17, GRC43, and PRT178 extracts with 84.35, 82.61, and 76.16%, respectively. These findings might help to explain why ESP105 saponin extracts have the strongest negative impact on Caco-2 cells in this study. The relatively low amount of glycosides of unknown aglycones 2-5 and soyasapogenol A in all extracts and the low amount of medicagenic acid glycosides in the ESP105 extract suggests that these saponins are not important contributors to the cytotoxic effect of the extracts.

To investigate the mechanism by which *M. truncatula* saponin extracts decrease Caco-2 cell proliferation, caspase-3 activation in response to saponin treatment was studied. The majority of saponins that have been shown to kill cancer cells do so through the induction of apoptosis (Bachran et al., 2008), and caspase-3 is an enzyme activated by both the intrinsic and extrinsic pathway of apoptosis (Ghavami et al., 2009). Apoptosis induction by saponin treatment has previously been shown in Caco-2 cells by assays that assess differential membrane composition in apoptotic cells and caspase 3/7 and 9 activation and by DAPI staining (Alam et al., 2015; Omoyeni et al., 2015). Apoptosis induction in Caco-2 cells treated with the well-established antitumor agent butyrate did not begin until 16 hr after treatment and had occurred in 60% of cells by 48 hr (Ruemmele et al., 2003). After treatment with *M. truncatula* saponin extracts at 100 µg/ml, Caco-2 cell proliferation did not begin to decline until 48 hr post-treatment. Apoptosis induction caused by treatment with saponin extracts from the four accessions at 100 µg/ml was evaluated in three independent experiments by assaying caspase-3 activation at 48 hr post-treatment, and no effect was found (Figure 16). Single experiments assaying caspase-3 activation 24 hr after 100 µg/ml treatments from the four accessions and 48 hr after 250 µg/ml ESP105 treatment also showed no effect of saponin extracts on apoptosis induction (data not shown). These results suggest that *M. truncatula* saponins do not decrease cell proliferation through apoptosis induction but are not conclusive. Saponins have been shown to decrease cell proliferation through a variety of mechanisms that are likely to be related to structure (Bachran et al., 2008; Podolak et al., 2010). In these experiments, Caco-2 cells

were treated with a complex mixture of structurally diverse saponins that could each decrease proliferation through a variety of mechanisms. Saponins that decrease cell proliferation through apoptosis induction could have been present in the extracts but at such low concentrations that the level of apoptosis they might have caused was not sufficient to be detected by the assay used. The experiments performed also lacked a positive control. Treatment of Caco-2 cells with a compound previously shown to induce apoptosis, such as butyrate, would have confirmed that the particular Caco-2 cells used were responding to treatment and that the particular assay used was able to measure Caco-2 apoptosis induction. Additional investigation into the effects of *M. truncatula* saponin extracts on apoptosis induction using assays that involve other aspects of apoptosis besides caspase-3 could also be helpful.

The existence of saponins that do not induce apoptosis and that act as anticancer agents through other mechanisms is not unprecedented. Soybean saponin treatment of HT-29 colon cancer cells resulted in decreased cell proliferation, suppressed protein kinase C activity, and induction of cell differentiation but did not induce apoptosis (Oh and Sung, 2001). Soybean produces some of the same sapogenins and saponins as *M. truncatula*, including soyasapogenol A, B and E and soyasaponin I (Jin et al., 2006). The soybean saponin treatment was later shown to decrease HT-29 colon cancer cell proliferation by suppressing inflammatory responses (Kim et al., 2004). Treatment of HCT-15 colon cancer cells with purified soybean B-group saponin extract decreased cell proliferation, caused cell cycle arrest, and induced autophagic Type II non-apoptotic programmed cell death (Ellington et al., 2005). A mixture of two steroidal saponins isolated from *Balanites aegyptiaca* demonstrated anti-proliferative effects on several cancer cells but did not induce apoptosis, instead causing major actin cytoskeleton disorganization by depleting ATP (Gnoula et al., 2008). Avicin D, a plant triterpenoid that has been shown to trigger apoptosis, was able to induce autophagic programmed cell death in tumor cells treated with an apoptosis inhibitor and in apoptosis-resistant knockout cells (Xu et al., 2007).

Through HPLC-MS analysis, we found that extracts from the foliar tissue of *M. truncatula* A17, ESP105, PRT178, and GRC43 accession plants contained varying levels and types of triterpenoid saponins, and that treatment of Caco-2 human colon cancer cells with these extracts decreased cell proliferation in a dose- and time-dependent manner. This anti-proliferative effect did not appear to be mediated through caspase-3 activation leading to induction of apoptotic programmed cell death. The

remarkable complexity of cancer and the many mechanisms it uses to outwit the immune system and proliferate throughout the body, as well as the negative effects of many currently used chemotherapeutic drugs, has led to a constant demand for novel anticancer compounds. With their extensively studied cytotoxic effects, saponins may be able to satisfy this demand. Numerous studies on the *in vitro* effects of saponins on cancer cells have been conducted, and research on their *in vivo* effects, primarily in mice, is growing. For example, after being shown to significantly inhibit viability of human hepatoma, glioma, a gastric, breast, and colon carcinoma cell cultures, the hederagenin saponin macranthoside B isolated from *Lonicera macranthoides* was used to treat human hepatoma HepG2 xenograft tumors in nude mice by intravenous injection. Treatment significantly reduced tumor weight and volume as compared to the solvent-treated control and with comparable strength to cyclophosphamide, a chemotherapeutic drug (Wang et al., 2009). The steroidal saponin diosgenin isolated from fenugreek was incorporated into the diet of rats that had been injected with the carcinogen azoxymethane one-week prior to induce colon carcinogenesis. The diosgenin diet significantly suppressed invasive and non-invasive colon tumor incidence and decreased tumor multiplicity (Malisetty et al., 2005). In a similar study, a mixture of saponins isolated from soybean flour was used to supplement the diet of azoxymethane-treated mice and untreated mice. The saponin diet was not only able to significantly reduce incidence of precancerous aberrant crypt foci in azoxymethane-treated mice, but also showed no adverse effects on the growth and overall health of the untreated mice (Koratkar and Rao, 1997).

