1
Biosynthesis and Chemical Properties of Natural Substances in Plants

The number of known so-called "secondary metabolites" (also referred to as "natural products") that have been discovered to date is increasing at a constant rate. Yet, it is not only plants (as described in this book) that produce these bioactive compounds; rather, other organisms such as bacteria, fungi, sponges, as well as animals, are also capable of synthesizing a plethora of these metabolites.

Whilst some of these metabolites are discussed in Chapters 4 and 5, a large number remain undiscovered. Moreover, secondary metabolites often possess interesting pharmacological properties, and therefore their characterization is very important. It should not be forgotten that plants synthesize these compounds as part of their own survival strategies, typically as defense compounds or as signals for pollinators or symbionts. In addition, recent evidence has pointed to additional roles for secondary metabolites in plant development. Although the term "secondary metabolites" perhaps infers a less important role for these compounds than those involved in primary metabolism, this is not the case. In fact, many essential and nonessential compounds in this group are found in plants, and even so-called "nonessential materials" can play a role in a plant's responses against abiotic and biotic stress. In this situation, the deletion of a biosynthetic pathway would cause damage to the plant, even if the pathway was not needed under favorable conditions. Interest in the secondary metabolites of plants was further increased when more sensitive analytical instruments became available, as well as genome sequence data for many plant species. Together, these tools allow the details of biosynthetic pathways to be investigated, an example being biotechnological manipulation. Notably, attention also began to be focused on the evolutionary aspects of secondary metabolite synthesis.

In general, secondary metabolites occur as complex mixtures (Figure 1.1). Their biosynthesis can be influenced by a variety of factors during development, in addition to stress, which makes the determination of their complete pattern essentially impossible. Whilst secondary metabolites can occur in the tissues as active compounds, they can also be synthesized as inactive compounds that must be transformed into active products; such metabolites that pre-exist are known as "phytoanticipins." Compounds that are biosynthesized under stress conditions are typically not detectable in unstressed tissues; when they are synthesized after the invasion of plants by various pests these metabolites are termed "phytoalexins."

Plant Natural Products: Synthesis, Biological Functions and Practical Applications, First Edition.
Herwig O. Gutzeit and Jutta Ludwig-Müller.

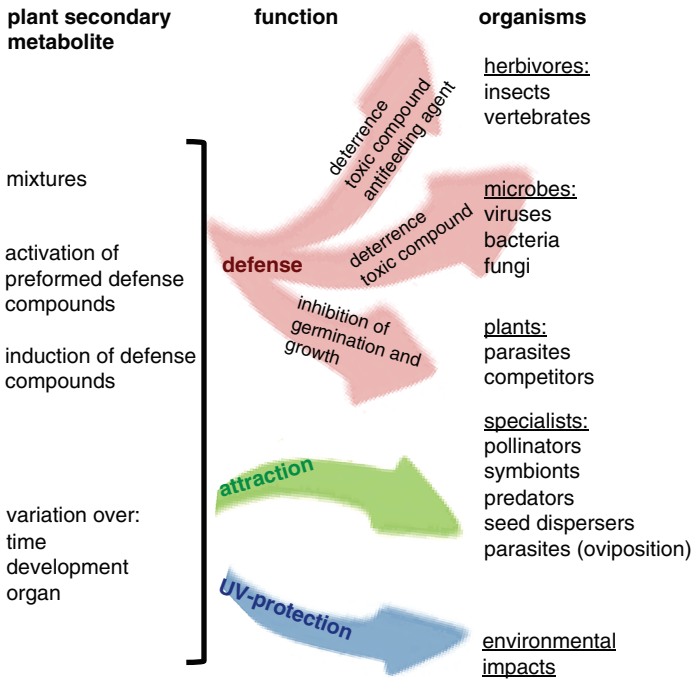


Figure 1.1 Correlation between secondary metabolites, their occurrence, and various functions. Secondary metabolites occur always as complex mixtures, and each plant has a specific set of compounds. The composition varies over time, development, and in different tissues and organs. In addition, the metabolites can be inactive and must be metabolized to the active form. Others will be produced only in response to biotic stress factors and other stress situations. The latter have a function in deterring a variety of herbivores, but can also attract insects for pollination.

Secondary metabolites play a variety of roles in the interaction of plants with their environment, but they may also be involved in plant development. They can function in defense mechanisms against herbivores such as nematodes, insects or vertebrates, as well as against microbes and competing plants. Metabolites may also play important roles in the attraction of symbionts, such as insects for pollination or zoophagous insects in multitrophic (when more than two organisms are involved) defense mechanisms. Animals that are required for seed-dispersal purposes are attracted by color and aroma compounds; however, some specialist insects and vertebrates have developed mechanisms to cope with the toxic effects of these compounds, and may generate an ecological niche in recognizing a specific host plant for either feeding or oviposition. Abiotic signals also contribute to the specific patterns of secondary metabolites; for example, some compounds are good protectors against ultraviolet (UV) light and photo-oxidation (i.e., oxidation reactions induced by light).

In order to understand the complexity of these compounds it is essential to have a good knowledge of their biosynthesis and the regulation of genes that

encode the enzymes involved in syntheses, under changing conditions. A brief description of the biosynthesis of selected compounds is provided in the following sections of this chapter, but for additional information relating to the enzymes and their biochemistry, the reader is referred to respective textbooks on plant physiology and plant biochemistry.

1.1 Selected Classes of Secondary Metabolites

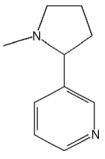
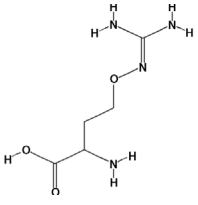
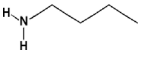
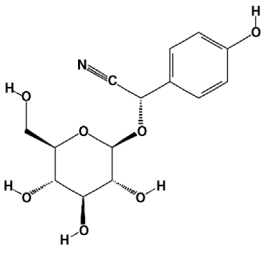
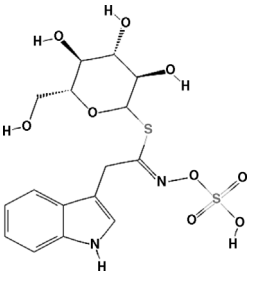
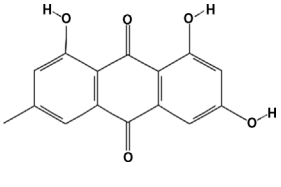
1.1.1 Occurrence and Compartmentation

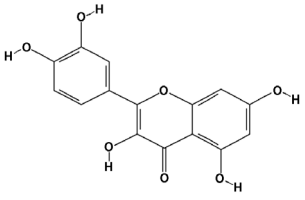
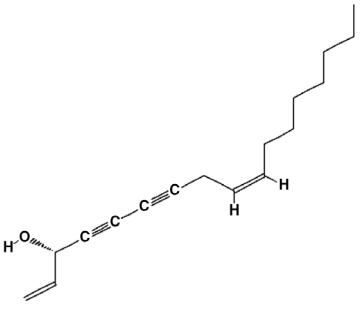
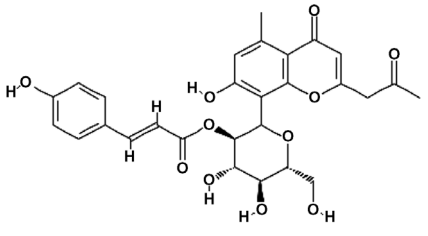
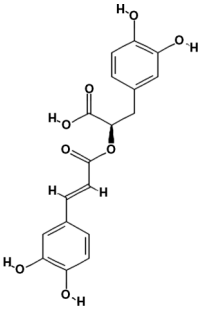
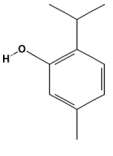
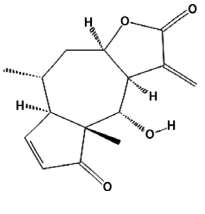
The patterns of secondary metabolites will differ depending on the species, the developmental stages, and the stress situations occurring in that species. The differences in pattern are the result of the metabolite biosynthesis, and the regulation of genes encoding the biosynthetic enzymes, transcription factors and transport proteins (see Sections 1.1.3 and 1.1.4). It is therefore difficult to generalize the metabolic patterns for a given plant species. The occurrence of some metabolites can be confined to a certain plant species, such as betalains in the Caryophyllales, glucosinolates in the Brassicales (with few exceptions), and polyacetylenes mainly in the Asteraceae and Apiaceae families. In the latter case, the compounds are important as bioactive substances in food crops (see Chapter 6) such as carrot (*Daucus carota*), celery (*Apium graveolens*) and fennel (*Foeniculum vulgare*). A compound from the Asteraceae *Carlina acaulis* has a strong antimicrobial activity and is also toxic for trypanosomes. Based on their chemical structures, the secondary metabolites can be placed into different classes (Table 1.1).

The synthesis of secondary metabolites can occur in all plant organs, including the roots, shoots, leaves, flowers, fruit, and seeds. Some metabolites are stored in specific compartments, which may be either whole organs or specialized cell types. Within these compartments the concentration of toxic secondary metabolites may be very high, so that they can exert an efficient defense against herbivores. For example, the glandular trichomes are often rich in toxic compounds, and some pathways – an example being the synthesis of the labdane diterpene Z-abienol in tobacco – occur exclusively in these trichomes. The glandular trichomes of peppermint leaves also contain large amounts of aromatic oils. The latex of *Euphorbia* is a good example of a toxic mixture of compounds being stored in specialized cells, the laticifers (Figure 1.2).

Lipophilic compounds can be found in the membranes of organelles or in the endoplasmic reticulum (ER), although specific storage vesicles have been reported for some compounds, such as the alkaloid berberin. Hydrophilic compounds are stored preferentially in vacuoles in the inner parts of the organelles and in the cytosol, as well as in the extracellular space (apoplast). Biosynthesis can occur in the cytosol, organelles and the ER, whereas transcriptional control takes place in the nucleus.

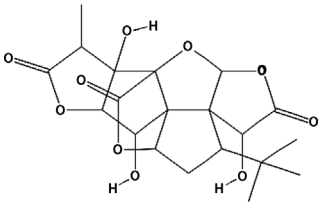
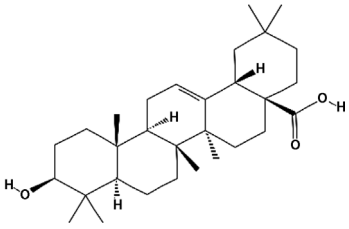
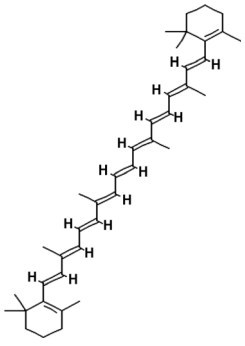
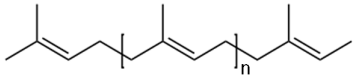
Table 1.1 The major classes of secondary metabolites based on their chemical structures. Chemical structures are from The PubChem Project.

Class of compound	Example	Structure
N-containing		
Alkaloids	Nicotine	
Nonproteinogenic amino acid	Canavanine	
Amines	Butylamine	
Cyanogenic glycosides	Dhurrin	
N- and S-containing		
Glucosinolates	Glucobrassicin	
Without N		
Anthraquinones	Emodin	

Flavonoids	Quercetin	
Polyacetylenes	Falcarinol	
Polyketides	Aloeresin	
Phenylpropanoids	Rosmarinic acid	
Terpenes		
Mono-	Thymol	
Sesqui-	Helenaline	

(*continued*)

Table 1.1 (*Continued*)

Class of compound	Example	Structure
Di-	Gingkolide	
Tri-	Oleanolic acid	
Tetra-	ß-Carotin	
Poly-	Rubber	

Some metabolites occur only in specialized subcellular compartments. In such cases, the tissue- and cell-specific localizations will depend on the solubility of the compound, notably whether it is lipophilic or hydrophilic (Table 1.2). Most hydrophilic compounds accumulate in vacuoles, and are typically present as inactive precursor substances that must be activated either chemically and/or enzymatically; this is the case for cyanogenic glycosides and glucosinolates. Antimicrobial tannins, which exert their protective effects in the extracellular space, are located in the apoplast, while the laticifers contain both hydrophilic and lipophilic compounds. The metabolite patterns are species-specific, but all are toxic; if human skin comes into contact with a latex extract, a severe irritation results. Lipophilic storage compartments also contain antimicrobial and/or defense compounds. Metabolites located

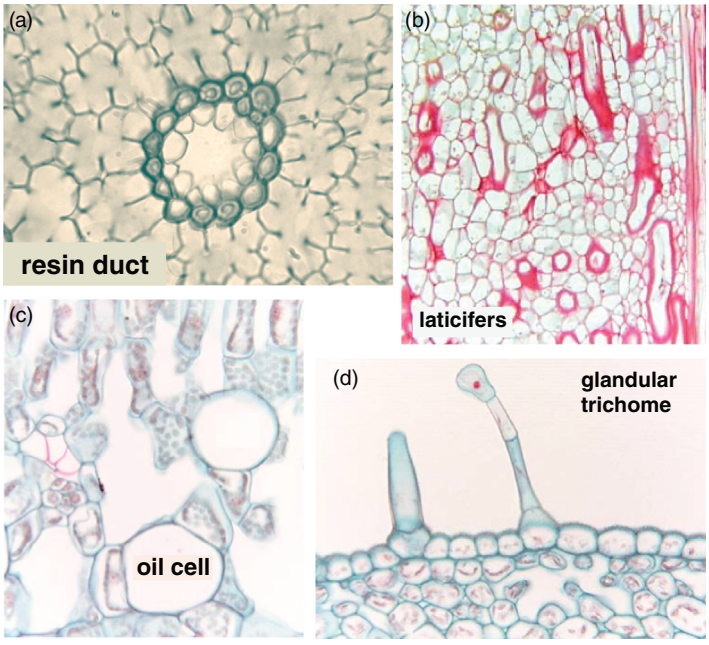


Figure 1.2 Some examples of secondary metabolite-containing plant compartments or cells. (a) Resin duct; cross-section through the needle of a *Pinus* species; (b) Laticifers; longitudinal section through the shoot of an *Euphorbia* species; (c) Oil cells; cross-section through a magnolia (*Magnolia* sp.) leaf; (d) Trichome; section through a foxglove (*Digitalis* sp.) leaf with glandular trichome. Image (a) from Bernd Schulz and Jutta Ludwig-Müller, Technische Universität Dresden, Germany; images (b–d) from James D. Mauseth, University of Texas, USA.

in the membranes of the plastids and mitochondria are necessary for primary metabolism, an example being electron-transport chains and light absorbance. Plants secrete volatile compounds either to attract insects for pollination purposes, or zoophagous insects to defend themselves against herbivorous insects (Figure 1.1; see also Chapter 2). Plants may also deliver nonvolatile compounds into the soil, thereby altering the rhizosphere (termed allelopathy; see Section 2.6) that leads to an inhibition of the growth of other, competing, plants.

