Legume Natural Products: Understanding and Manipulating Complex Pathways for Human and Animal Health[1]

Richard A. Dixon* and Lloyd W. Sumner

Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73401

"To every thing there is a season" (Ecclesiastes, 3), and interest in plant natural products is certainly undergoing a renaissance at the present time. This is particularly true in the case of the natural products of the Leguminosae, which have long been studied as important taxonomic markers for this complex and economically important family (Wink and Waterman, 1999). In the face of the vast number of natural products collectively produced by plants, the study of specific pathways had been viewed as somewhat esoteric, and attempts to obtain a more global understanding of natural product biosynthesis seemed beyond easy grasp. Those views have been changing in recent years due to the realization of the importance of natural products for plant, animal, and human health, and the impact of genomics technologies on all areas of biology. At least 25% of the genome of Arabidopsis encodes enzymes of metabolism, and the number may be similar or even higher in legumes, several of which now have extensive genomics resources (Quackenbush et al., 2000; Bell et al., 2001). Whole genome-level DNA sequence information, coupled with improved methods for profiling natural products, now make possible combined genetic and biochemical approaches for addressing natural product function, deciphering biosynthetic pathways, and engineering novel pathways in transgenic plants. Several of the following case studies highlight these approaches.

ALKALOIDS AND NONPROTEIN AMINO ACIDS (NPAAS)

Within the approximately 650 genera and more than 18,000 species of legumes, quinolizidine (characteristic of *Lupinus* species; Fig. 1), dipiperidine, pyrrolizidine, β-carboline, phenylethylamine, and indole alkaloids have been reported. The Tyr-derived *Erythrina* alkaloids appear to be found only in the large genus *Erythrina*. NPAAs are also common within the Leguminosae, with canavanine, pipecolic acid, and djencolic acid derivatives the most important groups. NPAAs are often highly toxic, and are responsible for several serious human toxicoses, among the best known of which is lathyrism, a nonprogressive motor neuron disease associated with high consumption of grasspeas. As early as the 5th century BC, writers described the irreversible weakness in the legs of the inhabitants of ancient cities during times of war and starvation, when they were forced to eat a diet containing a high proportion of pulses (Retief and Cilliers, 2002). Grasspeas, which are ideally suited to arid regions such as Ethiopia and the Indian subcontinent, contain high levels of ODPA (Fig. 1) in their seeds, and this compound is responsible for the neurological symptoms and also for deleterious effects on bone formation, particularly in children. Although low-ODPA lines of grasspea have been developed through traditional breeding and selection that appear suitable as supplementary material for animal feeds (Hanbury et al., 2000), removal of the neurotoxin from the seed by transgenic approaches is yet to be reported. More work is needed on the molecular biology of the biosynthetic pathways leading to the many nitrogen-containing natural products of the Leguminosae.

ISOFLAVONOIDS: NATURAL PRODUCTS FOR PLANT AND HUMAN HEALTH

Flavonoids are found throughout the plant kingdom, whereas isoflavonoids are more restricted. Isoflavonoids are particularly prevalent in the Papilonoideae subfamily of the Leguminosae, in which they are widely distributed and function as preformed or inducible antimicrobial or anti-insect compounds, as inducers of the nodulation genes of symbiotic *Rhizobium* bacteria, or as allelopathic agents (Dixon, 1999). Pterocarpan-type phytoalexins such as medicarpin and constitutive isoflavone malonyl glycosides (Fig. 1) are typical of the isoflavonoids from these species. A large body of literature has reported temporal and spatial correlations between phytoalexin accumulation and disease resistance in legumes, but a role for isoflavonoids in disease resistance has only recently been confirmed by genetic approaches (Dixon, 2001).

[1] This work was supported by the Samuel Roberts Noble Foundation, by the Oklahoma Center for the Advancement of Science and Technology, and by the National Science Foundation.

* Corresponding author; e-mail radixon@noble.org; fax 580–224–6692.

www.plantphysiol.org/cgi/doi/10.1104/pp.102.017319.