To the author's knowledge, this is the first study to show that saponin extracts from *M. truncatula* plants have cytotoxic activity and lays the foundation for further study of *M. truncatula* saponins and their possible future use as chemotherapeutic drugs. The saponin extracts used in this study, either from all four accessions or from the most anti-proliferative accessions, could be fractionated by preparative HPLC, and the cytotoxicity of those fractions assayed to identify the strongest anti-proliferative individual compounds. The mechanism through which the purified saponin compounds mediate their cytotoxic activity could then be evaluated without the interference of other saponins present in a complex mixture, and information on the structure of the purified saponin compounds could add to our knowledge of the relationship between saponin structure and cytotoxic activity. The effects of the *M. truncatula* saponin extracts used in this study or compounds purified from them on additional cancer cell types grown in

culture and the effect of *M. truncatula* saponin supplemented diets on colon carcinogenesis in azoxymethane-treated mice could be tested to further characerize the possibility of utilization of *M. truncatula* saponins as chemotherapeutic drugs.

Literature Cited

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W., and Dixon, R.A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. The Plant Journal *41*, 875–887.

Aharoni, A., Giri, A.P., Deuerlein, S., Griepink, F., Kogel, W.-J. de, Verstappen, F.W.A., Verhoeven, H.A., Jongsma, M.A., Schwab, W., and Bouwmeester, H.J. (2003). Terpenoid Metabolism in Wild-Type and Transgenic Arabidopsis Plants. Plant Cell *15*, 2866–2884.

Alam, F., Saqib, Q.N., and Waheed, A. (2015). Effect of crude saponins from Gaultheria trichophylla extract on growth inhibition in human colorectal cancer cells. Bangladesh Journal of Pharmacology *10*, 160–165.

Applebaum, S.W., Marco, S., and Birk, Y. (1969). Saponins as possible factors of resistance of legume seeds to the attack of insects. Journal of Agricultural and Food Chemistry *17*, 618–622.

Arase, Y., Ikeda, K., Murashima, N., Chayama, K., Tsubota, A., Koida, I., Suzuki, Y., Saitoh, S., Kobayashi, M., and Kumada, H. (1997). The long term efficacy of glycyrrhizin in chronic hepatitis C patients. Cancer *79*, 1494–1500.

Armah, C.N., Mackie, A.R., Roy, C., Price, K., Osbourn, A.E., Bowyer, P., and Ladha, S. (1999). The Membrane-Permeabilizing Effect of Avenacin A-1 Involves the Reorganization of Bilayer Cholesterol. Biophysical Journal *76*, 281–290.

Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry *72*, 435–457.

Artursson, P., and Karlsson, J. (1991). Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. Biochemical and Biophysical Research Communications *175*, 880–885.

Bachran, C., Bachran, S., Sutherland, M., Bachran, D., and Fuchs, H. (2008). Saponins in tumor therapy. Mini Reviews in Medicinal Chemistry *8*, 575–584.

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., and Werck-Reichhart, D. (2011). Cytochromes P450. The Arabidopsis Book e0144.

Bang, S.-C., Lee, J.-H., Song, G.-Y., Kim, D.-H., Yoon, M.-Y., and Ahn, B.-Z. (2005). Antitumor Activity of Pulsatilla koreana Saponins and Their Structure–Activity Relationship. Chemical and Pharmaceutical Bulletin *53*, 1451–1454.

Barker, D.G., Pfaff, T., Moreau, D., Groves, E., Ruffel, S., Lepetit, M., Whitehand, S., Maillet, F., Nair, R.M., and Journet, E.-P. (2007). Growing M. truncatula: choice of substrates and growth conditions. In The Medicago Truncatula Handbook.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., et al. (2008). A gene expression atlas of the model legume Medicago truncatula. The Plant Journal *55*, 504–513.

Bialy, Z., Jurzysta, M., Mella, M., and Tava, A. (2006). Triterpene saponins from the roots of Medicago hybrida. Journal of Agricultural and Food Chemistry *54*, 2520–2526.

Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., Scotti, C., and Calderini, O. (2015). CYP72A67 Catalyzes a Key Oxidative Step in Medicago truncatula Hemolytic Saponin Biosynthesis. Molecular Plant *8*, 1493–1506.

Bora, K.S., and Sharma, A. (2011). Phytochemical and pharmacological potential of *Medicago sativa* : A review. Pharmaceutical Biology *49*, 211–220.

Broeckling, C.D., Huhman, D.V., Farag, M.A., Smith, J.T., May, G.D., Mendes, P., Dixon, R.A., and Sumner, L.W. (2005). Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. Journal of Experimental Botany *56*, 323–336.