The process of compartmentation has three important implications:

1) The plant itself may not be in contact with the toxic compounds, because it is only after wounding or attack by other organisms that the metabolites will be released from the storage compartments.
2) The biosynthetic pathways are also compartmentalized, but this necessitates the involvement of transport processes for metabolite movement.
3) The target molecules for the secondary metabolites may be present in all compartments; however, the target is not necessarily colocalized with an active compound.

Table 1.2 Tissue-specific and subcellular compartmentation of hydrophilic and lipophilic compounds.

Characteristics	Storage compartment		Class of compounds
	Tissue	Cell	
Hydrophilic	Laticifer		Some alkaloids
			Nonproteinogenic amino acids
			Digitalis glycosides
			Cyanogenic glycosides
		Vacuole	Many alkaloids
			Nonproteinogenic amino acids
			Saponins
			Glycosides
			Flavonoids
			Anthocyanidins
			Betalains
			Tannins
			Cyanogenic glycosides
			Glucosinolates
			Amines
		Apoplast	Tannins
		Cytoplasm	Most hydrophilic compounds during their biosynthesis
		Nucleus	Flavonoids
Lipophilic	Trichome		Terpenoids
	Resin duct		Flavonoids
			Terpenoids
	Laticifer		Diterpenes
			Quinones
			Flavonoids
			Polyterpenes
	Oil cell		Anthraquinones
			Terpenoids
		Cuticula	Wax
			Flavonoids
			Terpenoids
		Mitochondrion	Alkaloids (*Conium*)
		Plastid	Terpenes
			Alkaloids (coniin, quinolizidine, coffein)
		Specialized vesicles	Alkaloids (protoberberin)
		Plastid membrane	Ubiquinones
Tetraterpenes			
		Endoplasmic reticulum	Lipophilic substances during enzymatic hydroxylation steps

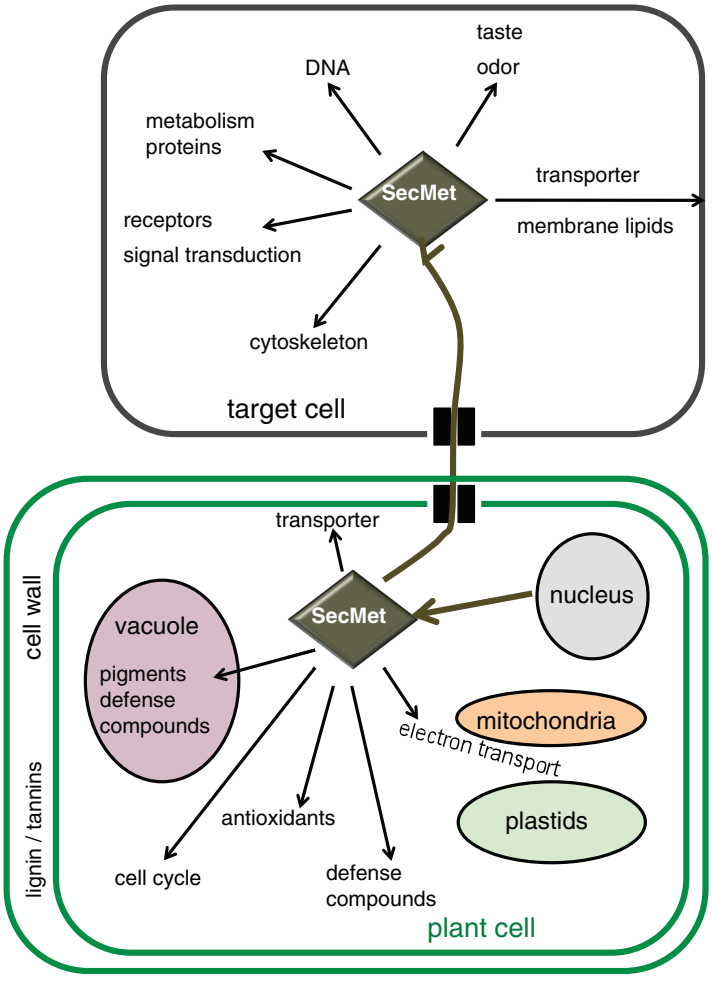


Figure 1.3 Cellular and molecular targets for secondary metabolites in plants and other organisms. In the plant cell (green), secondary metabolites (SecMet) are synthesized and can exert their effects in different compartments, such as the electron-transport chain in organelles, transporters at the plasma membrane, as antioxidants, and as fortification elements in the cell walls. They are stored in vacuoles or cell wall compartments as coloring or defense compounds. In the cells of the target organisms (brown), into which the metabolites must be transported, the targets can be in all compartments. Target molecules encompass transporters and channels, enzymes and proteins, nucleic acids, the cytoskeleton, receptor molecules and signal transduction chains, as well as membrane lipids.

A few cellular targets will be discussed in the following section (see Figure 1.3), and further descriptions are provided in Chapters 5 and 6, with details of the compounds' biological activities. Many compounds that are used to protect plants against microbes and herbivores exert their biological activities outside the plant

tissues; however, if these compounds were to exert any toxic effect on the plant, their compartmentation would provide protection against self-poisoning.

Membrane targets can be found at the plasma membrane, ER or Golgi apparatus, where the functions of molecules such as transporter and channels, and also of membrane receptors, can be altered. The electron-transport chains of both plastids and mitochondria may also serve as targets, and plant compounds may inhibit their function. In addition, the structural proteins of the cell wall and of enzymes may be affected, while the cytoskeleton, with its component proteins myosin and tubulin, may be a target for secondary metabolites and result in an altered cell cycle (see Figure 2.5). Within the nucleus and the organelles, desoxyribonucleic acid (DNA) is a target for interaction with secondary metabolites; it is well known that some molecules alkylate DNA, which in turn affects its replication, transcription, and repair mechanisms. It should be noted that, in order to be active in these compartments, the metabolites must be transported across the respective membranes (see Section 1.1.4).

Although the major role of these compounds seems to be the defense of plants against other organisms, humans are also affected by secondary metabolites (see Chapters 3.2 and 6). When defending plants against microbes, the toxic effect of the secondary metabolites on one or several target molecules constitutes a major role, although their "bitter-tasting" properties will generally deter attacks by herbivores. In these organisms, the targets may be located as indicated in Figure 1.3, with the exception of plastids, which only play a role as target in allelopathy. This might not be the case for parasitic plants, however, which are dependent on the host photosynthesis for their nutrient acquisition. Many secondary metabolites, such as sterols, can influence membrane fluidity (see Figure 2.28), while other compounds may interact with or intercalate in DNA, and are thus cytotoxic; mitosis may also be inhibited (see Figure 2.5).

Secondary metabolites can also inhibit the activity of various enzymes. Alkaloids and amines are often similar to the neurotransmitters of vertebrates, or to hormones involved in the development of insects, and thus mimic their effects. Cyanogenic glycosides release toxic cyanide, which inhibits cytochrome c oxidase in the mitochondrion and, as a consequence, energy production in the form of adenosine triphosphate (ATP). Isothiocyanates, which are released from glucosinolates, can influence membrane fluidity. Terpenes also interact with membranes, though some demonstrate similarities to biologically important sterols in fungi and animals. For example, the Na^+-, K^+-ATPase, which is important in animals, can be inhibited by several groups of metabolites, including the pyrrolizidine alkaloid monocrotalin, the digitalis glycosides, and/or polyketides such as anthraquinones. The latter also interfere with the formation of cyclic AMP (cAMP) by inhibiting adenylate cyclase, which itself influences many signal transduction pathways. Interestingly, plants contain cyclic guanosine monophosphate (cGMP) rather than cAMP. Flavonoids also have the ability to inhibit enzymes, but may cause additional mutagenic and toxic effects on DNA via alkylation.

1.1.2 Biosynthesis

The biosynthetic pathways derive from various precursors of primary metabolism (Figure 1.4). The precursor is defined as a molecule used by a biosynthetic enzyme as a substrate and converted to a product. The product can be an intermediate in the pathway, and in this case it is used as precursor for the next biosynthetic enzyme, or it is the final product of the reaction chain.

substrate (precursor) → intermediate → intermediate → product
⇕
precursor → intermediate → product

In a complex reaction scheme, which has many junctions, an intermediate is simultaneously also a precursor for another part of the pathway. In Figure 1.4, shikimic acid would be an intermediate for the amino acid metabolism, and also

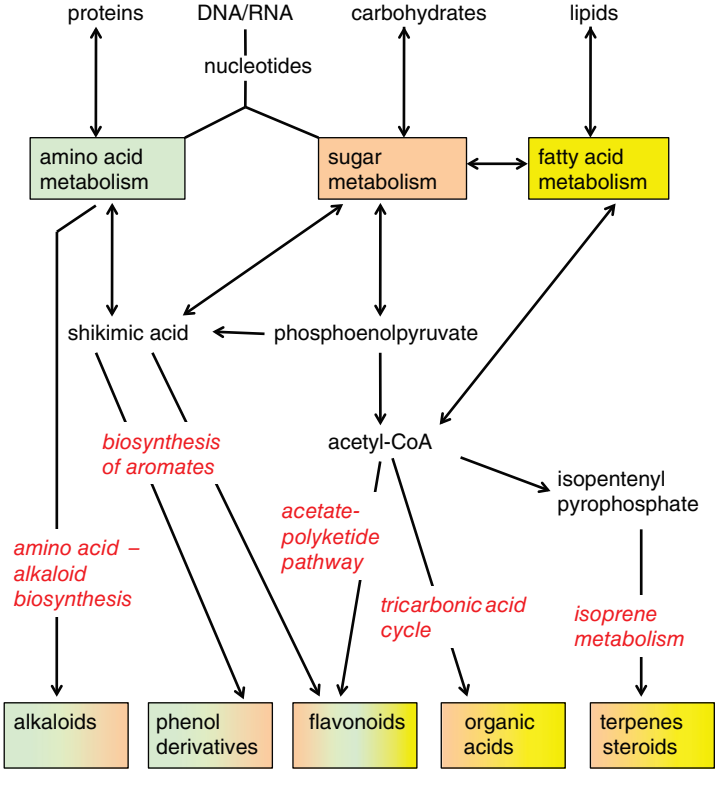


Figure 1.4 General scheme of biosynthetic pathways and precursors for the major classes of secondary metabolites. Pathways deriving from aromatic amino acids are shown in green; pathways deriving from carbohydrates are shown in brown; pathways deriving from lipids are shown in yellow. Modified from Larcher, W. (1995) *Physiological Plant Ecology*, 3rd edn, Springer-Verlag.

a precursor for the biosynthesis of aromatic secondary metabolites. Similar precursors are used within one class of compounds for the biosynthesis, but the same precursors can be also used for a range of different metabolites.