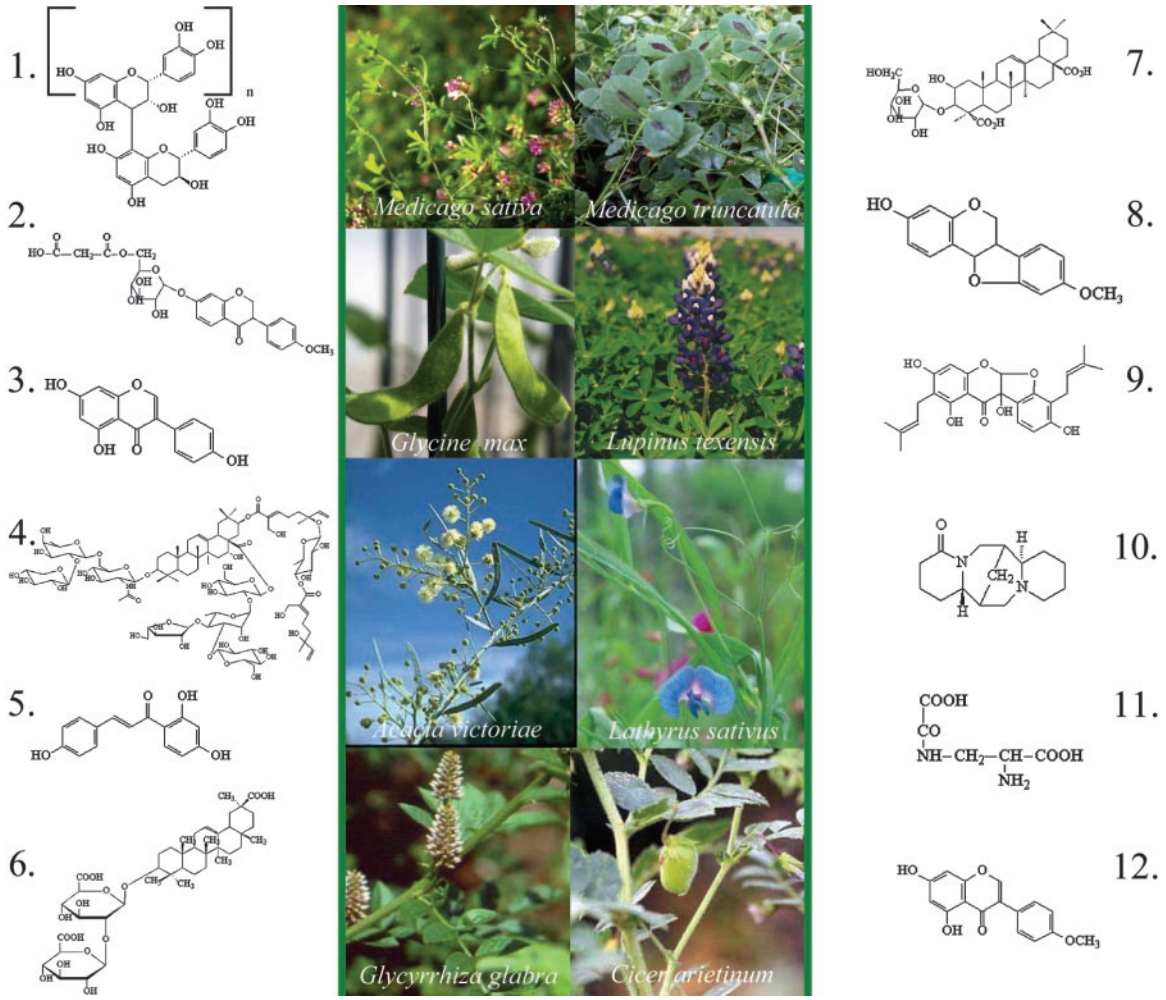

Figure 1. A small fraction of the biochemical diversity of legumes is shown in this selection of natural products from eight species. The compounds are seed coat proanthocyanidin from alfalfa (*Medicago sativa*; 1); formononetin malonyl glucoside, a constitutive isoflavone conjugate from roots of alfalfa and barrel medic (*Medicago truncatula*; 2); genistein, an isoflavone from seeds of soybean (*Glycine max*; 3); avicin D, a complex triterpene saponin from seed pods of *Acacia victoriae* (4); isoliquritigenin, a chalcone from roots of licorice (*Glycyrrhiza galbra*; 5); glycyrrhizin, a triterpene saponin from roots of licorice (6); medicagenic acid glucoside, a triterpene saponin from roots of alfalfa and barrel medic (7); medicarpin, a pterocarpan phytoalexin from fungally infected barrel medic and alfalfa (8); a prenylated isoflavone (lupinol A) from roots of *Lupinus* species (9); lupanine, a quinolizidine alkaloid from roots of *Lupinus* species (10); 3-*N*-oxalyl-L-2,3-diaminopropanoic acid (ODPA) from seed of grasspea (*Lathyrus sativus*; 11); and biochanin A, an isoflavone from seeds of chickpea (*Cicer arietinum*; 12).

Isoflavonoids are formed from flavanones (ubiquitously present in plants) by an unusual aryl migration reaction catalyzed by the cytochrome P450 enzyme CYP93C1 (2-hydroxyisoflavanone synthase, commonly termed isoflavone synthase [IFS]; Steele et al., 1999; Jung et al., 2000). It would appear that the *IFS* gene has arisen independently during evolution in taxonomically distinct families, because, in addition to their general occurrence in papilonoid legumes, isoflavonoids have also been reported in a few members of other families, including the Rosaceae, Chenopodiaceae, Apocynaceae, and Pinaceae.

Isoflavones exhibit estrogenic, antiangiogenic, antioxidant, and anticancer activities (Dixon, 1999; Dixon and Ferreira, 2002), and are now popular as dietary supplements (Palevitz, 2000). Genistein (Fig. 1) has been the subject of over 3,600 published studies (listed in Biological Abstracts) in the last 10 years. Major sources of isoflavones for humans are seed products of soybean (daidzein and genistein) and chickpea (biochanin A, Fig. 1), and the health-promoting activity of high-soy diets is believed to reside in their isoflavone components (Setchell and Cassidy, 1999; Lamartiniere, 2000; Palevitz, 2000). Epidemiological studies suggest a link between consumption of soy isoflavones and reduced risks of breast and prostate cancers in humans (Adlercreutz, 1998; Setchell and Cassidy, 1999; Lamartiniere, 2000).

Isoflavones may possess other health-promoting activities, including chemoprevention of osteoporo-

sis, and prevention of other postmenopausal disorders and cardiovascular disease (Alekel et al., 2000; MerzDemlow et al., 2000; Uesugi et al., 2001). A recent study indicated that a high-soy diet may even help improve cognitive function in students presented with a variety of complex mental tasks (File et al., 2001). Plants containing certain prenylated isoflavones have been used by the Zulus of South Africa for the treatment of impotency, and they appear to be active in improving erectile dysfunction (a kind of "phytoviagra"; Drewes et al., 2002).