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi, M., Piano, E., Arcioni, S., et al. (2011). Medicago truncatula CYP716A12 Is a Multifunctional Oxidase Involved in the Biosynthesis of Hemolytic Saponins. Plant Cell *23*, 3070–3081.

Carelli, M., Biazzi, E., Tava, A., Losini, I., Abbruscato, P., Depedro, C., and Scotti, C. (2015). Sapogenin content variation in Medicago inter-specific hybrid derivatives highlights some aspects of saponin synthesis and control. New Phytologist *206*, 303–314.

Cheeke, P.R. (1996). Biological Effects of Feed and Forage Saponins and Their Impacts on Animal Production. In Saponins Used in Food and Agriculture, G.R. Waller, and K. Yamasaki, eds. (Springer US), pp. 377–385.

Cheeke, P.R. (2000). Actual and potential applications of and saponins in human and animal nutrition. Journal of Animal Science *77*, 1–10.

Cheeke, P.R. (2009). Applications of saponins as feed additives in poultry production. In 20th Annual Australian Poultry Science Symposium, p. 50.

Choi, H.-K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J.-H., Kalo, P., Penmetsa, R.V., Seres, A., Kulikova, O., et al. (2004a). A Sequence-Based Genetic Map of Medicago truncatula and Comparison of Marker Colinearity with M. sativa. Genetics *166*, 1463–1502.

Choi, H.-K., Mun, J.-H., Kim, D.-J., Zhu, H., Baek, J.-M., Mudge, J., Roe, B., Ellis, N., Doyle, J., Kiss, G.B., et al. (2004b). Estimating Genome Conservation between Crop and Model Legume Species. Proceedings of the National Academy of Sciences of the United States of America *101*, 15289–15294.

Cinatl, J., Morgenstern, B., Bauer, G., Chandra, P., Rabenau, H., and Doerr, H. (2003). Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus. The Lancet *361*, 2045–2046.

Coleman, J.J., Okoli, I., Tegos, G.P., Holson, E.B., Wagner, F.F., Hamblin, M.R., and Mylonakis, E. (2010). Characterization of Plant-Derived Saponin Natural Products against Candida albicans. ACS Chemical Biology *5*, 321–332.

Croteau, R., Ketchum, R.E.B., Long, R.M., Kaspera, R., and Wildung, M.R. (2006). Taxol biosynthesis and molecular genetics. Phytochemistry Reviews *5*, 75–97.

D'Addabbo, T., Carbonara, T., Leonetti, P., Radicci, V., Tava, A., and Avato, P. (2011). Control of plant parasitic nematodes with active saponins and biomass from Medicago sativa. Phytochemistry Reviews *10*, 503–519.

De Geyter, N., Gholami, A., Goormachtig, S., and Goossens, A. (2012). Transcriptional machineries in jasmonate-elicited plant secondary metabolism. Trends in Plant Science *17*, 349–359.

Delalande, M., Greene, S., Hughes, S., Nair, R., Huguet, T., Aouani, M.E., and Prosperi, J.-M. (2007). Wild accessions / populations. In The Medicago Truncatula Handbook.

Dixon, R.A., and Sumner, L.W. (2003). Legume Natural Products: Understanding and Manipulating Complex Pathways for Human and Animal Health. Plant Physiology. *131*, 878–885.

Drake, J.J.-P., Butterworth, K.R., Gaunt, I.F., Hooson, J., Evans, J.G., and Gangolli, S.D. (1982). Long-term toxicity study of quillaia extract in rats. Food and Chemical Toxicology *20*, 15–23.

Elias, P.M., Goerke, J., Friend, D.S., and Brown, B.E. (1978). Freeze-fracture identification of sterol-digitonin complexes in cell and liposome membranes. The Journal of Cell Biology *78*, 577–596.

Ellington, A.A., Berhow, M., and Singletary, K.W. (2005). Induction of macroautophagy in human colon cancer cells by soybean B-group triterpenoid saponins. Carcinogenesis *26*, 159–167.

Felsenstein, J. (1989). PHYLIP--Phylogeny Inference Package (Version 3.2). Cladistics *5*, 164–166.

Francis, G., Kerem, Z., Makkar, H.P.S., and Becker, K. (2002). The biological action of saponins in animal systems: a review. British Journal of Nutrition *88*, 587–605.

Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., Saito, K., and Muranaka, T. (2011). CYP716A Subfamily Members are Multifunctional Oxidases in Triterpenoid Biosynthesis. Plant and Cell Physiology *52*, 2050–2061.

Fukushima, E.O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., and Muranaka, T. (2013). Combinatorial Biosynthesis of Legume Natural and Rare Triterpenoids in Engineered Yeast. Plant and Cell Physiology *54*, 740–749.

Gauthier, C., Legault, J., Girard-Lalancette, K., Mshvildadze, V., and Pichette, A. (2009). Haemolytic activity, cytotoxicity and membrane cell permeabilization of semi-synthetic and natural lupane- and oleanane-type saponins. Bioorganic & Medicinal Chemistry *17*, 2002–2008.

Ghavami, S., Hashemi, M., Ande, S.R., Yeganeh, B., Xiao, W., Eshraghi, M., Bus, C.J., Kadkhoda, K., Wiechec, E., Halayko, A.J., et al. (2009). Apoptosis and cancer: mutations within caspase genes. Journal of Medical Genetics *46*, 497–510.