The final products are derived from three major classes of compounds and marked in the same color in Figure 1.4. The major precursors are derived from protein (amino acids), carbohydrate (sugars) and lipid (fatty acid) metabolism. The biosynthetic pathway for aromatic amino acids is one of the major sources of aromatic compounds such as phenols, flavonoids and some alkaloids. Acetyl-CoA is a central metabolite formed by glycolysis and also via the β-oxidation of fatty acids, and is used in the tricarboxylic acid cycle in the synthesis of organic acids, which are also precursors for secondary metabolites. In addition, acetyl-CoA is involved in the synthesis of terpenes, which form a distinct class of metabolites. In the following sections some selected biosynthetic pathways will be discussed, but for more detail the reader is referred to textbooks on plant biochemistry. Knowledge of the biosynthetic pathways of natural compounds is essential for the targeted manipulation of these pathways in biotechnology. In addition to knowledge on the enzymes and genes involved in these biosyntheses, it is important to understand the transcriptional control of the genes. Therefore, a few examples will be provided where the transcriptional control of pathways has been well elucidated, notably of the glucosinolates and flavonoids (Section 1.1.3). Finally, details on biosynthetic pathways and corresponding genes will help to provide an understanding of how these secondary metabolites have evolved (Section 1.2).

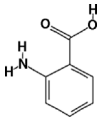
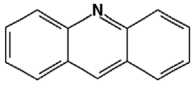
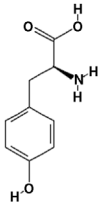
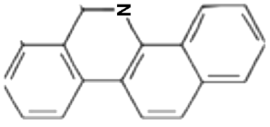
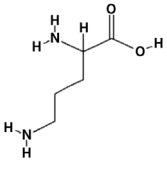
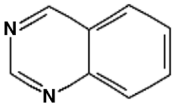
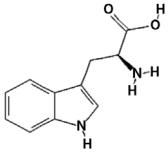
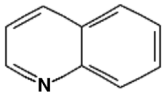
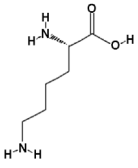
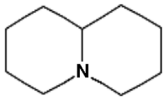
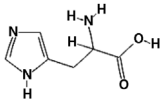
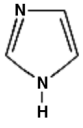
When the basic skeleton of a secondary metabolite has been synthesized, it is the many enzymatic modifications (which are added later to the structure) that form the main reason for the large variations among these molecules. The chemical reactions involved include oxidations, hydroxylations, reductions, acylations, methylations, prenylations and glycosylations (see also Table 2.5).

1.1.2.1 **Alkaloids**

Within the group of alkaloids many compounds with stimulating, hallucinogenic and analgesic properties can be found (see Chapter 5). Many of these compounds are toxic, or they can be converted from nontoxic to toxic compounds. Depending on the chemical structure of an alkaloid, different precursors are required for the biosynthesis. Examples of the major groups of alkaloids, together with a typical chemical structure, a plant in which the compound is found and the main precursor(s) involved in the biosynthesis, are shown in Table 1.3.

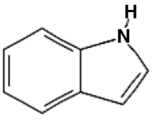
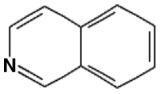
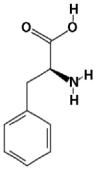
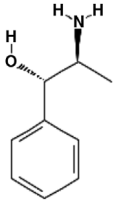
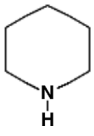
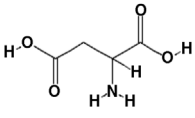
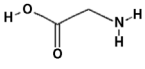
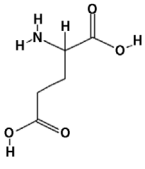
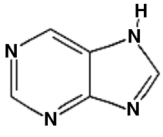
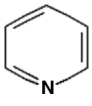
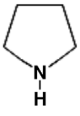
The biosynthetic pathway for an individual compound is very complex. Many different precursors are involved for the various pathways, including not only aromatic amino acids such as tryptophan, tyrosine and phenylalanine but also aspartate, glutamine, lysine, glycine and valine for other alkaloids (Figures 1.5 and 1.6; Table 1.3). In addition, the nonproteinogenic amino acid ornithine is an important precursor for various alkaloids. Anthranilic acid, from which tryptophan is synthesized, is the precursor for acridine alkaloids, and thus the shikimate pathway (which is responsible for the biosynthesis of all aromatic amino acids) is involved. For several alkaloids two different precursors are needed for the biosynthetic

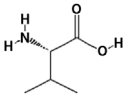
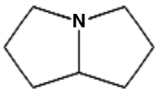
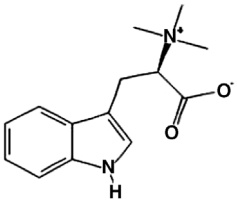
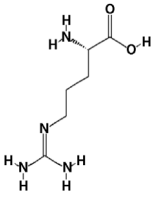
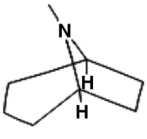
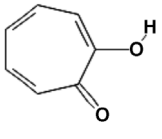
Table 1.3 The major classes of alkaloids and their precursors, according to their chemical structure. One example, with chemical structure, and one plant species where it occurs, is given for each class. Chemical structures are from The PubChem Project.

Basic structure	Precursor	Example of structure (Plant)	Formula basic structure
Acridine	Anthranilate 	OH-N-Methylacridone (*Ruta graveolens*)	
Benzophenanthrinidine	Tyrosine 	Sanguinarine (*Eschscholzia californica*)	
Quinazoline	Ornithine 	Peganine (*Peganum harmala*)	
Quinoline	Anthranilate Tryptophan 	Quinine (*Cinchona officinalis*)	
Quinolizidine	Lysine 	Lupanine (*Lupinus polyphyllus*)	
Imidazole	Histidine 	Pilocarpine (*Pilocarpus jaborandi*)	

(*continued*)

Table 1.3 (*Continued*)

Basic structure	Precursor	Example of structure (Plant)	Formula basic structure
Indole	Tryptophan	Serotonin (*Musa x paradisiaca*)	
Isoquinoline	Tyrosine	Morphine (*Papaver somniferum*)	
Phenylalkylamine	Phenylalanine 	Cathinone (*Catha edulis*)	
Piperidine	Lysine	Coniine (*Conium maculatum*)	
Purine	Aspartate  Glycine  Glutamate 	Caffeine (*Coffea arabica*)	
Pyridine	Aspartate Ornithine	Nicotine (*Nicotiana tabacum*)	
Pyrrolidine	Aspartate Ornithine	Muscarine (*Amanita muscaria*)	

Pyrrolizidine	Ornithine Valine 	Senecionine (*Senecio jacobaea*)	
Terpene indole	Tryptophan Isoprene 	Physostigmine (*Physostigma venenosum*)	
Tropane	Ornithine Arginine 	Cocaine (*Erythroxylon coca*)	
Tropolone	Tyrosine Phenylalanine	Colchicine (*Colchicum autumnale*)	

pathways. In the case of terpene indole alkaloids (Figures 1.5 and 1.6; see also Figure 1.34), it is not only tryptophan that is involved as a precursor for the indole moiety, but also monoterpenes for the synthesis of side chains. Another example is the biosynthesis of the tropane alkaloids hyoscyamine and scopolamine, where ornithine and phenylalanine are required for the different parts of the molecule (Figure 1.5; Table 1.3).

Although many alkaloids are of importance to humans (Chapter 5), their biosynthetic pathways are often not fully understood. However, it is important to know as much as possible about the precursors, enzymes and limiting steps in the pathway, if these secondary metabolites should be changed in abundance by breeding or biotechnological methods employed in plants (Section 1.3). It is important that changes in one pathway do not lead to limitation of a precursor for a second pathway, in case the two pathways compete for the same precursor. This can be observed especially for aromatic amino acids, which play many roles in different pathways in the plant; examples include their requirement in proteins, a role in the synthesis of cell wall components, and as precursors for the plant hormone indole-3-acetic acid (IAA). In secondary metabolite synthesis, amino acids

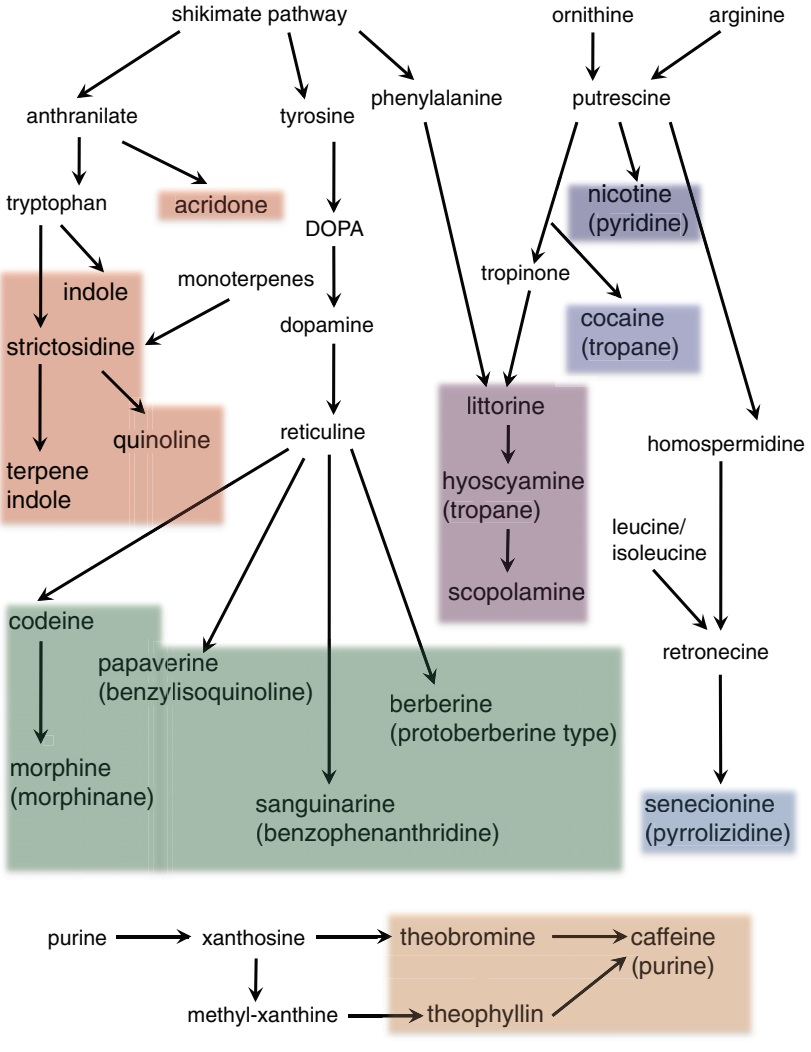


Figure 1.5 Overview on the biosynthesis of selected alkaloids. The shikimate pathway leads to the synthesis of many different alkaloids given in colors depending on their origin (red = anthranilate/tryptophan, green = tyrosine). In addition, ornithine and arginine via putrescine result in the synthesis of pyridine (dark blue), tropane (blue) and pyrrolizidine alkaloids (light blue). Phenylalanine together with ornithine is needed for the synthesis of a second group of tropane alkaloids (violet). Caffeine and related substances is derived from purine (brown). The class of compounds is given in brackets.

are involved in the synthesis of alkaloids, of phenolic compounds, and also of pigments such as anthocyanins and betalains. This indicates that, even under natural conditions, alkaloid biosynthesis competes with many other pathways, because the other amino acids are also needed for protein synthesis.