Is it possible to introduce genistein or other isoflavones into vegetables, grains, and fruits for dietary disease prevention? Soybean IFS has been expressed in Arabidopsis, corn (*Zea mays*), and tobacco (*Nicotiana tabacum*). However, in all cases, only small amounts of genistein glycoconjugates were formed (Jung et al., 2000; Yu et al., 2000). Limiting factors for obtaining significant isoflavone accumulation in a heterologous target plant include limitation of IFS activity itself, limitation of precursor pools, and, most importantly, competition between IFS and other enzymes, such as flavanone 3-hydroxylase, that use the same substrate (naringenin; Liu et al., 2002). This competition may be indicative of metabolic channeling at the branch points for the formation of the various classes of flavonoids (Winkel-Shirley, 1999). Armed with this knowledge, it should now be possible to optimize isoflavonoid biosynthesis in nonlegumes to expand the delivery of dietary isoflavones and to develop new sources for the more complex bioactive isoflavonoids.

Important areas for future research on isoflavonoids include understanding flux control between isoflavonoid biosynthesis and competing pathways, deciphering the physical basis for association of biosynthetic enzymes in metabolic channels, and validating the various health-promoting effects ascribed to dietary provision of isoflavones. The latter point is of great importance if transgenic foods with value added health benefits are ever to make it to the market place.

TRITERPENE SAPONINS: COMPLEX MOLECULES WITH COMPLEX ACTIVITIES

All known classes of terpenoids have been reported within the Leguminosae. Particularly interesting are the triterpene saponins, whose biological activities can positively and negatively impact plant traits. Some saponins display allelopathic, antimicrobial, and anti-insect activity, but they can also be toxic to monogastric animals, act as antipalatability factors, or reduce forage digestibility in ruminants (Oleszek, 1996; Small, 1996; Oleszek et al., 1999). Monogastric animals often avoid consuming foods that contain saponins, and, therefore, development of saponin-free alfalfa is an agronomic target.

Saponins also have useful pharmacological activities. Many are anticholesterolemic or can act as adjuvants. The roots of the licorice plant (*Glycyrrhiza glabra*) are one of the oldest known botanicals in Chinese medicine. Health beneficial activities include anti-inflammation, antiulcer, antiallergy, and anticarcinogenesis, and the triterpene saponin glycyrrhizin (Fig. 1) may account for many of these properties, although licorice also contains bioactive chalcones (Fig. 1), isoflavans, diketones, and hydroxy-phenols (Wang and Nixon, 2001). Desert shrubs of the genus *Acacia* contain complex triterpene saponins, known as avicins, within the developing seedpods, where they presumably protect the seeds from predation. These compounds, which consist of an acacic acid triterpene skeleton conjugated to eight sugars and two linear monoterpenes (Fig. 1), are now under development as anticancer agents in view of their ability to induce cell cycle arrest in mammalian cells (Haridas et al., 2001). Their mode of action in target cells appears to involve induction of apoptosis by mitochondrial perturbation.

Most of the steps in the biosynthesis of triterpene saponins remain uncharacterized at the molecular level. The model legume barrel medic contains a complex mixture of saponins (Huhman and Sumner, 2002), including glycosides of medicagenic acid (Fig. 1), some of which have also previously been found in soybean. The first committed step in their biosynthesis is catalyzed by a specific oxidosqualene cyclase, β-amyrin synthase. β-Amyrin synthase has been functionally characterized from several plants, including pea (*Pisum sativum*; Morita et al., 2000) and barrel medic (Suzuki et al., 2002), and is closely related to plant cycloartenol synthase involved in sterol biosynthesis. The steps between β-amyrin and the various saponin aglycones produced in *Medicago* and soybean involve a series of oxidative reactions that, by analogy to similar reactions in brassinosteroid biosynthesis, probably are catalyzed by cytochrome P450 enzymes. The aglycones are subsequently converted to the saponins by the action of a series of glycosyltransferases (GTs). To date, only a single GT involved in saponin biosynthesis in soybean has been characterized biochemically (Kurosawa et al., 2002). This pathway is a prime candidate for functional genomics approaches (see below).

Important areas for future research on triterpene saponins for legume improvement and commercial exploitation include obtaining a basic understanding of their biosynthesis from initial cyclization to final conjugation, discovering regulatory genes for coordinated up-regulation of triterpene pathways, and using transgenic approaches to learn more about triterpene function as a basis for genetic modification studies.

NATURAL PRODUCTS AFFECTING THE NUTRITIONAL QUALITY OF FORAGE LEGUMES

Condensed tannins (CTs, also known as proanthocyanidins) are polymers or oligomers of flavan-3-ol

units derived from the flavonoid pathway. They are common components of seed coats throughout the plant kingdom. CTs are found in many legumes with a tree-like habit, and occur in the leaves of some forage legumes such as bird's foot trefoil (*Lotus corniculatus*) and sanfoin (*Onobrychis viciifolia*). Their structures can be quite variable, among the commonest being a series of four to eight linked (-)-epicatechin units terminating in a catechin unit (Fig. 1), as found in the alfalfa seed coat (Koupai-Abyazani et al., 1993). CTs most likely play a protective function within the plant, but are now attracting attention because of their widespread effects on human health and ruminant nutrition. They are powerful antioxidants with beneficial effects on cardiac health and immunity (Bagchi et al., 2000; Lin et al., 2002). CTs from fruits such as cranberry (*Vaccinum macrocarpon*) protect against urinary tract infections (Foo et al., 2000), and the CTs and their precursors (catechins and epicatechins) are important for determining flavor and astringency in wines and tea, while at the same time conferring potential health beneficial effects to these beverages (Ahmad et al., 2000; Bagchi et al., 2000). Chocolate contains CTs, and the health-promoting effects of this important (for some!) dietary component is now being publicized by the industry.