Gholami, A., Geyter, N.D., Pollier, J., Goormachtig, S., and Goossens, A. (2014). Natural product biosynthesis in Medicago species. Natural Product Reports *31*, 356–380.

Gnoula, C., Mégalizzi, V., Neve, N.D., Sauvage, S., Ribaucour, F., Guissou, P., Duez, P., Dubois, J., Ingrassia, L., Lefranc, F., et al. (2008). Balanitin-6 and-7: diosgenyl saponins isolated from Balanites aegyptiaca Del. display significant anti-tumor activity in vitro and in vivo. International Journal of Oncology *32*, 5–16.

Gögelein, H., and Hüby, A. (1984). Interaction of saponin and digitonin with black lipid membranes and lipid monolayers. Biochimica et Biophysica Acta (BBA) - Biomembranes *773*, 32–38.

Graham, P.H., and Vance, C.P. (2003). Legumes: Importance and Constraints to Greater Use. Plant Physiology *131*, 872–877.

Güçlü-Üstündağ, Ö., and Mazza, G. (2007). Saponins: Properties, Applications and Processing. Critical Reviews in Food Science and Nutrition *47*, 231–258.

Gurfinkel, D., and Rao, A. (2003). Soyasaponins: the relationship between chemical structure and colon anticarcinogenic activity. Nutrition & Cancer *47*, 24–33.

Haridas, V., Higuchi, M., Jayatilake, G.S., Bailey, D., Mujoo, K., Blake, M.E., Arntzen, C.J., and Gutterman, J.U. (2001). Avicins: Triterpenoid saponins from Acacia victoriae (Bentham) induce apoptosis by mitochondrial perturbation. PNAS *98*, 5821–5826.

Higgins, D.G., Bleasby, A.J., and Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. Computer Applications in the Biosciences *8*, 189–191.

Holtshausen, L., Chaves, A.V., Beauchemin, K.A., McGinn, S.M., McAllister, T.A., Odongo, N.E., Cheeke, P.R., and Benchaar, C. (2009). Feeding saponin-containing Yucca schidigera and Quillaja saponaria to decrease enteric methane production in dairy cows1. Journal of Dairy Science *92*, 2809–2821.

Howe, G.A., and Jander, G. (2008). Plant Immunity to Insect Herbivores. Annual Review of Plant Biology *59*, 41–66.

Huhman, D.V., and Sumner, L.W. (2002). Metabolic profiling of saponins in Medicago sativa and Medicago truncatula using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry *59*, 347–360.

Huhman, D.V., Berhow, M.A., and Sumner, L.W. (2005). Quantification of Saponins in Aerial and Subterranean Tissues of Medicago truncatula. Journal of Agricultural and Food Chemistry *53*, 1914–1920.

Hummelbrunner, L.A., and Isman, M.B. (2001). Acute, Sublethal, Antifeedant, and Synergistic Effects of Monoterpenoid Essential Oil Compounds on the Tobacco Cutworm, Spodoptera litura (Lep., Noctuidae). Journal of Agricultural and Food Chemistry *49*, 715–720.

Hwang, D.-C., and Damodaran, S. (1994). Selective Precipitation of Fat Globule Membranes of Cheese Whey by Saponin and Bile Salt. Journal of Agricultural and Food Chemistry *42*, 1872–1878.

Irmler, S., Schröder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000). Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. The Plant Journal *24*, 797–804.

Iturbe-Ormaetxe, I., Haralampidis, K., Papadopoulou, K., and Osbourn, A.E. (2003). Molecular cloning and characterization of triterpene synthases from Medicago truncatula and Lotus japonicus. Plant Molecular Biology *51*, 731–743.

Jin, M., Yang, Y., Su, B., and Ren, Q. (2006). Rapid quantification and characterization of soyasaponins by high-performance liquid chromatography coupled with electrospray mass spectrometry. Journal of Chromatography A *1108*, 31–37.

Johnson, I.T., Gee, J.M., Price, K., Curl, C., and Fenwick, G.R. (1986). Influence of saponins on gut permeability and active nutrient transport in vitro. J Nutr *116*, 2270–2277.

Kamphuis, L.G., Lichtenzveig, J., Peng, K., Guo, S.-M., Klingler, J.P., Siddique, K.H.M., Gao, L.-L., and Singh, K.B. (2013). Characterization and genetic dissection of resistance to spotted alfalfa aphid (Therioaphis trifolii) in Medicago truncatula. Journal of Experimental Botany *64*, 5157–5172.

Kang, R.K.L., Zyzak, L.L., and Nakatsu, T. (1999). Flavored product additive and method for using same.

Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of Saponins in Aerial Parts of Barrel Medic (Medicago truncatula) by Liquid Chromatography−Electrospray Ionization/Mass Spectrometry. Journal of Agricultural and Food Chemistry *53*, 7654–7660.

Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene Saponins from Barrel Medic (Medicago truncatula) Aerial Parts. Journal of Agricultural and Food Chemistry *53*, 2164–2170.

Kennelly, E.J., Suttisri, R., and Kinghorn, A.D. (1996). Novel Sweet-Tasting Saponins of the Cycloartane, Oleanane, Secodammarane, and Steroidal Types. In Saponins Used in Food and Agriculture, G.R. Waller, and K. Yamasaki, eds. (Springer US), pp. 13–24.

Kessler, A., and Baldwin, and I.T. (2002). PLANT RESPONSES TO INSECT HERBIVORY: The Emerging Molecular Analysis. Annual Review of Plant Biology *53*, 299–328.