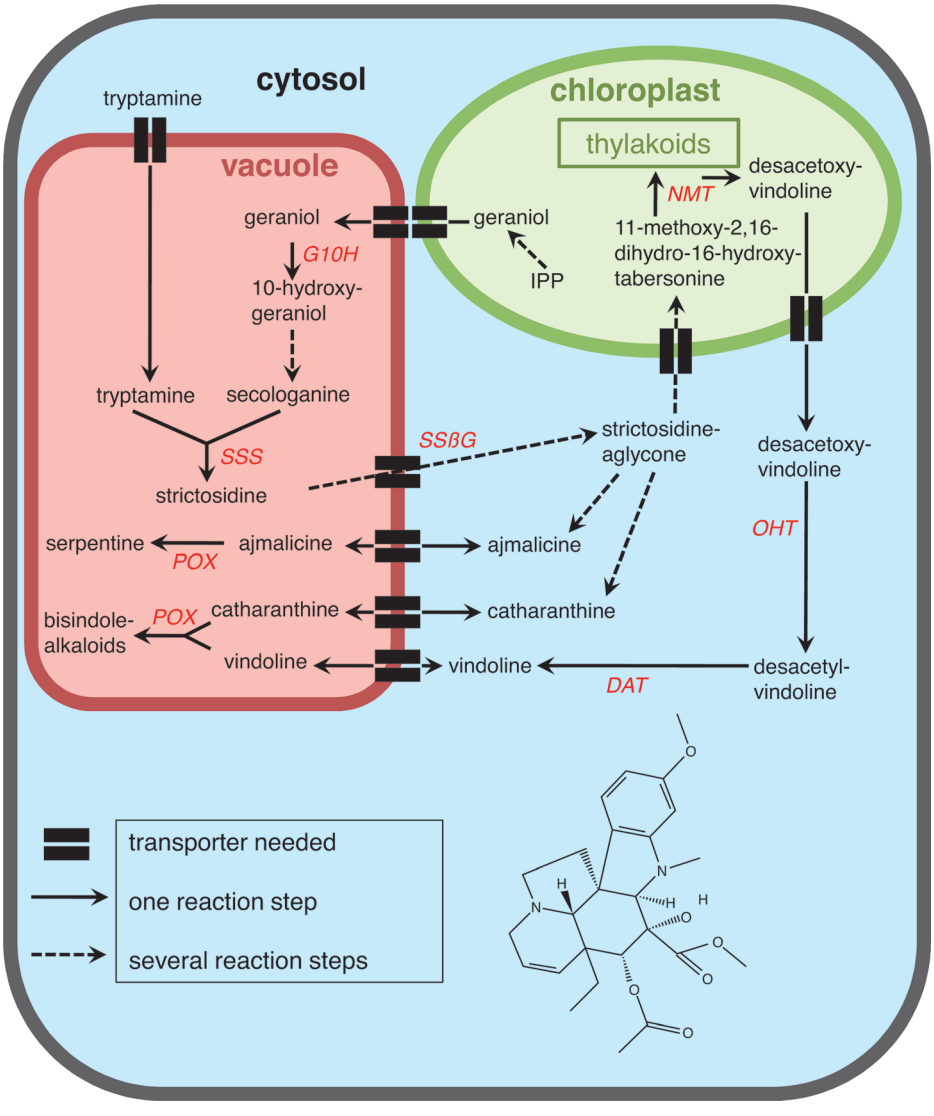


Figure 1.6 Compartmentation of terpene indole alkaloid biosynthesis in *Catharanthus roseus*. The cellular compartments (plastids, cytosol and vacuole) participate in the biosynthetic steps. Enzymatic reactions with dashed arrows are hypothetical. Circles symbolize membrane-associated enzymes. Putative transporters are also indicated. The structure for the terpene indole alkaloid vindoline is shown. IPP: isopentenylpyrophosphate; G10H: geraniol-10-hydroxylase; NMT: *S*-adenosyl-methionine hydroxytabersonine-*N*-methyltransferase; DAT: acetyl-coenzyme-A deacetylvindoline 17-*O*-acetyltransferase; OHT: 2-oxoglutarate-dependent dioxygenase; POX: peroxidase; SSβG: strictosidine-β-glycosidase; SSS: strictosidine synthase. Modified from Verpoorte, R. (2000) *Metabolic engineering of plant secondary metabolism*. Kluwer Academic Publisher, pp. 1–29. Chemical structure is from The PubChem Project.

The shikimate pathway contributes to the synthesis of acridone and indole alkaloids via anthranilic acid or tryptophan. Tyrosine is a precursor for a series of compounds, namely the protoberberine type, benzylisoquinoline, morphinane, and benzophenanthridine alkaloids; the intermediate is reticuline (Figure 1.5). Pyridine (e.g., nicotine) and tropane alkaloids (e.g., cocaine) are derived from ornithine, where the polyamine putrescine is a common precursor; however, the pathways diverge before the intermediate tropolone is formed. For a second class of tropane alkaloids, the amino acid phenylalanine is required in addition to ornithine for synthesis of the backbone of the molecule littorine, which is an intermediate on the route to hyoscyamine and scopolamine. Putrescine is converted to homospermidine, which is then metabolized to retronecine, the precursor for the group of pyrrolizidine alkaloids (Figure 1.5). The reaction converting homospermidine is catalyzed by homospermidine synthase, the first specific enzyme for this pathway. For an additional part of the molecule – the complex ester structure of mono- or dibasic acids – either leucine or isoleucine are mainly involved as precursors. Pyrrolizidine alkaloids are synthesized mainly in the roots and then transported to the upper parts of the plant. The problem here for humans and animals is the hepatotoxic potential of these compounds.

Although quinolizidine alkaloids are structurally very similar to pyrrolizidine alkaloids, they are synthesized from the proteinogenic amino acid lysine. An important intermediate in this pathway is cadaverine (another polyamine), the biosynthesis of which occurs in the chloroplasts of mesophyll cells. A third group with structural similarity to the above-mentioned compounds are the indolizidine alkaloids; these are found in the orchid genus *Dendrobium*, but also occur in the South American poison dart frog of the genus *Dendrobates*. Even though the synthesis of the two major compounds swainsonine and castanospermine derives from the amino acid lysine, a different intermediate is found, namely pipecolic acid. The second part of the ring structure is derived from acetate.

In the way that the same precursor is used for a range of metabolic pathways, intermediates can also serve in the synthesis of a variety of compounds (Figure 1.5). The intermediate reticuline serves as a precursor for different groups of alkaloids, such as isoquinoline-, benzophenanthridine- or protoberberine-type compounds; examples are morphines, sangurinarine, and berberine. Within these pathways there are alternatives to synthesize the same metabolite; for example, nicotine can be produced from either ornithine alone or by a fusion of ornithine with putrescine. Morphine (for structure, see Figure 5.30) can either be synthesized via thebaine, oripavin and morphinone (the latter is the direct precursor for morphine), or by the conversion of thebaine to neopinone, which is then converted via codeinone to the bioactive codeine, a known remedy against cough (see Section 4.1); ultimately, codeine is converted to morphine.

Amphetamine-like compounds such as cathinone from *Catha edulis* are also derived from an aromatic amino acid (phenylalanine in this case). The mild stimulating activity of the plant can be achieved by chewing leaves, whereby cathinone is absorbed via the mucosa of the mouth. The genus *Ephedra* produces ephedrine and pseudoephedrine, which are used in medicine to reduce nasal congestions

(Section 5.3.7). The slightly stimulating activity is, again, due to the similarity with amphetamine. The biosynthesis might proceed via a route similar to phenylpropanoid synthesis, but with benzoic acid or benzoyl-CoA as intermediates.

Simple indole alkaloids such as serotonin most likely derive from tryptophan via tryptamine as intermediate. The complex structure of terpene indole alkaloids is also dependent on the terpene pathway for the side-chain moiety. The biosynthesis of terpene indole alkaloids is also an example of a highly compartmented pathway (Figure 1.6). On the one hand, the localization of enzymes is critical, but on the other hand the solubility of the compounds plays a role and these factors necessitate the compartmentalization of the compounds (Figure 1.3; Table 1.2), and therefore transporters are essential for the respective intermediates. Tryptophan is synthesized in the plastids, whereas tryptamine is formed in the cytosol. The latter must be transported between the two compartments, and the same is true for the monoterpene moiety, which is again synthesized in the plastids (see Figure 1.14). Tryptamine is further converted in the cytosol, whereas the first step to the terpene indole alkaloid takes place in the vacuole. Although further enzymatic modification steps are observed again in the cytosol, the later steps also occur in the vacuole (Figure 1.6).

1.1.2.2 **Phenol Derivatives, Especially Flavonoids**

As a representative group for bioactive phenols, the biosynthesis of flavonoids is described in more detail. The biosynthetic pathways derive from the shikimate pathway (Figure 1.7), which is shared by indoles, and by several alkaloids and betalains. The precursor for the cinnamic acid derivatives and flavonoids is phenylalanine, which is converted by the key enzyme for this part of the pathway, phenylalanine ammonia lyase (PAL), to cinnamic acid. Simple molecules (e.g., cinnamic acids) as well as very complicated molecules (e.g., rosmarinic acid) are synthesized via this pathway. Rosmarinic acid has a high antioxidative potential and also good aromatic qualities. Coumarin occurs in woodruff as an aroma and an intoxicating (poisonous) substance. The cinnamic acid derivatives also serve as precursors for polymers such as lignin, which is synthesized via cinnamaldehydes and the respective alcohols, termed monolignols. Flavonoids and stilbenes, such as resveratrol from grapes, are also formed via this pathway. Activation by coenzyme-A plays an important role in this biosynthesis network; the cinnamic acid is activated and acts then, as cinnamoyl-CoA, as a precursor for other pathways in the synthesis of flavonoids and coumarins.

Flavonoid biosynthesis has been quite well elucidated in the model plant *Arabidopsis thaliana*, in which only a small set of secondary metabolites are synthesized. Much information is also available from maize and legumes, the latter also contain isoflavonoids. The scheme for the biosynthetic steps leading to the major flavonoid compounds includes information derived from several plant species (Figure 1.8). The key enzyme for the flavonoid pathway is chalcone synthase, which catalyzes the reaction to naringenin chalcone, the first flavonoid in the pathway. Chalcone synthase mutants of *Arabidopsis* have a transparent seed coat (testa), and are therefore referred to as *transparent testa* (*tt*) mutants. However, other mutants in the pathway of, for example, the next enzyme encoding chalcone

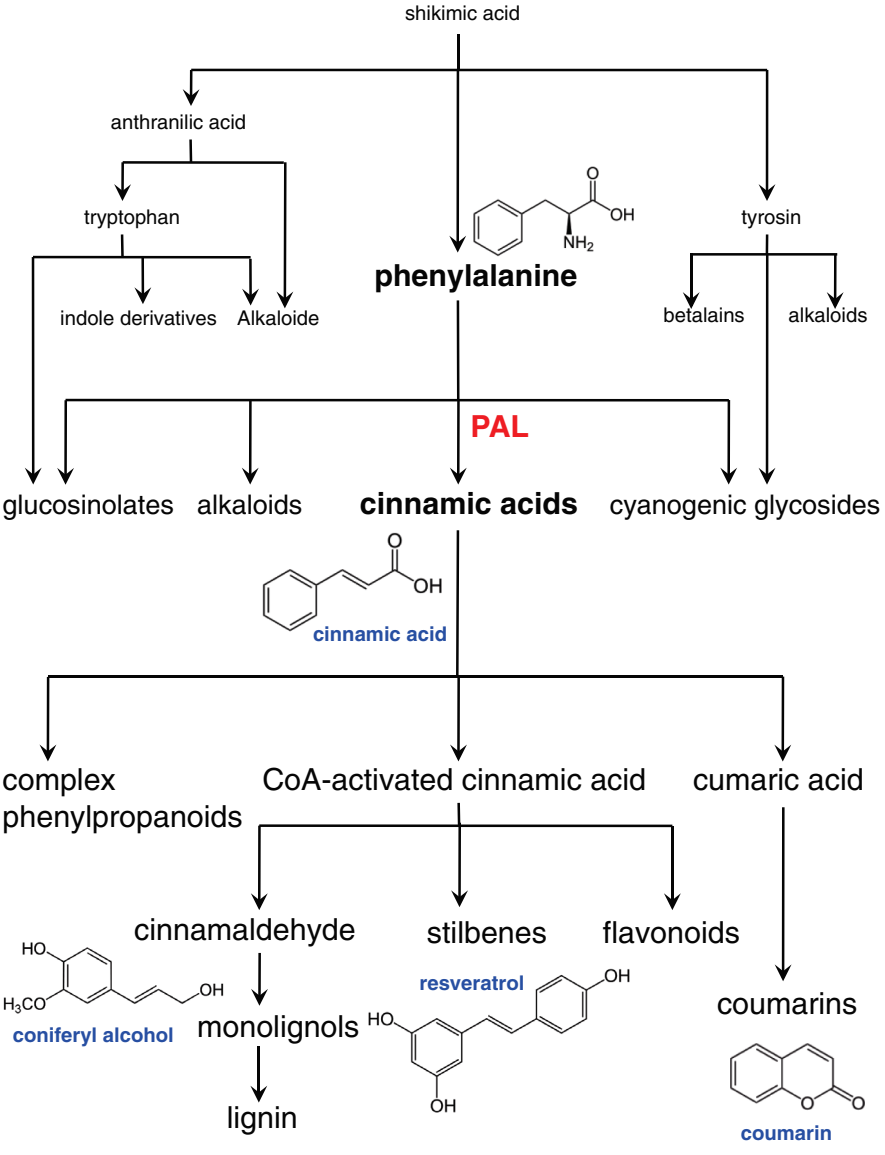


Figure 1.7 Schematic overview for the biosynthetic pathways of selected phenols from phenylalanine as precursor (bold). A key enzyme, phenylalanine ammonia lyase (PAL), is shown (red). Some example structures are depicted (blue). Chemical structures are from The PubChem Project.

isomerase (which is responsible for the synthesis of naringenin), also show this phenotype, and consequently the mutations were numbered consecutively, starting with "1." Mutations in the transcription factors (see Section 1.1.3) that control the synthesis of flavonoids have similar phenotypes.