The most important agronomic trait involving CTs is their ability to prevent the "bloating" characteristics of forage legumes such as alfalfa and white clover (*Trifolium repens*) that lack CTs in the consumed aerial portions (Aerts et al., 1999; McMahon et al., 2000). Farmers, ranchers, and those city dwellers who have read Thomas Hardy's "Far from the Madding Crowd" will be well aware of the suffering and economic loss caused by pasture bloat. CTs bind to dietary proteins in the rumen and thereby slow down their rate of degradation, preventing bloat and improving the animal's nitrogen nutrition by increasing the amount of dietary protein exiting the rumen (Aerts et al., 1999; Barry and McNabb, 1999; Douglas et al., 1999). This can lead to increased body weight and wool production (Douglas et al., 1999). In addition to reducing bacterial degradation of proteins in the rumen, CTs can also slow down protein degradation during ensiling of forage legumes, thereby improving the nitrogen nutritional value of the feed.

Although the flavonoid pathway has been extensively studied by chemists, biochemists, and geneticists for over 60 years, the enzymatic formation of the 2,3-cis-flavan-3-ol [(-)-epicatechin] unit that forms the major portion of most CTs has, until recently, remained a mystery. Mutations in the *BANYULS* (*BAN*, named after the color of a French red wine) gene in Arabidopsis result in a transparent testa (*tt*), associated with a lack of CTs and precocious accumulation of anthocyanins in the seed coat (Devic et al., 1999). On the basis of this phenotype and the amino acid sequence similarity of BAN to a reductase of flavonoid biosynthesis (dihydroflavonol reductase), it was suggested that *BAN* encodes leucoanthocyanidin reductase (Devic et al., 1999), a yet poorly characterized enzyme proposed to convert flavan-3,4-diols to 2,3-trans-flavan-3-ols such as (+)-catechin, the "starter unit" for CT condensation. It has now been shown that the *BAN* genes from Arabidopsis and barrel medic encode a new enzyme, anthocyanidin reductase, that converts cyanidin to 2,3-cis-(-)-epicatechin (Xie et al., 2002). Therefore, anthocyanins are not, as previously believed, only end products of flavonoid metabolism. Although *BAN* expression in barrel medic is primarily limited to young seed coats, transgenic expression of barrel medic BAN in tobacco leads to accumulation of CTs throughout the pigmented portions of the petals, with concomitant reduction in anthocyanin levels (Xie et al., 2002). These results suggest that it should soon be possible to engineer CT accumulation in forage legumes for protection of animals against pasture bloat.

Lignin is a phenylpropanoid polymer found in all higher plants, and an important factor affecting cell wall digestibility in forage legumes. Lignin levels increase with progressive maturity in stems of many forage legumes, including alfalfa (Jung et al., 1997). In addition, the lignin composition often changes with advanced maturity toward a progressively higher syringyl to guaiacyl (S/G) ratio, reflecting an increased degree of methylation of the lignin. Some studies have linked decreased forage digestibility to increased S/G ratio as a function of increased plant maturity (Buxton and Russell, 1988; Grabber et al., 1992), whereas others have questioned the effect of lignin composition on digestibility (Grabber et al., 1997). It has been estimated by the U.S. Dairy and Forage Research Center that a 10% increase in fiber digestibility for forage legumes and grasses would result in an annual $350 million increase in milk/beef production and a 2.8 million ton reduction in annual manure solids (http://dfrc.wisc.edu/research).

Genetic manipulation of lignin levels in forage legumes has, to date, targeted just three of the 10 or more enzymes involved in the formation of the guaiacyl and syringyl lignin monomers (monolignols). Antisense reduction of caffeic acid 3-*O*-methyltransferase (COMT) to less than 5% of wild-type values in the tropical pasture legume *Stylosanthes humilis* resulted in no apparent reduction in overall lignin levels but in a strong reduction in S lignin (Rae et al., 2001). In vitro digestibility of stem material in rumen fluid was increased by up to 10% in the transgenic plants exhibiting strongest COMT down-regulation. Up to 30% decreases in Klason lignin levels, near elimination of S lignin, and appearance of novel benzodioxane units in the lignin fraction, were observed in transgenic alfalfa in which COMT down-regulation was targeted using the vascular tissue-specific bean *PAL2* promoter (Guo et al.,

2000; Marita et al., 2003). Forage material from COMT down-regulated alfalfa plants had significantly increased neutral detergent fiber, acid detergent fiber, and in vitro true digestibility. In-rumen digestibility was increased by up to 4% in a series of replicated analyses in fistulated steers (Guo et al., 2001).

Near elimination of caffeoyl coenzyme A 3-*O*-methyltransferase (CCoAOMT) activity in transgenic alfalfa reduced G lignin by up to 50% in some lines, but had no effect on S lignin (Guo et al., 2000). CCoAOMT-down-regulated plants had a significant decrease in overall lignin content and increased neutral detergent fiber, acid detergent fiber, and in vitro true digestibility. In-rumen digestibility of CCoAOMT-down-regulated alfalfa forage was increased by up to 6%.

Antisense down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa to approximately 30% of wild-type level led to a red coloration of the stem and a reduction in lignin S/G ratio primarily due to a decrease in S units (Baucher et al., 1999). The most strongly down-regulated plants exhibited increased in situ digestibility of dry matter in cannulated sheep. Taken together, the above studies indicate the success of transgenic approaches for improvement of forage quality in alfalfa, and new cultivars incorporating these traits may soon reach the market.