Kevei, Z., Lougnon, G., Mergaert, P., Horváth, G.V., Kereszt, A., Jayaraman, D., Zaman, N., Marcel, F., Regulski, K., Kiss, G.B., et al. (2007). 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase1 Interacts with NORK and Is Crucial for Nodulation in Medicago truncatula. Plant Cell *19*, 3974–3989.

Kikuchi, Y., Sasa, H., Kita, T., Hirata, J., Tode, T., and Nagata, I. (1991). Inhibition of human ovarian cancer cell proliferation in vitro by ginsenoside Rh2 and adjuvant effects to cisplatin in vivo. Anti-Cancer Drugs *2*, 63–68.

Kim, H.-Y., Yu, R., Kim, J.-S., Kim, Y.-K., and Sung, M.-K. (2004). Antiproliferative crude soy saponin extract modulates the expression of IκBα, protein kinase C, and cyclooxygenase-2 in human colon cancer cells. Cancer Letters *210*, 1–6.

Kliebenstein, D.J. (2012). Plant Defense Compounds: Systems Approaches to Metabolic Analysis. Annual Review of Phytopathology *50*, 155–173.

Koratkar, R., and Rao, A.V. (1997). Effect of soya bean saponins on azoxymethane‐induced preneoplastic lesions in the colon of mice. Nutrition and Cancer *27*, 206–209.

Kucukkurt, I., Akkol, E.K., Karabag, F., Ince, S., Süntar, I., Eryavuz, A., and Sözbilir, N.B. (2016). Determination of the regulatory properties of Yucca schidigera extracts on the biochemical parameters and plasma hormone levels associated with obesity. Revista Brasileira de Farmacognosia *26*, 246–250.

Kuzina, V., Ekstrøm, C.T., Andersen, S.B., Nielsen, J.K., Olsen, C.E., and Bak, S. (2009). Identification of Defense Compounds in Barbarea vulgaris against the Herbivore Phyllotreta nemorum by an Ecometabolomic Approach. Plant Physiology *151*, 1977–1990.

Lacaille-Dubois, M.A. (2007). Bioactive Saponins from plants: Recent Development (Handbook of Medicinal plants. India: Harworth Press).

Li, M., Wei, S.-Y., Xu, B., Guo, W., Liu, D.-L., Cui, J.-R., and Yao, X.-S. (2008). Pro-apoptotic and microtubule-disassembly effects of ardisiacrispin (A+B), triterpenoid saponins from Ardisia crenata on human hepatoma Bel-7402 cells. Journal of Asian Natural Products Research *10*, 729–736.

Li, Y.-H., Niu, Y.-B., Sun, Y., Zhang, F., Liu, C.-X., Fan, L., and Mei, Q.-B. (2015). Role of phytochemicals in colorectal cancer prevention. World Journal of Gastroenterology *21*, 9262–9272.

Lin, F., and Wang, R. (2009). Hemolytic mechanism of dioscin proposed by molecular dynamics simulations. Journal of Molecular Modeling *16*, 107–118.

Lindl, T., and Steubing, R. (2013). Atlas of living cell cultures (Weinheim, Germany: Wiley-Blackwell).

Liu, J. (1995). Pharmacology of oleanolic acid and ursolic acid. Journal of Ethnopharmacology *49*, 57–68.

Liu, W.K., Xu, S.X., and Che, C.T. (2000). Anti-proliferative effect of ginseng saponins on human prostate cancer cell line. Life Sciences *67*, 1297–1306.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods *25*, 402–408.

Lorent, J., Lins, L., Domenech, Ò, Quetin-Leclercq, J., Brasseur, R., and Mingeot-Leclercq, M.-P. (2014). Domain formation and permeabilization induced by the saponin α-hederin and its aglycone hederagenin in a cholesterol-containing bilayer. Langmuir *30*, 4556–4569.

Mader, T.L., and Brumm, M.C. (1987). Effect of feeding sarsaponin in cattle and swine diets. Journal of Animal Science *65*, 9–15.

Maffei, M.E., Gertsch, J., and Appendino, G. (2011). Plant volatiles: Production, function and pharmacology. Natural Product Reports *28*, 1359.

Malisetty, V.S., Patlolla, J.M.R., Raju, J., Marcus, L.A., Choi, C.-I., and Rao, C.V. (2005). Chemoprevention of colon cancer by diosgenin, a steroidal saponin constituent of fenugreek. Cancer Research *65*, 580–580.

Mallinckrodt, B., Thakur, M., Weng, A., Gilabert-Oriol, R., Dürkop, H., Brenner, W., Lukas, M., Beindorff, N., Melzig, M.F., and Fuchs, H. (2014). Dianthin-EGF is an effective tumor targeted toxin in combination with saponins in a xenograft model for colon carcinoma. Future Oncology *10*, 2161–2175.

Man, S., Gao, W., Zhang, Y., Yan, L., Ma, C., Liu, C., and Huang, L. (2009). Antitumor and antimetastatic activities of Rhizoma Paridis saponins. Steroids *74*, 1051–1056.

Matsuda, H., Li, Y., Murakami, T., Yamahara, J., and Yoshikawa, M. (1999). Structure-Related Inhibitory Activity of Oleanolic Acid Glycosides on Gastric Emptying in Mice. Bioorganic & Medicinal Chemistry *7*, 323–327.