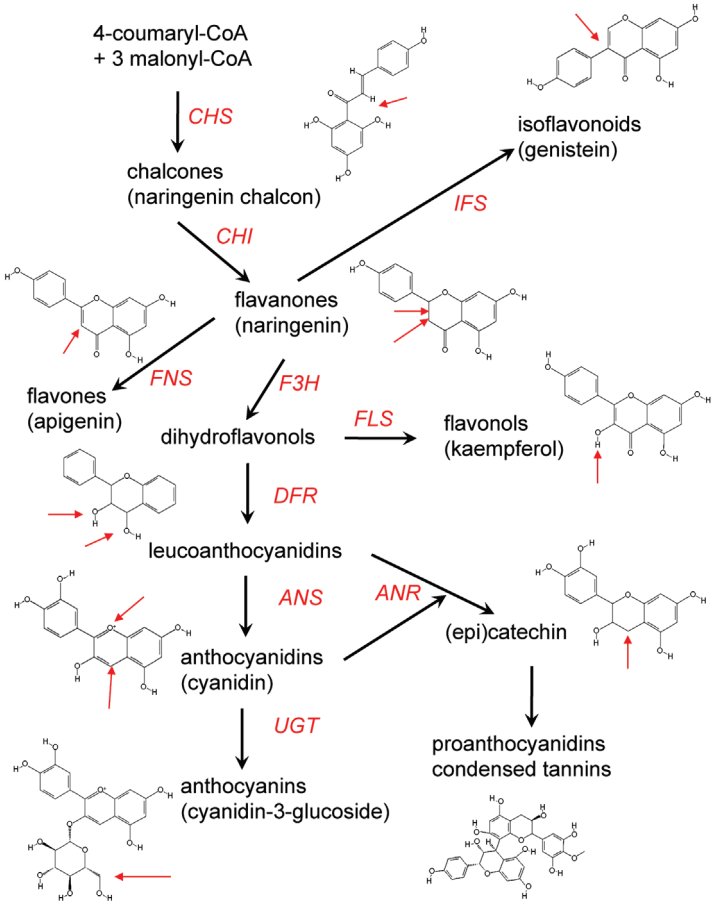


Figure 1.8 The main pathways for flavonoid synthesis derived from different plant species. Because of the complexity of the different biosynthetic steps, only the basic features of the pathways are illustrated. The names in brackets are examples for each class of compounds, which is then also presented in a chemical structure. The red arrows indicate changes in the individual flavonoid groups compared to naringenin. CHS: chalcone synthase; CHI: chalcone isomerase; IFS: isoflavonoid synthase; FNS: flavone synthase; F3H: flavanone-3-hydroxylase; FLS: flavonol synthase; DFR: dihydroflavonol reductase; ANS: anthocyanidin synthase; UGT: glycosyltransferase; ANR: anthocyanidin reductase. Chemical structures are from The PubChem Project.

Several biosynthetic pathways originate from naringenin, which lead to (i) the isoflavonoids of legumes, (ii) flavones, and (iii) flavonols, which constitute the largest group among the flavonoids (Table 1.4). The flavonols in the model organism for molecular biology, *Arabidopsis thaliana*, are kaemfperol and quercetin, and their metabolites (glycosides). Dihydroflavonols are intermediates for the formation of flavonols, but are also converted to the anthocyanin pigments. In addition, they give rise to the proanthocyanidins (also known as "condensed tannins"),

Table 1.4 The major classes of flavonoids with the estimated number of compounds and some functions (see also Chapter 2).

Class	No. of known compounds	Biological function in plants
Anthocyanins	250	Red and blue pigments[a)]
Aurones	20	Yellow pigments
Biflavonoids	65	Not known yet
Catechins	40	Antimicrobial
Chalcones	60	Yellow pigments
Dihydrochalcones	10	Feeding deterrents
Flavones	350	Light yellow pigments
Flavonols	350	Feeding deterrent
Isoflavonoids	15	Antimicrobial, attraction
Proanthocyanidins	50	Antimicrobial

a) Pigmentation can have the function in flowers of pollination, but also in fruit of seed dispersal or the deterrence of herbivores in combination with bitter-tasting compounds. In leaves, the pigmentation can serve as a protection against high levels of light.

which are found in the seed coats of many plants. The catechins are derived from intermediates of the anthocyanin pathway, the (leuco)anthocyanidins. The highly diverse structures are achieved by glycosylation reactions with various sugars, where mostly monoglycosides or diglycosides are formed (see Figure 2.13). The sugar moiety is added either as a monosaccharide or disaccharide to one position, though the sugars can be also added at different positions of the molecule. Glycosylations also render the molecules more hydrophilic, and this may change the storage compartment involved. Further details on the water-soluble anthocyan(in)s are provided in Chapter 2.

1.1.2.3 **Betalains**

Betalains fulfill the function of anthocyanins in the Caryophyllales as pigments and UV-protecting compounds. They have never been detected in combination with anthocyanins in any plant, and the reason for that remains a mystery. In addition to plants, betalains occur in some mushrooms, such as fly agaric (*Amanita muscaria*). Betalain-containing plant species cannot convert flavane-3-dioles to the respective anthocyanin, whereas flavonoids up to the step of leucoanthocyanidins occur in betalain-containing plants.

The group of N-containing betalains are derived from tyrosine, though an alternative pathway based on tyramine has been reported for some plant species. In contrast to anthocyanins and carotenoids, the biosynthetic pathway of betalains is only partially understood. Like anthocyanins, they also occur in a glycosylated form, but the color does not depend on the pH. Betalains are comprised of two major groups: the yellow betaxanthins and the red-pink betacyanins (Figure 1.9). The more than 50 betacyanins known to date are water-soluble and stored in vacuoles. The most important intermediates are DOPA (L-3,4-dihydroxyphenylalanine; see Figure 1.9) and betalamic acid. Next to betanidin, amararanthin is a prominent compound in

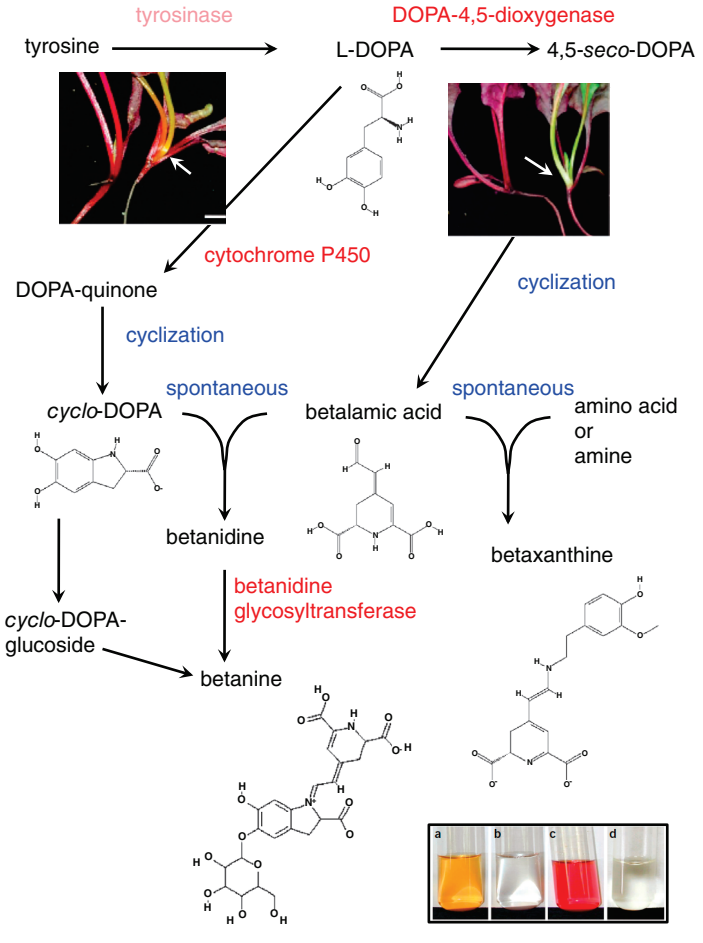


Figure 1.9 Betalain biosynthesis from the aromatic amino acid tyrosine. Enzymatic steps are shown in red, spontaneous reactions in blue. The inferred enzyme tyrosinase is shown in light red. DOPA = (2*S*)-2-amino-3-(3,4-dihydroxyphenyl)propanoic acid; DODA = DOPA 4,5-dioxygenase. Insert: Expression of the cytochrome P450 gene and the *DODA* gene in yeast (a. *DODA* expression + DOPA as substrate; b. *DODA* expression – DOPA; c. both genes + DOPA as substrate; d. both genes – DOPA). Photographs from Hatlestad, G.J. *et al.* (2012) *Nat. Genet.*, **44**, 81. Reprinted with permission from Nature Publishing Group. Chemical structures are from The PubChem Project.

Amaranthus species, while betalamic acid serves as intermediate for the betaxanthins and betanidins. The reaction is nonenzymatic in both cases; rather, a spontaneous addition occurs of either cyclo-DOPA to yield betanidin, or of an amino acid or amine, which results in the formation of a betaxanthin. The glycosylated betanin is formed by a glycosyltransferase reaction with either cyclo-DOPA or betanidin. The enzyme tyrosinase is known from the synthesis of melanin in animals, and in this

case two reactions are being catalyzed by the one enzyme; but this enzyme has not been characterized in plants. In contrast, evidence has been advanced that for these two steps a different enzyme is required.

A novel gene involved in the biosynthesis of betacyanins in *Beta vulgaris* has been isolated with the help of yellow mutants encoding a novel cytochrome P450 protein that is involved in the conversion of DOPA to DOPA-quinone, and most likely again by the spontaneous cyclization to cyclo-DOPA. The biosynthesis of betalamic acid is dependent on a DOPA 4,5-dioxygenase (DODA), which results in the formation of both betacyanins and betaxanthins. Functional proof of this was derived from the silencing of red wild-type *B. vulgaris* plants using both genes alternatively. Whereas loss of the cytochrome P450 gene resulted in yellow-pigmented plants, loss of the *DODA* gene yielded white (unpigmented) plants; this showed that in beet, betalamic acid is indeed the precursor for both the yellow and red pigments (Figure 1.9). Complementation of the yellow beet mutant with the cytochrome P450 gene led to red tissues. The expression of these genes in yeast, with subsequent addition of the substrate DOPA, resulted in a red color in the case of both enzymes, in a yellow color when only the *DODA* gene was expressed, and in no color when the DOPA substrate was omitted. The results of this experiment indicated that both genes are essential for the red betacyanin formation, whereas the *DODA* gene alone is sufficient to yield the yellow betaxanthins.

1.1.2.4 Cyanogenic Glycosides, Glucosinolates, and other Indole Derivatives

While alkaloids, phenols, flavonoids and terpenes are present in most plant families – albeit with their plant-specific patterns – glucosinolates occur mainly within the crucifer family (Brassicaceae). This family includes many crop plants, vegetables and spices are found, such as cabbages, radish, mustard, or oilseed rape (canola) (Figure 1.10); the model plant *A. thaliana* is also a member of the Brassicaceae. Although glucosinolates and cyanogenic glycosides share similar intermediates in their biosynthetic pathways, it has been hypothesized that both have evolved separately (Section 1.2), and both classes of metabolites have been identified only in a few plant species, among them in papaya (*Carica papaya*).

The cyanogenic glycosides and glucosinolates are both present as glycosides in the vacuoles; however, they are inactive and must first be metabolized to form the active compounds, either via an enzymatic reaction or spontaneously. This process occurs only following damage of the tissue due to colonization by microbes, or by the feeding of herbivores, and consequently the glycosides and glucosinolates are considered as preformed defense compounds (see Section 2.5). Although the glucosinolates are localized in the vacuole, the degrading enzyme – β-thioglycosidase (myrosinase) – is compartmentalized in the cytosol. Following tissue disruption, however, both components of the system come together, whereupon the glucosinolates are metabolized to their toxic active components that are referred to as "mustard oils"; hence, the whole system is known as "The Mustard Oil Bomb." As these compounds might also be toxic towards the plant cell, compartmentation is essential in order to maintain healthy tissues. The application of immunohistochemical techniques has shown that myrosinases also occur in the vacuoles, albeit in specialized "myrosin cells," named

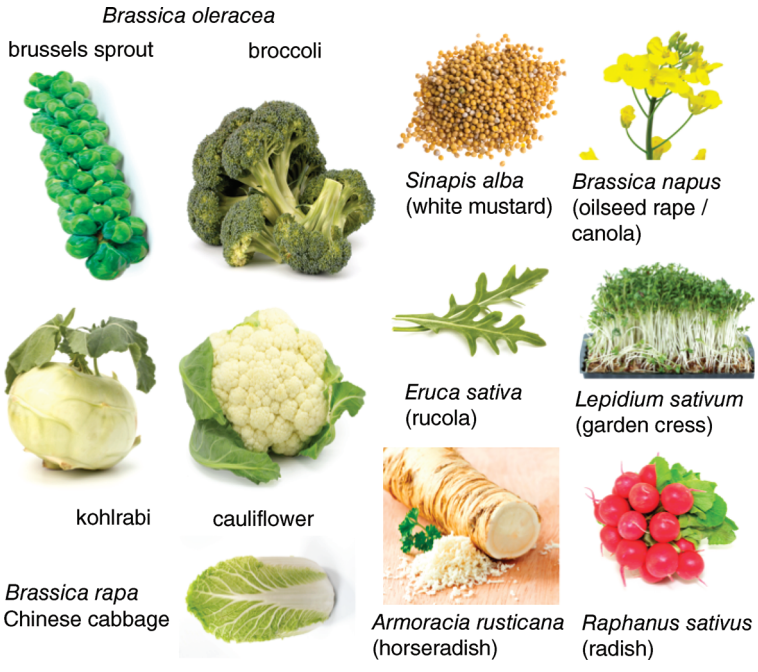


Figure 1.10 Some examples of crop, vegetable and spice plants belonging to the Brassicaceae. Only the species, not cultivars, are given. Source: Fotolia.com (Chinese cabbage, © Bernd, S.; oilseed rape, © Schliemer; radish, © photocrew; cauliflower, © dimakp; Brussels sprout, © narumol pug; cress, © eyetronic; mustard, © Andrey Starostin; rucola, © hvoya; horseradish, © Hetizia; broccoli, kohlrabi, © Natika).

after the enzyme myrosinase. In contrast, the glucosinolates are localized in so-called S-cells (Figure 1.11), so-named due to their high sulfur (S) content. This specific cell-type localization of the glucosinolate–myrosinase system has been identified in the inflorenscences of *A. thaliana* and *Brassica napus* (oilseed rape).