Major areas for future research on lignins and proanthocyanidins for forage legume improvement include development of improved analytical methods for determining the content and composition of polymeric phenylpropanoids and flavonoids, understanding the exact relationships between lignin and proanthocyanidin content and composition and forage quality, developing better and more predictable approaches for engineering the lignin polymer, and understanding all the factors necessary for synthesis and assembly of proanthocyanidins in leaf tissues.

GENOMICS AND METABOLOMICS AS KEY TECHNOLOGIES FOR DECIPHERING LEGUME NATURAL PRODUCT BIOSYNTHESIS

The success of large-scale genome and expressed sequence tag (EST) sequencing projects has greatly expanded the scale on which natural product biosynthesis and biological systems in general can be addressed. Extensive DNA sequence resources are currently available for soybean (primarily in the form of EST sequences; http://129.186.26.94/) and barrel medic (EST and genomic sequence; https://xgi.ncgr.org/mgi/; http://www.tigr.org/tdb/tgi/mtgi/; http://www.genome.ou.edu/medicago.html; http://www.medicago.org/; and http://sequence.toulouse.inra.fr/). The sequences of many genes encoding enzymes of natural product biosynthesis are already present in these databases. The question is how to identify them. One answer is to apply functional genomic approaches that encompass global assessment of the transcriptome and the metabolome. The comprehensive profiling of large numbers of metabolites can be used to assess gene function and to query holistic responses of biological systems to external stimuli (Fiehn, 2002). This approach is the key means to qualitatively and quantitatively defining the chemical phenotype (chemotype) of a genetically or environmentally perturbed biological system.

The current and prevailing opinion is that no single technique will provide a comprehensive assessment of the chemically complex metabolome, particularly when considering the chemical diversity of natural products; thus, multiple tools must be used (Hall et al., 2002). These include thin-layer chromatography, infrared spectroscopy, NMR, gas chromatography/mass spectrometry (GC/MS), liquid chromatography with UV or MS detection, liquid chromatography/MS/MS, capillary electrophoresis, and capillary electrophoresis/MS (Sumner et al., 2002, 2003).

The best methods for chemical profiling of flavonoids and isoflavonoids use HPLC for separation coupled to UV absorption and/or mass selective detection (Sumner et al., 1996; Lin et al., 2000; Bednarek et al., 2001; Liu et al., 2002). The need to profile intact glycosidic conjugates, the relatively high M_r of the conjugates, and the multiplicity of polar hydroxyl groups obviate against the use of GC separation because of the need for extensive derivatization and the limited *m/z* range of most commercial GC/MS instruments. In addition to a plethora of reports on the extraction and identification of individual compounds, specific protocols for the routine profiling of the flavonoid and isoflavonoid complements of various legume species, including red clover (*Trifolium pratense*), soybean, and lupin, a rich source of prenylated isoflavonoids (Fig. 1), have been published (Graham, 1991; Lin et al., 2000; Bednarek et al., 2001).

Saponins contain poor chromophores; thus, the preferred method for profiling the triterpene glycoside complement of legumes such as alfalfa and barrel medic is reversed-phase HPLC coupled with electrospray-ionization MS (Huhman and Sumner, 2002; Sumner et al., 2002). Using this technique, it has been demonstrated that the model legume barrel medic contains a more complex mixture of triterpenes than found in the closely related and previously well-studied species alfalfa. Five different β-amyrin-derived triterpene aglycones, soyasapogenol B, soyasapogenol E, medicagenic acid, hederagenin, and bayogenin were found to be the core of the 27 barrel medic saponins identified (Huhman and Sumner, 2002).

Analysis of lignin presents far greater technical challenges than for most other natural products because it is a complex insoluble heteropolymer. The

reader is referred elsewhere for a description of the problems and some current approaches (Dean, 1997; Lu and Ralph, 1997; Marita et al., 2002). Profiling of proanthocyanidin polymers is likewise challenging, and most studies rely on simple chemical extraction and colorimetric determination (Schofield et al., 2001), protocols that do not provide structural information. Structures have been determined for several legume proanthocyanidins (e.g. Koupai-Abyazani et al., 1993), but the methods fall far short of high throughput profiling and new approaches are needed.

The following example outlines the utility of genomics coupled to metabolomics in deciphering a biosynthetic pathway, the formation of the β-amyrin-derived triterpenes in barrel medic, for which the exact route of biosynthesis is experimentally undetermined. In-depth targeted metabolite profiling with the approaches outlined above is first used to determine the exact complement of the metabolites of interest and their potential precursors. The best approach compares tissues that make the compound(s) in question with those that do not (Dixon, 2001). Therefore, the biological system can be a particular species, set of ecotypes, or group of species. Alternatively, an inducible system such as elicited roots or cell cultures can be studied. In the case of the triterpenes, elicitation of cell suspension cultures with methyl jasmonate results in a striking induction of the compounds and, presumably, their biosynthetic enzymes (Suzuki et al., 2002). Based on the metabolite profiles and perhaps already existing knowledge, a tentative pathway can be proposed; this consists essentially of P450- and GT-catalyzed reactions in the case of triterpene saponins (Suzuki et al., 2002). If not already available, cDNA libraries are then made from tissues of the plant in which the particular chemistry is active, and high-throughput EST sequencing performed. In barrel medic, there are more than 250 expressed cytochrome P450s and nearly 300 expressed GTs, based on EST counting in the more than 30 cDNA libraries sequenced to date. Bioinformatic approaches such as use of self-organizing maps for in silico expression analysis of EST libraries, coupled with DNA array analysis of transcripts from at least two cell or tissue types (chemical producers and nonproducers), can then provide a shortcut to candidate gene identification from among, in this particular case, the approximately 600 candidate P450s and GTs. The number of candidates revealed in this way (approximately 20 of each class; L. Achnine and R.A. Dixon, unpublished data) is small enough for direct expression studies in a heterologous system such as *Escherichia coli* or yeast. When dealing with complex pathways for which intermediates are unavailable, parallel approaches such as stable or transient down-regulation of candidate genes coupled with metabolite profiling may also have to be used.