Mertens, J., Pollier, J., Bossche, R.V., Lopez-Vidriero, I., Franco-Zorrilla, J.M., and Goossens, A. (2016). The bHLH Transcription Factors TSAR1 and TSAR2 Regulate Triterpene Saponin Biosynthesis in Medicago truncatula. Plant Physiology *170*, 194–210.

Milgate, J., and Roberts, D.C.K. (1995). The nutritional & biological significance of saponins. Nutrition Research *15*, 1223–1249.

Mimaki, Y., Kuroda, M., Kameyama, A., Yokosuka, A., and Sashida, Y. (1998). Steroidal saponins from the rhizomes of Hosta sieboldii and their cytostatic activity on HL-60 cells. Phytochemistry *48*, 1361–1369.

Mithöfer, A., and Boland, W. (2012). Plant Defense Against Herbivores: Chemical Aspects. Annual Review of Plant Biology *63*, 431–450.

Misawa, N. (2011). Pathway engineering for functional isoprenoids. Current Opinion in Biotechnology *22*, 627–633.

Moses, T., Papadopoulou, K.K., and Osbourn, A. (2014a). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. Critical Reviews in Biochemistry and Molecular Biology *49*, 439–462.

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M.A., Martins, J.C., Thevelein, J.M., and Goossens, A. (2014b). Combinatorial biosynthesis of sapogenins and saponins in Saccharomyces cerevisiae using a C-16α hydroxylase from Bupleurum falcatum. Proceedings of the National Academy of Sciences *111*, 1634–1639.

Moses, T., Thevelein, J.M., Goossens, A., and Pollier, J. (2014c). Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. Phytochemistry *108*, 47–56.

Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L., Thevelein, J.M., Geelen, D., and Goossens, A. (2015). Unraveling the Triterpenoid Saponin Biosynthesis of the African Shrub Maesa lanceolata. Molecular Plant *8*, 122–135.

Naoumkina, M.A., Modolo, L.V., Huhman, D.V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L.W., and Dixon, R.A. (2010). Genomic and Coexpression Analyses Predict Multiple Genes Involved in Triterpene Saponin Biosynthesis in Medicago truncatula. Plant Cell *22*, 850–866.

Nasrollahi, V., Mirzaie-asl, A., Piri, K., Nazeri, S., and Mehrabi, R. (2014). The effect of drought stress on the expression of key genes involved in the biosynthesis of triterpenoid saponins in liquorice (Glycyrrhiza glabra). Phytochemistry *103*, 32–37.

Nishikawa, M., Nojima, S., Akiyama, T., Sankawa, U., and Inoue, K. (1984). Interaction of Digitonin and Its Analogs with Membrane Cholesterol. Journal of Biochemistry *96*, 1231–1239.

Oh, Y.J., and Sung, M.K. (2001). Soybean saponins inhibit cell proliferation by suppressing PKC activation and induce differentiation of HT-29 human colon adenocarcinoma cells. Nutrition and Cancer *39*, 132–138.

Omoyeni, O.A., Hussein, A., Meyer, M., Green, I., and Iwuoha, E. (2015). Pleiocarpa pycnantha leaves and its triterpenes induce apoptotic cell death in Caco-2 cells in vitro. BMC Complementary and Alternative Medicine *15*.

Ozawa, R., Shimoda, T., Kawaguchi, M., Arimura, G., Horiuchi, J., Nishioka, T., and Takabayashi, J. (2000). Lotus japonicus Infested with Herbivorous Mites Emits Volatile Compounds That Attract Predatory Mites. Journal of Plant Research *113*, 427–433.

Park, J.C., Lee, Y.J., Choi, H.Y., Shin, Y.K., Kim, J.D., and Ku, S.K. (2014). In Vivo and In Vitro Antitumor Effects of Platycodin D, a Saponin Purified from Platycodi Radix on the H520 Lung Cancer Cell. Evidence-Based Complementary & Alternative Medicine *2014*, 1–17 17p.

Pen, B., Sar, C., Mwenya, B., Kuwaki, K., Morikawa, R., and Takahashi, J. (2006). Effects of Yucca schidigera and Quillaja saponaria extracts on in vitro ruminal fermentation and methane emission. Animal Feed Science and Technology *129*, 175–186.

Petit, P., Sauvaire, Y., Ponsin, G., Manteghetti, M., Fave, A., and Ribes, G. (1993). Effects of a fenugreek seed extract on feeding behaviour in the rat: Metabolic-endocrine correlates. Pharmacology Biochemistry and Behavior *45*, 369–374.

Phillips, M.A., and Croteau, R.B. (1999). Resin-based defenses in conifers. Trends in Plant Science *4*, 184–190.

Pichersky, E., and Lewinsohn, E. (2011). Convergent Evolution in Plant Specialized Metabolism. Annual Review of Plant Biology *62*, 549–566.

Podolak, I., Galanty, A., and Sobolewska, D. (2010). Saponins as cytotoxic agents: a review. Phytochemistry Reviews *9*, 425–474.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011a). Metabolite profiling of triterpene saponins in Medicago truncatula hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. Journal of Natural Products *74*, 1462–1476.

Pollier, J., Moses, T., and Goossens, A. (2011b). Combinatorial biosynthesis in plants: A (p)review on its potential and future exploitation. Natural Product Reports *28*, 1897.