Among the *Brassica* species used as food or spices, compartmentation is especially important because it gives rise to the special taste of these crops. Some of the metabolites also have anticancerogenic potential, so that their formation is desirable (Chapter 6). Other cell types contain the so-called "specifier proteins"; these are important when determining the metabolic fate of the glucosinolate as either volatile (iso)thiocyanates, or as less-toxic nitriles that can be further converted to carboxylic acids (Figure 1.12).

The precursors required for the biosynthesis of all these compounds are amino acids (Figure 1.12). The glucosinolates can be divided into three large groups, depending on the respective precursor amino acid: (i) aliphatic glucosinolates deriving from methionine; (ii) aromatic glucosinolates deriving from phenylalanine and tyrosine; and (iii) indolic glucosinolates originating from tryptophan. Not all brassicaceous plants contain all groups of glucosinolates. For example, the

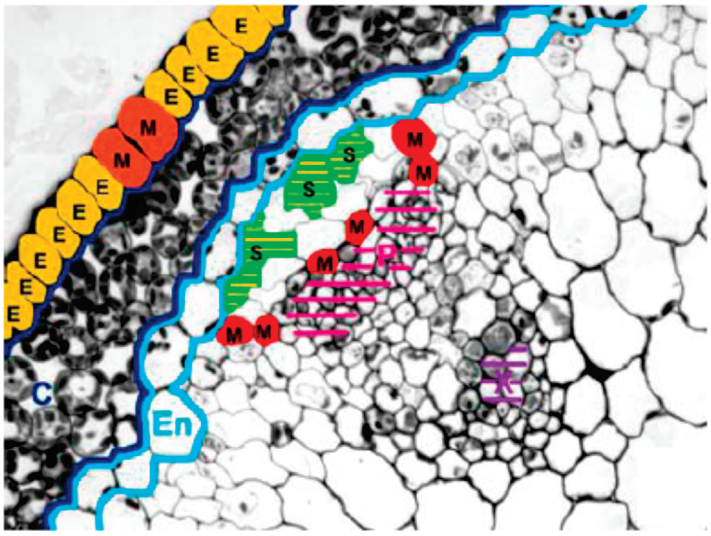


Figure 1.11 Cellular compartmentation of the glucosinolate–myrosinase system in inflorescences of *Arabidopsis thaliana*. A light microscopy cross-section shows the different cell types labeled in different colors: glucosinolate-containing S-cells (green), the myrosin cells (M) containing the β-thioglycosidase myrosinase (red and orange), the epidermal cells (E), which contain epithiospecifier proteins (yellow), the cortex (C) (dark blue), the endodermis (En; cyan), phloem (P; pink) and xylem (X; violet). From Kissen, R. *et al.* (2009) *Phytochem. Rev.*, **8**, 69 (see Further Reading). Reprinted with permission from Springer.

model plant *Arabidopsis* contains aliphatic and indole glucosinolates, but no aromatic representatives. On the other hand, this is the major glucosinolate group in the garden cress (*Lepidium sativum*) and in nasturtium (*Tropaeolum majus*) which, in addition, contains indole glucosinolates.

The major variation in the aliphatic glucosinolates is derived from enzymatic chain elongation reactions of the amino acid methionine. These chain-elongated precursors are then used for the formation of the next intermediate aldoxime. This enzymatic step, catalyzed by cytochrome P450-dependent monooxygenases, is quite specific for the amino acid substrate. In contrast, the later steps – that is, the addition of a sulfur group and glycosylation – are not very specific, and the respective enzymes accept a variety of different glucosinolate structures. This implies that only a modification of the specific step could lead to novel compounds in other plant species, because they can transform the new substance by their set of enzymes to the final glucosinolate. This has been achieved with *Arabidopsis*, where the biosynthesis of benzylglucosinolate was engineered by transformation with the gene encoding the protein for the aldoxime synthetic reaction. The late modification steps, such as hydroxylation or methylation, are again specific for the individual glucosinolates. The range of possible enzymatic reactions can differ greatly between the aliphatics on the one hand and indole glucosinolates on the other hand (Figure 1.12). Finally, unusual modification reactions have been reported for glucosinolates, such

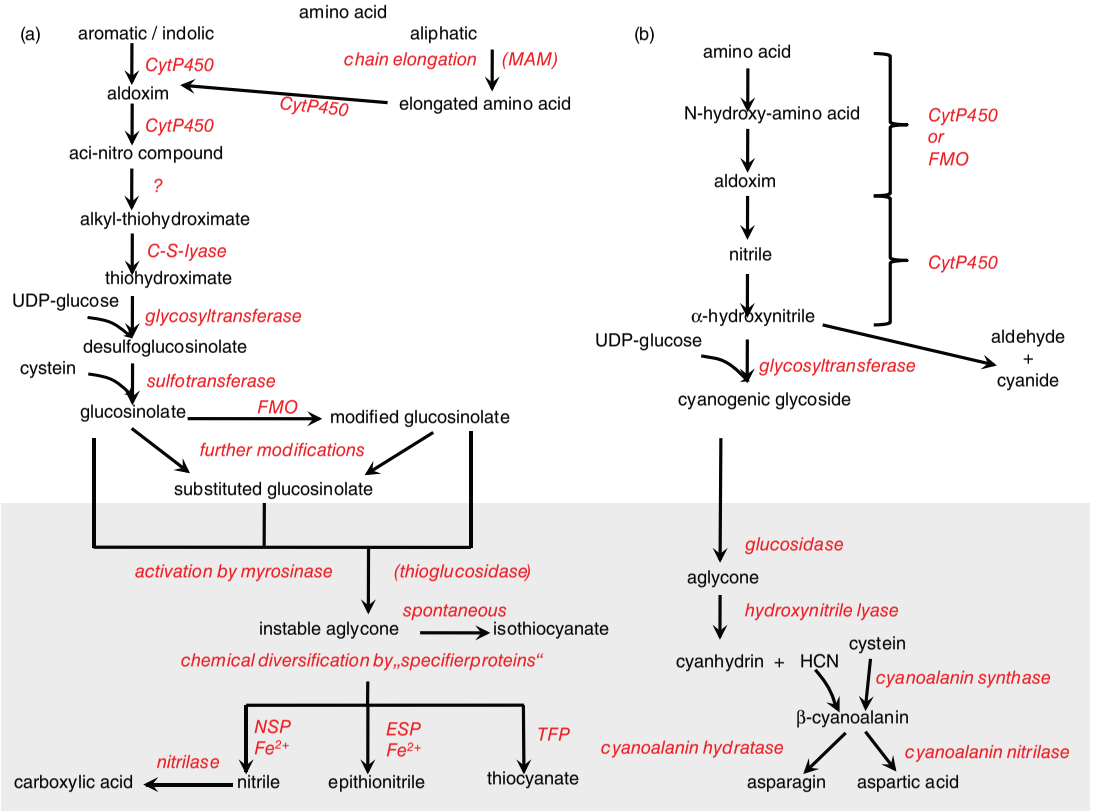


Figure 1.12 Biosynthetic pathways (light gray) and metabolism (gray) of glucosinolates and cyanogenic glycosides in comparison. For the glucosinolates the general scheme is shown for the different amino acids as precursors (a) and for cyanogenic glycosides the main pathway is shown (b). In both cases the metabolism leads to toxic compounds. For cyanogenic glycosides the deglycosylation yields cyanide, whereas for glucosinolates the metabolism is more complicated involving so-called specifier proteins. MAM = methylthioalkylmalate synthase; FMO = flavin-dependent monooxygenase; CytP450 = cytochrome P450-dependent monooxygenase; "?" indicates a reaction step for which no enzyme has yet been identified; ESP = epithiospecifier protein; NSP = nitrile-specifier protein; TFP = thiocyanate-forming protein.

as benzoylation and sinapoylation. In these cases, a benzyl or sinapoyl moiety is attached following the esterification of hydroxylated glucosinolates to the core structure.

Closely related to the indole glucosinolate pathway is the biosynthesis of other indolic compounds, for example the indole-type phytoalexins in *Brassica* species. In *Arabidopsis* there is one phytoalexin (camalexin) with an indole moiety present, whereas other *Brassica* species contain several indole phytoalexins, such as cyclobrassinin, brassilexin, rutalexin, or spirobrassinin (see Section 2.5). The biosynthetic pathway of camalexin is a branch of the indole-3-acetaldoxime pathway leading to indole glucosinolates (see Figure 1.19), whereas the biosynthesis of the *Brassica* indole phytoalexins derives from either tryptophan, indole or indole-3-glycerolphosphate, thus representing an earlier branch point in the metabolic network of indole metabolites.

Apart from the indole glucosinolates and indole phytoalexins of the Brassicaceae, other indole derivatives – which are also considered as defense compounds – are synthesized in different plant species. Among these are the terpene indole alkaloids (Figure 1.6) and a volatile compound in maize (DIMBOA; 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one). A detailed description of the biosynthetic pathway of this compound is provided in Section 1.2 (see Figure 1.25). Briefly, the precursor is also indole-3-glycerol phosphate, which is converted in two enzymatic steps to DIBOA (2,4-dihydroxy-2*H*-1,4-benzoxazin-3(4*H*)-one). DIBOA is glycosylated by a uridine diphosphate (UDP)-dependent glycosyltransferase; the glycosylated DIBOA is then converted to DIMBOA-glycoside, which is stored in the vacuoles. As in the case of glucosinolates, the glycosylated form is not toxic and must be deglycosylated to yield the toxic metabolite (the chemical structure is shown in Figures 2.30 and 3.12).

The cyanogenic glycosides are also synthesized from an amino acid (see Figure 1.12) although, in contrast to the glucosinolates, they do not occur predominantly in one plant family. The best-understood pathway is that yielding dhurrin, the cyanogenic glycoside of millet (*Sorghum bicolor*). In general, the amino acid is converted to an aldoxime, which is reminiscent of a biosynthetic step in glucosinolate synthesis. The aldoxime is further converted via a nitrile as an intermediate to the cyanogenic glycoside. The nitrile can also be converted to an aldehyde and cyanide. Detoxification of cyanide is therefore an essential part of the metabolic pathways leading to these compounds. In contrast to the formation of aldoxime in the glucosinolate pathway, the enzyme catalyzing the first step in cyanogenic glycoside synthesis is dependent on NADPH. Metabolic channeling was also shown to occur for the plasma membrane-bound enzymes, which means that the metabolites are not released from the enzymes after each step of the reaction. Nonetheless, this is a very efficient procedure in terms of the availability of precursors in low concentrations.

The degradation of glucosinolates and cyanogenic glycosides to toxic metabolites closely resembles one another in the first steps (see Figure 1.12). Metabolic conversion is especially vital for glucosinolates to become bioactive, and their fate is mainly determined by protein factors, which influence the outcome of the reaction. While the β-thioglucosidase myrosinase catalyzes the deglycosylation, the so-called "specifier proteins" determine the next reaction product. Isothiocyanates

may be formed spontaneously. An epithiospecifier protein is involved in the formation of an epithionitrile, a nitrile specifier protein in the formation of a nitrile, which is less toxic. These reaction products can be further metabolized to the corresponding carboxylic acid by nitrilases (see Figures 1.12 and 1.24). Finally, a thiocyanate-forming protein gives rise to the toxic thiocyanate. If none of these specifier proteins is present, then the isothiocyanate is mostly released, because of its volatility. Cyanogenic glycosides are also deglycosylated, which results in an aglycone that is further converted to cyanohydrin and the toxic cyanide (HCN) by the enzyme hydroxynitrile lyase. The cyanide group can be transferred enzymatically to the amino acid cysteine, such that β-cyanoalanine is the final product. The latter may then be converted by a specialized nitrilase to aspartate, or by a nitrile hydratase to asparagine (see Figure 1.24). These reactions are important to the plant, because of the extreme toxicity of the cyanide.

1.1.2.5 **Terpenes**

Among the terpenes some intoxicating or hallucinogenic compounds can be found; examples are the cannabinoids (see Section 5.3.2), which are synthesized predominantly via the methylerythritol–phosphate (MEP) pathway in the plastids. Terpenes can be classified as either essential or nonessential compounds (Figure 1.13). Essential terpenes include the carotenoids, which play an important role in photosynthesis as components of light-harvesting complexes, and also as protective compounds against high light intensities, but also antioxidative compounds such as tocopherol. Many plant hormones also belong to the terpenoid family, but as these are essential signaling molecules for growth and development and occur in rather small amounts, they are not usually viewed as secondary metabolites. Among nonessential terpenes antimicrobial substances such as the monoterpenes can be found. The term nonessential means that the plant is at an advantage if it can produce these substances under stress conditions; however, under normal conditions nonessential terpenes will not result in a phenotype or prove to be fatal if the biosynthetic pathway is mutated.