STRUCTURAL BIOLOGY: A KEY TO THE FUTURE OF NATURAL PRODUCT BIOSYNTHESIS

Although a bewildering array of plant natural products exists in nature, most are constructed using a relatively small number of enzyme types, e.g. polyketide synthase, terpene cyclase, reductase, acyltransferase, *O*-methyltransferase, etc. A better understanding of the relationship between amino acid sequence and catalytic activity is the key to a better prediction of function for enzymes of natural product biosynthesis based on primary DNA sequence information. This requires knowledge of structure-function relationships among the various classes of enzymes involved in natural product biosynthesis. This knowledge is just beginning to appear. For example, the three-dimensional structures of a number of natural product pathway enzymes from alfalfa, including three *O*-methyltransferases of flavonoid, isoflavonoid, and lignin biosynthesis, have recently been solved by x-ray crystallography (Zubieta et al., 2001, 2002). Information of this type not only increases predictive ability for functional genomics, but also facilitates structure-directed modification of catalytic activity (Jez et al., 2002; Zubieta et al., 2002) for the generation of novel natural products through in vivo transgenic approaches or, ultimately, in vitro combinatorial biochemistry.

CONCLUDING REMARKS

Functional and structural genomics, coupled with increases in the resolving power of metabolomics, are poised to make a huge impact on our understanding of plant natural product biosynthesis, and, therefore, on our ability to harness nature's wonderful chemical diversity for the benefit of humankind. These approaches take advantage of the rapidly decreasing costs of DNA sequencing to generate databases for natural product gene discovery. Legumes will be at the forefront of these endeavors because these species combine emerging genomic accessibility with chemistry that is of relevance for plant, human, and animal health.

Genomics will also provide new approaches for discovering the transcriptional regulators that control expression of natural product biosynthetic enzymes (Meissner et al., 1999). Whereas many will probably fall into the known classes of transcription factors currently known to regulate phenylpropanoid and terpenoid biosynthesis (Dixon et al., 2002; Vom Endt et al., 2002), new types of factors may perhaps exist in legumes (e.g. Lindsay et al., 2002). Metabolic engineering by ectopic expression of transcription factors (Borevitz et al., 2001; Vom Endt et al., 2002) holds great promise for exploiting legumes as factories for production of bioactive secondary metabolites.

Received November 6, 2002; returned for revision December 3, 2002; accepted December 10, 2002.

LITERATURE CITED

Adlercreutz M (1998) Epidemiology of phytoestrogens. Baillieres Clin Endocrinol Metab **12:** 605–623

Aerts RJ, Barry TN, McNabb WC (1999) Polyphenols and agriculture: beneficial effects of proanthocyanidins in forages. Agric Ecosystems Environ **75:** 1–12

Ahmad N, Gupta S, Mukhtar H (2000) Green tea polyphenol epigallocatechin-3-gallate differentially modulates nuclear factor κB in cancer cells versus normal cells. Arch Biochem Biophys **376:** 338–346

Alekel DL, StGermain A, Pererson CT, Hanson KB, Stewart JW, Toda T (2000) Isoflavone-rich soy protein isolate attenuates bone loss in the lumbar spine of perimenopausal women. Am J Clin Nutr **72:** 844–852

Bagchi D, Bagchi M, Stohs SJ, Das DK, Ray SD, Kuszynski CA, Joshi SS, Pruess HG (2000) Free radicals and grape seed proanthocyanidin extract: importance in human health and disease prevention. Toxicology **148:** 187–197

Barry TN, McNabb WC (1999) The implications of condensed tannins on the nutritive value of temperate forages fed to ruminants. Br J Nutr **81:** 263–272

Baucher M, BernardVailhe MA, Chabbert B, Besle JM, Opsomer C, VanMontagu M, Botterman J (1999) Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the effect on lignin composition and digestibility. Plant Mol Biol **39:** 437–447

Bednarek P, Franski R, Kerhoas L, Einhorn J, Wojtaszek P, Stobiecki M (2001) Profiling changes in metabolism of isoflavonoids and their conjugates in *Lupinus albus* treated with biotic elicitor. Phytochemistry **56:** 77–85

Bell C, Dixon RA, Farmer AD, Flores R, Inman J, Gonzales RA, Harrison MJ, Paiva NL, Scott AD, Weller JW et al. (2001) The *Medicago* genome initiative: a model legume database. Nucleic Acids Res **29:** 114–117

Borevitz J, Xia Y, Blount JW, Dixon RA, Lamb C (2001) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12:** 2383–2393

Buxton DR, Russell JR (1988) Lignin constituents and cell-wall digestibility of grass and legume stems. Crop Sci **28:** 553–558

Dean JFD (1997) Lignin Analysis. *In* WV Dashek, ed, Methods in Plant Biochemistry and Molecular Biology. CRC Press, Boca Raton, FL, pp 199–215

Devic M, Guilleminot J, Debeaujon I, Bechtold N, Bensaude E, Koornneef M, Pelletier G, Delseny M (1999) The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. Plant J **19:** 387–398