Pollier, J., Moses, T., González-Guzmán, M., De Geyter, N., Lippens, S., Bossche, R.V., Marhavý, P., Kremer, A., Morreel, K., Guérin, C.J., et al. (2013). The protein quality control system manages plant defence compound synthesis. Nature *504*, 148–152.

Price, K.R., Griffiths, N.M., Curl, C.L., and Fenwick, G.R. (1985). Undesirable sensory properties of the dried pea (Pisum sativum). The rôle of saponins. Food Chemistry *17*, 105–115.

Richardson, T., and Jimenez-Flores, R. (1994). Process to remove cholesterol from dairy products.

Quetin-Leclercq, J., Elias, R., Balansard, G., Bassleer, R., and Angenot, L. (1992). Cytotoxic Activity of Some Triterpenoid Saponins. Planta Medica *58*, 279–281.

Rogers, A.M. (2015). A comparative study of Medicago truncatula transcriptomes as they relate to saponin defenses against insect pests. University of Arkansas.

Rosenqvist, E., Michaelsen, T.E., and Vistnes, A.I. (1980). Effect of streptolysin O and digitonin on egg lecithin/cholesterol vesicles. Biochimica et Biophysica Acta (BBA) - Biomembranes *600*, 91–102.

Roy, D., Kommalapati, R.R., Mandava, S.S., Valsaraj, K.T., and Constant, W.D. (1997). Soil Washing Potential of a Natural Surfactant. Environmental Science and Technology *31*, 670–675.

Ruemmele, F.M., Schwartz, S., Seidman, E.G., Dionne, S., Levy, E., and Lentze, M.J. (2003). Butyrate induced Caco-2 cell apoptosis is mediated via the mitochondrial pathway. Gut *52*, 94–100.

Salyer, J., Eswaranandam, S., and Lee, S.-O. (2013). Soyasaponin I, III, and Soyasapogenol B Inhibit Proliferation and Modulate PKC Expression in Caco-2 Human Colon Cancer Cells. Journal of Food Research *2*, p81.

Sambuy, Y., De Angelis, I., Ranaldi, G., Scarino, M.L., Stammati, A., and Zucco, F. (2005). The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. Cell Biology and Toxicology *21*, 1–26.

San Martín, R., and Briones, R. (1999). Industrial uses and sustainable supply of Quillaja saponaria (Rosaceae) saponins. Economic Botany *53*, 302–311.

Saniewska, A., Jurzysta, M., and Biały, Z. (2013). Differential antifungal activity of alfalfa (Medicago santva L.) saponins originated from roots and aerial parts for some ornamental plant pathogens. Acta Agrobotanica *54*, 31–43.

Scott, G.I., Colligan, P.B., Ren, B.H., and Ren, J. (2001). Ginsenosides Rb1 and Re decrease cardiac contraction in adult rat ventricular myocytes: role of nitric oxide. British Journal of Pharmacology *134*, 1159–1165.

Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., Fukushima, E.O., Akashi, T., Aoki, T., Saito, K., et al. (2011). Triterpene Functional Genomics in Licorice for Identification of CYP72A154 Involved in the Biosynthesis of Glycyrrhizin. Plant Cell *23*, 4112–4123.

Sidhu, G.S., and Oakenfull, D.G. (1986). A mechanism for the hypocholesterolaemic activity of saponins. British Journal of Nutrition *55*, 643–649.

Singh, B., and Sharma, R.A. (2014). Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. 3 Biotech *5*, 129–151.

Srinivasan, S., Koduru, S., Kumar, R., Venguswamy, G., Kyprianou, N., and Damodaran, C. (2009). Diosgenin targets Akt-mediated prosurvival signaling in human breast cancer cells. International Journal of Cancer *125*, 961–967.

Steel, C.C., and Drysdale, R.B. (1988). Electrolyte leakage from plant and fungal tissues and disruption of liposome membranes by α-tomatine. Phytochemistry *27*, 1025–1030.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques *28*, 1102–1104.

Sun, H.-X., Xie, Y., and Ye, Y.-P. (2009). Advances in saponin-based adjuvants. Vaccine *27*, 1787–1796.

Sun, S., Li, Y., Fang, W., Cheng, P., Liu, L., and Li, F. (2010). Effect and mechanism of AR-6 in experimental rheumatoid arthritis. Clinical and Experimental Medicine *10*, 113–121.

Sung, J., Han, K.-H., Zo, J.-H., Park, H.-J., Kim, C.-H., and Oh, B.-H. (2000). Effects of Red Ginseng upon Vascular Endothelial Function in Patients with Essential Hypertension. The American Journal of Chinese Medicine *28*, 205–216.

Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T., and Dixon, R.A. (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in Medicago truncatula. The Plant Journal *32*, 1033–1048.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P., and Dixon, R.A. (2004). Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume Medicago truncatula. Planta *220*, 696–707.

Tadege, M., Ratet, P., and Mysore, K.S. (2005). Insertional mutagenesis: a Swiss Army knife for functional genomics of Medicago truncatula. Trends in Plant Science *10*, 229–235.

Tava, A., and Odoardi, M. (1996). Saponins from Medicago spp.: chemical characterization and biological activity against insects. In Saponins Used in Food and Agriculture, (Springer), pp. 97–109.

Tava, A., and Pecetti, L. (2012). Chemical investigation of saponins from twelve annual Medicago species and their bioassay with the brine shrimp Artemia salina. Natural Product Communications *7*, 837–840.