The biosynthesis of terpenes can proceed via two different pathways: (i) the mevalonate (MVA) pathway; or (ii) the methylerythritol-phosphate (MEP) pathway (Figure 1.14) (these names derive from the major metabolites involved). The MVA pathway is located in the cytosol, while the MEP pathway is in the plastids. Although both reactions can occur in the plant simultaneously, some organisms use only one pathway for the synthesis of terpenes. The sesquiterpenes and triterpenes are synthesized in the cytosol, whereas the plastid is responsible for synthesis of the mono-, di-, and tetraterpene structures. When coupled to the ER, the oxidation steps are catalyzed by cytochrome P450-dependent monooxygenases. Terpenes can also be produced in non-green plastids, for example in the roots of plants inoculated with arbuscular mycorrhizal fungi (see Figure 2.23). It has been assumed that these two pathways interact with each other, as both are essential for synthesis of the so-called "building blocks" isopentenyl-diphosphate (IPP) and dimethylallyl-diphosphate (DMAP), which are used in the synthesis of higher-molecular-weight terpenes via head-to-tail additions. This reaction is catalyzed by prenyltransferases.

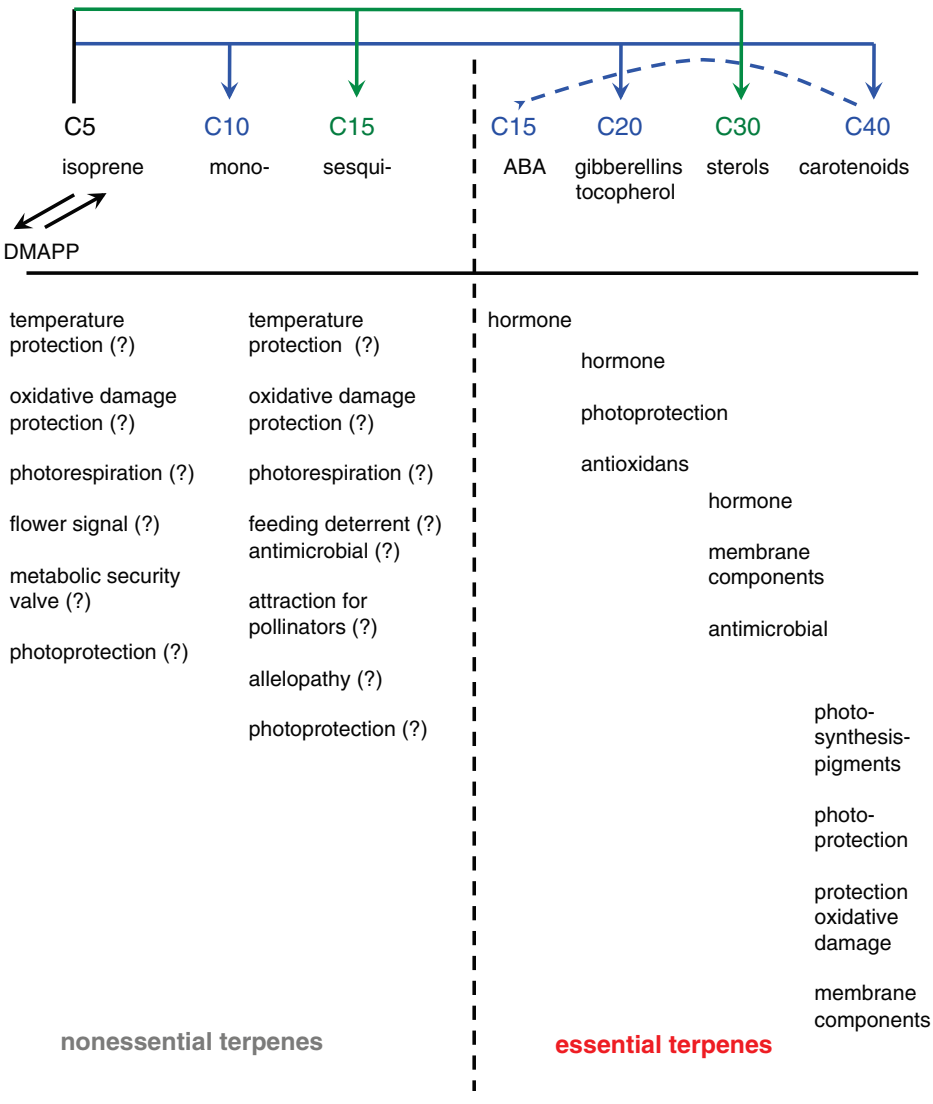


Figure 1.13 Classification of terpenes into essential and nonessential compounds for the plant. Examples of compounds and their biological function are given. Within the respective chemical scaffold, substances deriving from odd numbers of carbon atoms are shown in green; those deriving from even numbers of carbon atoms are shown in blue.

Longer-chain terpenoids are synthesized from two diphosphate molecules, where one phosphate group is lost. This reaction is typical for the synthesis of chlorophylls and ubiquinones. The key enzymes for the MVA pathway are hydroxymethylglutaryl-CoA synthase and -reductase, which catalyze the formation of mevalonic acid. Similarly, there are also key enzymatic steps for the MEP-pathway, involving 1-deoxy-D-

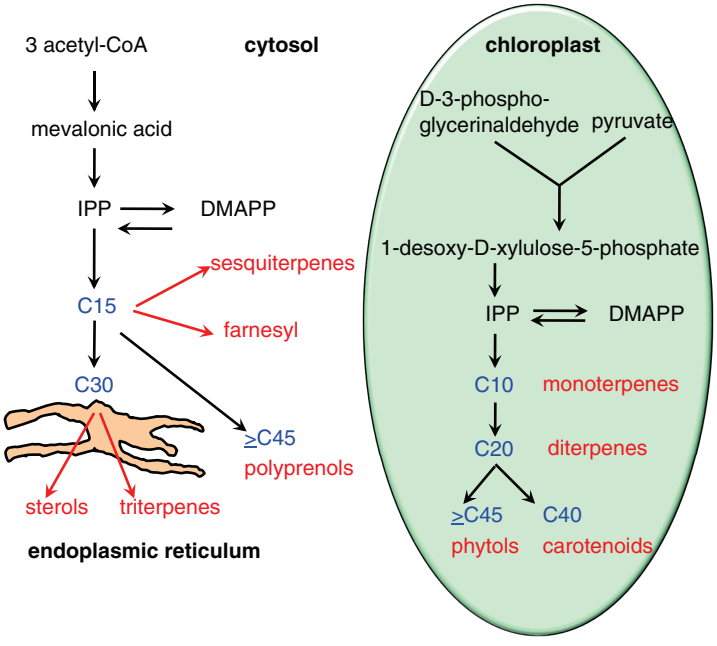


Figure 1.14 General scheme for terpene biosynthesis. The biosynthesis of terpenes can take place in two different compartments, the cytosol and chloroplast, using different pathways. The mevalonate (MVA) pathway in the cytosol delivers sesqui- and triterpenes, whereas the plastidic MEP (methylerythritol phosphate) pathway results in the synthesis of mono-, di-, and tetraterpenes. The endoplasmic reticulum hosts cytochrome P450-dependent monooxygenases, which catalyze various oxidation reactions. Both pathways are thought to interact with each other. IPP = isopentenylpyrophosphate; DMAPP = dimethylallylpyrophosphate.

xylulose-5-phosphate synthase and -reductoisomerase, which form 2-*C*-methyl-D-erythritol-4-phosphate. The later steps for both pathways are then catalyzed by terpene synthases (see Section 1.2) that are encoded by multigene families and are able to synthesize various longer-chain terpenes from diphosphate molecules. The number of C-atoms in the molecule is only dependent on the respective precursor, for example, geranyl-geranyl-diphosphate is converted to diterpenes by terpene synthases. Terpenes are also important precursors for the terpene indole alkaloids (Figure 1.6).

1.1.3 Transcriptional Control

The biosynthesis of secondary metabolites must be regulated in a timely and spatial manner, mainly because of the large energy costs associated with the process (Figure 1.15). As a plant cannot provide a complete range of secondary metabolites needed to survive stress situations at all times, regulation of the process is especially important for compounds that contain several elements, such as nitrogen-

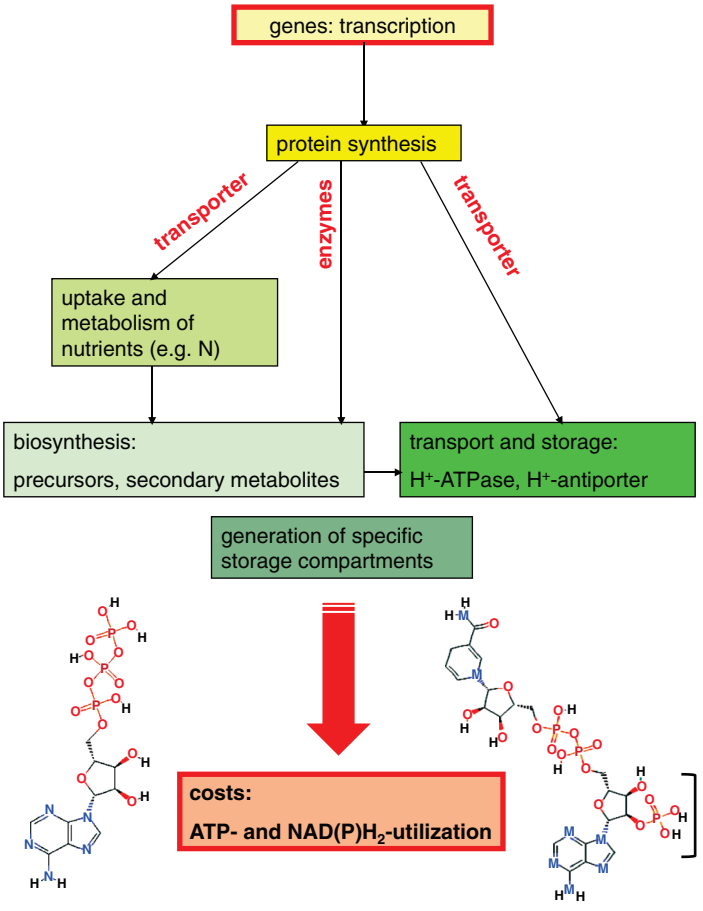


Figure 1.15 The biosynthesis of secondary metabolites is energy-consuming; therefore not all genes are transcribed at a given time point, and subsequently not all proteins present in the cells. Proteins are needed for biosynthesis, uptake and transport within the plant. For all these steps ATP and reduction equivalents, that is, NAD(P)H_2, are needed. The structures of adenosine triphosphate (ATP; left side) and nicotinamide-adenine dinucleotide phosphate (NADPH; right side) are shown. The bracket indicates the phosphate group not present in nicotinamide-adenine dinucleotide (NADH). Chemical structures are from The PubChem Project.

containing alkaloids and nitrogen- and sulfur-containing glucosinolates. Previously, a modeling approach has shown that the equivalent of about 15% of the total energy needed to synthesize all components of a leaf cell must be invested in the synthesis of all the glucosinolates that occur in *A. thaliana*. This results in a drain of oxygen, carbon dioxide, water, sulfur and nitrogen to the amino acid precursors needed to synthesize a secondary metabolite moiety. Clearly, the production of only one compound in large amounts under stress conditions might lead to disadvantages for the plant if a second stressor were to occur because, in the case of this

second stress, the "wrong" metabolite had been generated (see Sections 2.3 and 2.5). This of course implies that many metabolites must be generated simultaneously, while others would not be synthesized if not specifically required.

The energy to drive the chemical reactions must be provided by adenosine triphosphate (ATP) and reduction equivalents ($NAD(P)H_2$), which are needed not only for the enzymatic reactions and transport processes but also for the assimilation of nutrients. The tissue also requires specific storage compartments within the cell, or within an organ. Finally, messenger ribonucleic acid (mRNA) and proteins need to be synthesized. Consequently, regulation of the biosynthesis of enzymes and transporters on a transcriptional or post-transcriptional level is extremely important. For some compounds, the metabolites must be enzymatically activated (Figure 1.12), at both the right time and location.

Several options can be envisioned to regulate the synthesis of secondary metabolites, the most important being perhaps transcriptional control (see Figures 1.17 and 1.19). Transcription factors are DNA-binding proteins that interact with the regulatory promoter regions of the target genes and modulate the rate of transcriptional initiation by RNA polymerase. Several families of transcription factors have been shown to participate in controlling the biosynthesis and accumulation of secondary metabolites (Table 1.5). To date, some have only been found in one plant species, whilst others seem to occur more globally within the higher plants. Transcription factors belong mainly to seven distinct families, and their regulators integrate both internal (e.g., developmental) and external signals. These regulators bind to the corresponding control elements in the promoter (*cis*-elements) to either activate or repress the expression of enzyme-coding genes, while some will interact with other transcription factors to form a complex (see Figure 1.17).

In addition, both post-transcriptional (via microRNAs; Figure 1.16) and post-translational mechanisms (such as the activation of proteins by modifications; i.e., glycosylation, phosphorylation) can regulate the synthesis of secondary metabolites. Phosphorylation is often involved in the activation of transcription factors by protein kinases (Figure 1.19), while deactivation can be achieved by dephosphorylations catalyzed by protein phosphatases. Enzymes can also be activated or inhibited by substrates or reaction products.