Dixon RA (1999) Isoflavonoids: biochemistry, molecular biology and biological functions. *In* U Sankawa, ed, Comprehensive Natural Products Chemistry, Vol 1. Elsevier, Oxford, pp 773–823

Dixon RA (2001) Natural products and disease resistance. Nature **411:** 843–847

Dixon RA, Achnine L, Kota P, Liu C-J, Reddy MS, Wang L (2002) The phenylpropanoid pathway and plant defense: a genomics perspective. Mol Plant Pathol **3:** 371–390

Dixon RA, Ferreira D (2002) Molecules of interest: genistein. Phytochemistry **60:** 205–211

Douglas GB, Stienezen M, Waghorn GC, Foote AG, Purchas RW (1999) Effect of condensed tannins in birdsfoot trefoil (*Lotus corniculatus*) and sulla (*Hedysarum coronarium*) on body weight, carcass fat depth, and wool growth of lambs in New Zealand. New Zealand J Agric Res **42:** 55–64

Drewes SE, Horn MM, Munro OQ, Dhlamini JTB, Meyer JJM, Rakuambo NC (2002) Pyrano-isoflavones with erectile-dysfunction activity from *Eriosema kraussianum*. Phytochemistry **59:** 739–747

Fiehn O (2002) Metabolomics: the link between genotypes and phenotypes. Plant Mol Biol **48:** 155–171

File SE, Jarrett N, Fluck E, Duffy R, Casey K, Wiseman H (2001) Eating soya improves human memory. Psychopharmacology **157:** 430–436

Foo LY, Lu Y, Howell AB, Vorsa N (2000) The structure of cranberry proanthocyanidins which inhibit adherence of uropathogenic P-fimbriated *Escherichia coli* in vitro. Phytochemistry **54:** 173–181

Grabber JH, Jung GA, Abrams SM, Howard DB (1992) Digestion kinetics of parenchyma and sclerenchyma cell walls isolated from orchardgrass and switchgrass. Crop Sci **32:** 806–810

Grabber JH, Ralph J, Hatfield RD, Quideau S (1997) *p*-Hydroxyphenyl, guaiacyl, and syringyl lignins have similar inhibitory effects on wall degradability. J Agric Food Chem **45:** 2530–2532

Graham TL (1991) A rapid, high resolution high performance liquid chromatography profiling procedure for plant and microbial aromatic secondary metabolites. Plant Physiol **95:** 584–593

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2000) Down-regulation of caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa (*Medicago sativa* L.): impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell **13:** 73–88

Guo D, Chen F, Wheeler J, Winder J, Selman S, Peterson M, Dixon RA (2001) Improvement of in-rumen digestibility of alfalfa forage by genetic manipulation of lignin *O*-methyltransferases. Transgenic Res **10:** 457–464

Hall R, Beale M, Fiehn O, Hardy N, Sumner LW, Bino R (2002) Plant metabolomics as the missing link in functional genomics strategies. Plant Cell **14:** 1437–1440

Hanbury CD, White CL, Mullan BP, Siddique KHM (2000) A review of the potential of *Lathyrus sativus* L. and *L. cicera* L. grain for use as animal feed. Animal Feed Sci Technol **87:** 1–27

Haridas V, Higuchi M, Jayatilake GS, Bailey D, Mujoo K, Blake ME, Arntzen CJ, Gutterman JU (2001) Avicins: triterpenoid saponins from *Acacia victoriae* (Bentham) induce apoptosis by mitochondrial perturbation. Proc Natl Acad Sci USA **98:** 5821–5826

Huhman DV, Sumner LW (2002) Metabolic profiling of saponin glycosides in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry **59:** 347–360

Jez JM, Bowman ME, Noel JP (2002) Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. Proc Natl Acad Sci USA **99:** 5319–5324

Jung HG, Mertens DR, Payne AJ (1997) Correlation of acid detergent lignin and Klason lignin with digestibility of forage dry matter and neutral detergent fiber. J Dairy Sci **80:** 1622–1628

Jung W, Yu O, Lau S-MC, O'Keefe DP, Odell J, Fader G, McGonigle B (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat Biotechnol **18:** 208–212

Koupai-Abyazani MR, McCallum J, Muir AD, Lees GL, Bohm BA, Towers GHN, Gruber MY (1993) Purification and characterization of a proanthocyanidin polymer from seed of alfalfa (*Medicago sativa* cv. beaver). J Agric Food Chem **41:** 565–569

Kurosawa Y, Takahara H, Shiraiwa M (2002) UDP-glucuronic acid: soyasapogenol glucuronosyltransferase involved in saponin biosynthesis in germinating soybean seeds. Planta **215:** 620–629

Lamartiniere CA (2000) Protection against breast cancer with genistein: a component of soy. Am J Clin Nutr **71:** 1705S–1707S

Lin L-C, Kuo Y-C, Chou C-J (2002) Immunomodulatory proanthocyanidins from *Ecdysanthera utilis*. J Nat Prod **65:** 505–508

Lin LZ, He XG, Lindenmaier M, Yang J, Cleary M, Qiu SX, Cordell GA (2000) LC-ESI-MS study of the flavonoid glycoside malonates of red clover (*Trifolium pratense*). J Agric Food Chem **48:** 354–365

Lindsay WP, McAlister F, Zhu Q, He X-Z, Droge-Laser W, Hedrick S, Doerner P, Lamb C, Dixon RA (2002) KAP-2, a protein that binds to the H-box in a bean chalcone synthase promoter, is a novel plant transcription factor with sequence identity to the large subunit of human Ku autoantigen. Plant Mol Biol **49:** 503–514