Tava, A., Mella, M., Avato, P., Argentieri, M.P., Bialy, Z., and Jurzysta, M. (2005). Triterpenoid glycosides from leaves of Medicago arborea L. Journal of Agricultural and Food Chemistry *53*, 9954–9965.

Tava, A., Mella, M., Avato, P., Biazzi, E., Pecetti, L., Bialy, Z., and Jurzysta, M. (2009). New Triterpenic Saponins from the Aerial Parts of Medicago arabica (L.) Huds. Journal of Agricultural and Food Chemistry *57*, 2826–2835.

Tava, A., Scotti, C., and Avato, P. (2011). Biosynthesis of saponins in the genus Medicago. Phytochemistry Reviews *10*, 459–469.

Thakur, M., Mergel, K., Weng, A., von Mallinckrodt, B., Gilabert-Oriol, R., Dürkop, H., Melzig, M.F., and Fuchs, H. (2013). Targeted tumor therapy by epidermal growth factor appended toxin and purified saponin: An evaluation of toxicity and therapeutic potential in syngeneic tumor bearing mice. Molecular Oncology *7*, 475–483.

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourn, A. (2014). Triterpene Biosynthesis in Plants. Annual Review of Plant Biology *65*, 225–257.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. Nucleic Acids Research *22*, 4673–4680.

Thoquet, P., Ghérardi, M., Journet, E.-P., Kereszt, A., Ané, J.-M., Prosperi, J.-M., and Huguet, T. (2002). The molecular genetic linkage map of the model legume Medicago truncatula: an essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biology *2*, 1.

Tong, Q.-Y., Qing, Y., Shu, D., He, Y., Zhao, Y.-L., Li, Y., Wang, Z.-L., Zhang, S.-Y., Xing, Z., Xu, C., et al. (2011). Deltonin, a Steroidal Saponin, Inhibits Colon Cancer Cell Growth in Vitro and Tumor Growth in Vivo via Induction of Apoptosis and Antiangiogenesis. Cellular Physiology and Biochemistry *27*, 233–242.

Veerappan, V., Kadel, K., Alexis, N., Scott, A., Kryvoruchko, I., Sinharoy, S., Taylor, M., Udvardi, M., and Dickstein, R. (2014). Keel petal incision: a simple and efficient method for genetic crossing in Medicago truncatula. Plant Methods *10*, 11.

Walker, B.W., Manhanke, N., and Stine, K.J. (2008). Comparison of the interaction of tomatine with mixed monolayers containing phospholipid, egg sphingomyelin, and sterols. Biochimica et Biophysica Acta (BBA) - Biomembranes *1778*, 2244–2257.

Wang, J., Zhao, X.-Z., Qi, Q., Tao, L., Zhao, Q., Mu, R., Gu, H.-Y., Wang, M., Feng, X., and Guo, Q.-L. (2009). Macranthoside B, a hederagenin saponin extracted from Lonicera macranthoides and its anti-tumor activities in vitro and in vivo. Food and Chemical Toxicology *47*, 1716–1721.

Wang, Y., Zhang, Y., Zhu, Z., Zhu, S., Li, Y., Li, M., and Yu, B. (2007). Exploration of the correlation between the structure, hemolytic activity, and cytotoxicity of steroid saponins. Bioorganic & Medicinal Chemistry *15*, 2528–2532.

Watson, B.S., Bedair, M.F., Urbanczyk-Wochniak, E., Huhman, D.V., Yang, D.S., Allen, S.N., Li, W., Tang, Y., and Sumner, L.W. (2015). Integrated Metabolomics and Transcriptomics Reveal Enhanced Specialized Metabolism in Medicago truncatula Root Border Cells. Plant Physiology *167*, 1699–1716.

Weng, A., Thakur, Melzig, and Fuchs (2011). Chemistry and pharmacology of saponins: special focus on cytotoxic properties. Botanics: Targets and Therapy 19.

Windisch, W., Schedle, K., Plitzner, C., and Kroismayr, A. (2008). Use of phytogenic products as feed additives for swine and poultry. Journal of Animal Science *86*, E140–E148.

Xu, Z.-X., Liang, J., Haridas, V., Gaikwad, A., Connolly, F.P., Mills, G.B., and Gutterman, J.U. (2007). A plant triterpenoid, avicin D, induces autophagy by activation of AMP-activated protein kinase. Cell Death & Differentiation *14*, 1948–1957.

Yoshiki, Y., Kudou, S., and Okubo, K. (1998). Relationship between Chemical Structures and Biological Activities of Triterpenoid Saponins from Soybean. Bioscience, Biotechnology, and Biochemistry *62*, 2291–2299.

Zhang, J.-D., Cao, Y.-B., Xu, Z., Sun, H.-H., An, M.-M., Yan, L., Chen, H.-S., Gao, P.-H., Wang, Y., Jia, X.-M., et al. (2005). In Vitro and in Vivo Antifungal Activities of the Eight Steroid Saponins from Tribulus terrestris L. with Potent Activity against Fluconazole-Resistant Fungal. Biological and Pharmaceutical Bulletin *28*, 2211–2215.

Zhang, Y.-F., Fan, X.-J., Li, X., Peng, L.-L., Wang, G.-H., Ke, K.-F., and Jiang, Z.-L. (2008). Ginsenoside Rg1 protects neurons from hypoxic–ischemic injury possibly by inhibiting Ca2+ influx through NMDA receptors and L-type voltage-dependent Ca2+ channels. European Journal of Pharmacology *586*, 90–99.