Liu C-J, Blount JW, Steele CL, Dixon RA (2002) Bottlenecks for metabolic engineering of isoflavone glycoconjugates in *Arabidopsis*. Proc Natl Acad Sci USA **99:** 14578–14583

Lu FC, Ralph J (1997) Derivatization followed by reductive cleavage (DFRC method), a new method for lignin analysis: protocol for analysis of DFRC monomers. J Agric Food Chem **45:** 2590–2592

Marita JM, Ralph J, Hatfield RD, Guo D, Chen F, Dixon RA (2003) Structural and compositional modifications in lignin of transgenic alfalfa down-regulated in caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase. Phytochemistry **62:** 53–65

McMahon LR, McAllister TA, Berg BP, Majak W, Acharya SN, Popp JD, Coulman BE, Wang Y, Cheng KJ (2000) A review of the effects of forage condensed tannins on ruminal fermentation and bloat in grazing cattle. Can J Plant Sci **80:** 469–485

Meissner RC, Jin HL, Cominelli E, Denekamp M, Fuertes A, Greco R, Kranz HD, Penfield S, Petroni K, Urzainqui A, Martin C et al. (1999) Function search in a large transcription factor gene family in Arabidopsis: assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. Plant Cell **11:** 1827–1840

MerzDemlow BE, Duncan AM, Wangen KE, Xu X, Carr TP, Phipps WR, Kurzer MS (2000) Soy isoflavones improve plasma lipids in normocholesterolemic, premenopausal women. Am J Clin Nutr **71:** 1462–1469

Morita M, Shibuya M, Kushiro T, Masuda K, Ebizuka Y (2000) Molecular cloning and functional expression of triterpene synthases from pea (*Pisum sativum*). New alpha-amyrin-producing enzyme is a multifunctional triterpene synthase. Eur J Biochem **267:** 3453–3460

Oleszek W (1996) Alfalfa Saponins: Structure, Biological Activity, and Chemotaxonomy. Plenum Press, New York

Oleszek W, Junkuszew M, Stochmal A (1999) Determination and toxicity of saponins from *Amaranthus cruentus* seeds. J Agric Food Chem **47:** 3685–3687

Palevitz BA (2000) Soybeans hit main street. Scientist **14:** 8–9

Quackenbush J, Liang F, Holt I, Pertea G, Upton J (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res **28:** 141–145

Rae AL, Manners JM, Jones RJ, McIntyre CL, Lu DY (2001) Antisense suppression of the lignin biosynthetic enzyme, caffeate *O*-methyltransferase, improves in vitro digestibility of the tropical pasture legume, *Stylosanthes humilis*. Aust J Plant Physiol **28:** 289–297

Retief FP, Cilliers L (2002) Legumes and disease in the Graeco-Roman world. Suid Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie **20:** 9–13

Schofield P, Mbugua DM, Pell AN (2001) Analysis of condensed tannins: a review. Animal Feed Sci Technol **91:** 21–40

Setchell KDR, Cassidy A (1999) Dietary isoflavones: biological effects and relevance to human health. J Nutr **129:** 758S–767S

Small E (1996) Adaptations to herbivory in alfalfa (*Medicago sativa*). Can J Bot **74:** 807–822

Steele CL, Gijzen M, Qutob D, Dixon RA (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. Arch Biochem Biophys **367:** 147–150

Sumner L, Paiva NL, Dixon RA, Geno PW (1996) High-performance liquid chromatography/continuous-flow liquid secondary ion mass spectrometry of flavonoid glucosides in leguminous plant extracts. J Mass Spectrom **31:** 472–485

Sumner LW, Duran AL, Huhman DV, Smith JT (2002) Metabolomics: a developing and integral component in functional genomic studies of *Medicago truncatula*. *In* JT Romeo, RA Dixon, eds, Phytochemistry in the Genomics and Post-Genomics Eras. Elsevier, Oxford, pp 31–61

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry (in press)

Suzuki H, Achnine L, Xu R, Matsuda SPT, Dixon RA (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. Plant J **32:** 1033–1048

Uesugi T, Toda T, Tsuji K, Ishida H (2001) Comparative study on reduction of bone loss and lipid metabolism abnormality in ovariectomized rats by soy isoflavones, daidzin, genistin, and glycitin. Biol Pharm Bull **24:** 368–372

Vom Endt D, Kijne JW, Memelink J (2002) Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry **61:** 107–114

Wang ZY, Nixon DW (2001) Licorice and cancer. Nutr Cancer **39:** 1–11

Wink M, Waterman D (1999) Chemotaxonomy in relation to molecular phylogeny of plants. Annu Plant Rev **2:** 300–341

Winkel-Shirley B (1999) Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiol Plant **107:** 142–149

Xie D, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2002) Role of anthocyanidin reductase, encoded by BANYULS, in plant flavonoid biosynthesis. Science **299:** 396–399

Yu O, Jung W, Shi J, Croes RA, Fader GM, McGonigle B, Odell JT (2000) Production of the isoflavones genistein and daidzein in non-legume dicot and monocot tissues. Plant Physiol **124:** 781–794

Zubieta C, Dixon RA, Noel JP (2001) Crystal structures of chalcone *O*-methyltransferase and isoflavone *O*-methyltransferase reveal the structural basis for substrate specificity in plant *O*-methyltransferases. Nat Struct Biol **8:** 271–279

Zubieta C, Kota P, Ferrer J-L, Dixon RA, Noel J (2002) Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-*O*-methyltransferase. Plant Cell **14:** 1265–1277