MicroRNAs (miRNAs) are short, highly conserved noncoding RNAs that are capable of mediating the silencing of target mRNA molecules. This ability was first detected as a suppressive strategy of plants against viral infections. In addition, miRNAs are involved in the regulation of developmental processes, and can control the interaction of a plant with both abiotic and biotic environments. The same ability is also used to control post-transcriptional gene regulation *in vitro*, by transforming plants with miRNA precursors. Post-transcriptional regulation could also be involved in regulating secondary metabolism. The miRNAs are encoded by MIR genes, which are transcribed from the genomic DNA (Figure 1.16), and the resultant primary miRNA transcript then folds back into a double-stranded structure that is processed by a protein named "dicer" to produce a 21–30 nucleotide-long miRNA. The gene encoding a miRNA precursor has two features: one feature forms part of the complementary DNA, while the second feature forms a

Table 1.5 Families of transcription factors involved in the regulation of plant secondary metabolite genes. These belong mainly to seven classes: MYB, bHLH, AP2/ERF, WRKY, Zinc finger, DOF and NAC. For the listed transcription factors a plant species is given as an example for the presence of the respective transcription factor. The relevant metabolic pathway is also given. Compiled from Yang, C.Q. *et al.* (2012) *J. Integr. Plant Biol.*, **54**, 703–712.

Transcription factor	Metabolic pathway	Plant species
Superfamily MYB		
MYB	Anthocyanins	*Arabidopsis thaliana*
MYB	Anthocyanins	Asiatic hybrid lily
MYB	Anthocyanins	*Malus* × *domestica*
MYB	Anthocyanins	*Vitis vinifera*
MYB	Glucosinolates	*Arabidopsis thaliana*
MYB	Phenylpropanoids	*Nicotiana tabacum*
Ruby	Anthocyanins	*Camellia sinensis*
TT2	Proanthocyanidins	*Arabidopsis thaliana*
PAR	Proanthocyanidins	*Medicago truncatula*
Superfamily bHLH		
GL3	Anthocyanins	*Arabidopsis thaliana*
TT8	Anthocyanins	*Arabidopsis thaliana*
MYC	Terpene indole alkaloids	*Catharanthus roseus*
MYC	Terpene	*Arabidopsis thaliana*
MYC	Nicotine	*Nicotiana tabacum*
bHLH	Nicotine	*Nicotiana benthamiana*
Superfamily AP2/ERF		
ORCA	Terpene indole alkaloids	*Catharanthus roseus*
ERF	Artemisinin	*Artemisia annua*
Superfamily WRKY		
WRKY	Gossypol	*Gossypium arboreum*
WRKY	Artemisinin	*Artemisia annua*
WRKY	Camalexin	*Arabidopsis thaliana*
WRKY	Terpene indole alkaloids	*Catharanthus roseus*
Superfamily Zinc finger		
ZCT	Terpene indole alkaloids	*Catharanthus roseus*
Superfamily DOF		
OBP	Glucosinolates	*Arabidopsis thaliana*
DOF	Flavonoids	*Arabidopsis thaliana*
Superfamily NAC		
NAC	Camalexin	*Arabidopsis thaliana*

loop that has no complementary sequences but is structurally important for folding of the RNA. The miRNA∗ strand (see Figure 1.16) is later degraded, while the miRNA strand binds to its complementary mRNA molecule. The cell recognizes the partially double-stranded RNA, which is then degraded in a ribonucleoprotein complex, the so-called RNA-induced silencing complex (RISC), of which an essential component is the protein argonaute. Finally, the initially transcribed mRNA is degraded into short fragments. In the case of a virus infection, the viral nucleic

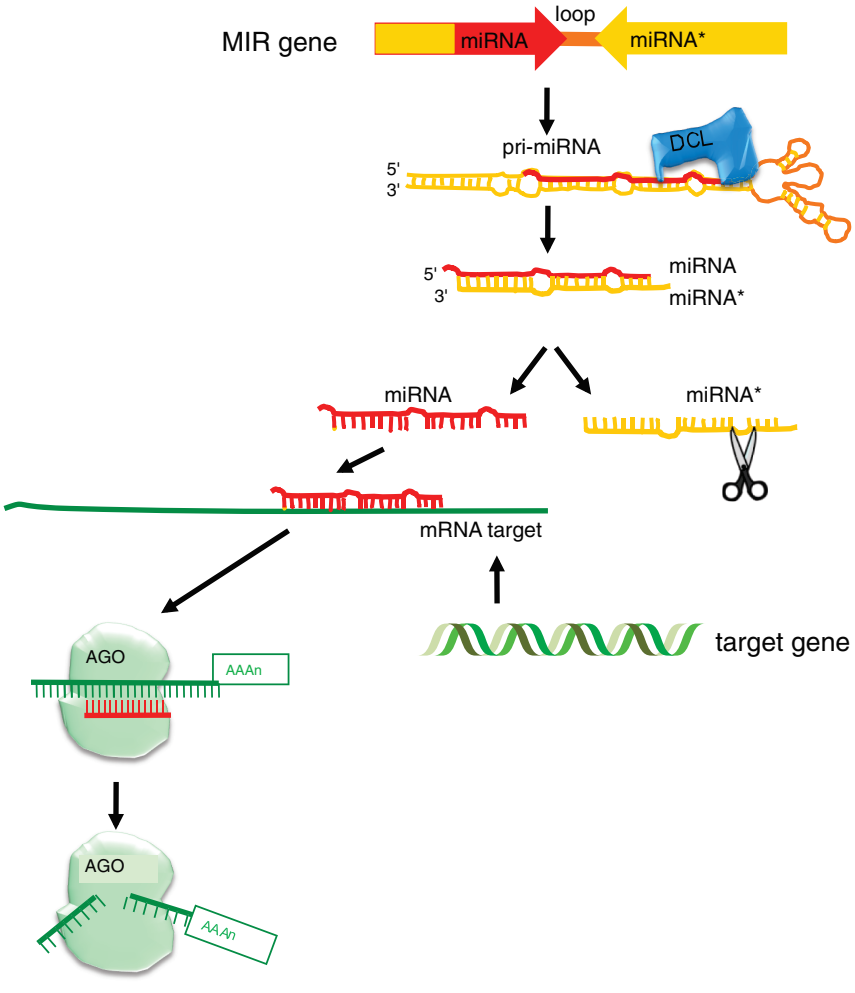


Figure 1.16 MicroRNA-mediated silencing of mRNA and translational repression. miRNAs are encoded by MIR genes. The primary miRNA (pri-miRNA) transcript folds back into a double-stranded structure, which is processed by dicer (DCL1). The miRNA* strand is degraded and the miRNA strands binds to the target gene (in green). Modified from *Teaching Tools in Plant Biology*, The American Society of Plant Biologists.

acids dimerize spontaneously when a certain number of molecules is present. This RNA duplex is also recognized by the plant, and degraded by dicer. The resultant short RNAs, which are referred to as short interfering RNAs (siRNAs), can bind further RNA molecules so that their numbers are rapidly decreased.

Many of the transcription factors involved in controlling secondary metabolite synthesis are also controlled by signaling molecules, such as jasmonic acid (Figures 1.17–1.19). Jasmonic acid and salicylic acid are both important signaling molecules in plant defense reactions against microbes and herbivores. Jasmonates, for

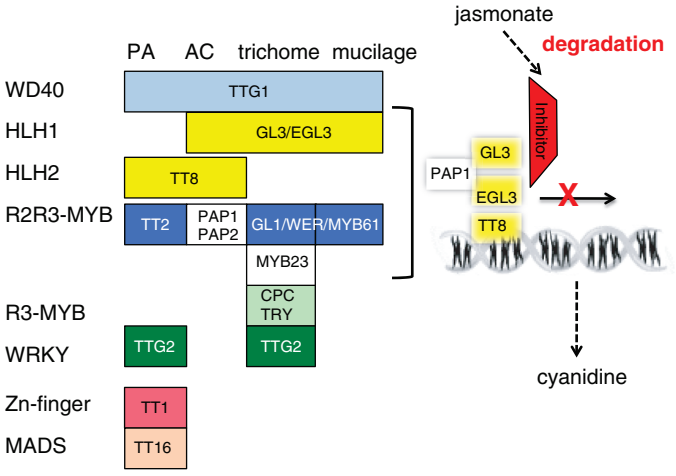


Figure 1.17 Transcriptional control of pigment and proanthocyanidin synthesis in *Arabidopsis thaliana*. The different families of transcription factors involved in the different biosynthetic pathways, proanthocyanidin (PA) and anthocyanidin (AC) synthesis are shown on the left. The developmental processes such as pigmentation, trichome and mucilage formation in *Arabidopsis* are shown at the top. The transcription factors sometimes need to form complexes to gain activity. They can also be regulated by inhibitor molecules (red). The inhibitor is degraded in a jasmonate-dependent manner (see also Figure 1.18). The gene names of some transcription factors are inspired by the phenotype of the mutant. TT = transparent testa; GL = glabrous (no trichomes); PAP = production anthocyanin pigment; WER = werewolf; EGL = enhancer of glabrous; TRY = triptychon; AN = anthocyanin.

example, can regulate many transcription factors such as the MYB family, AP/ERF family, and WRKY family, which implies that jasmonic acid can regulate the synthesis of a wide spectrum of secondary metabolites. Among these are quite different structures such as glucosinolates and camalexin in *Arabidopsis* (Figure 1.19), flavonoids and anthocyanidins in many plant species (Figure 1.17), terpene indole alkaloids in *Catharanthus roseus* (Figure 1.34), nicotine in tobacco, and artemisinin in *Artemisia*. Consequently, jasmonic acid is often used to elicit the synthesis of secondary metabolites in biotechnology (Table 1.9). The transcriptional control of secondary metabolite biosynthesis will be explained by using two classes of compounds as examples, namely the flavonoids and the glucosinolates.

The biosynthesis of flavonoids and anthocyanins is regulated in a compound-specific manner, and the induction is triggered either tissue- and/or stress-specifically (Figure 1.17). Often, several transcription factors need to interact with each other to exert their correct function, although in some cases one transcription factor may be active alone. Pigmentation is important not only for flowers and fruit coloring, but also for seed coatings to provide protection against high levels of irradiation. Developmental processes, such as the formation of trichomes or the production of mucilage, are also regulated by this group of transcription factors and flavonoids. The transcription factor complexes are able to regulate either one

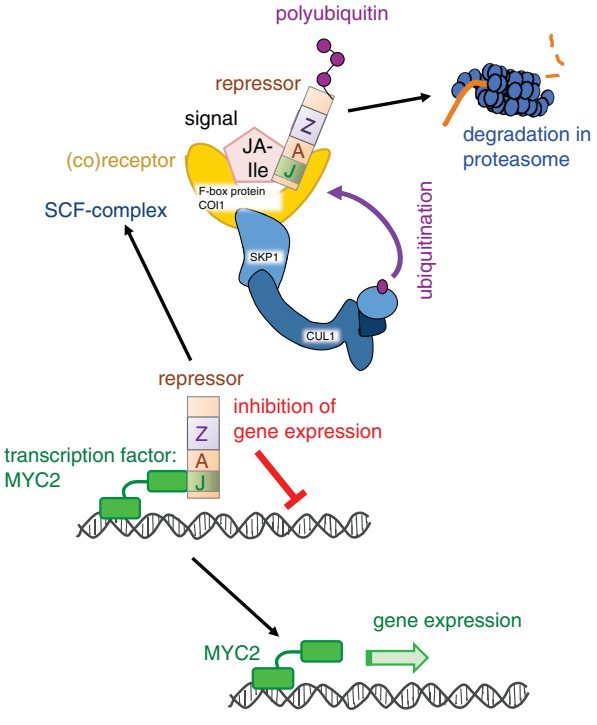


Figure 1.18 The COI signaling pathway of jasmonate-mediated gene expression regulation. Jasmonate-isoleucine (JA-Ile) binding by the COI1-JAZ coreceptor is shown. Ubiquitination and degradation of the JAZ inhibitor in the 26 S proteasome is the prerequisite for transcriptional activation. The repressor is targeted for degradation by addition of a poly-ubiquitin tail in the SCF-complex (SCF = SKP, CUL, F-box, where the F-box protein is the receptor COI1). Transcriptional activation is then exerted by MYC2 transcription factors and others. Modified from *Teaching Tools in Plant Biology*, American Society of Plant Biologists.

specific or several processes. For example, at least six transcription factors are involved in the synthesis of proanthocyanidins, whereas for mucilage formation only three are involved (Figure 1.17). However, the regulatory network is much more complex than depicted here, mainly because no environmental factors such as light, pathogens, or other abiotic stresses have been taken into account.

There are proteinaceous inhibitors of jasmonic acid-induced transcription, which have to be degraded before gene activation is possible (Figure 1.18). In this case, it is not jasmonic acid itself but rather its conjugate with the amino acid isoleucine that triggers this process. Under noninducing conditions a transcriptional repressor (JAZ) binds to the promoter of jasmonate-inducible genes, and this prevents transcriptional activation by the MYC transcription factors. In order for the repressor to be degraded, the amino acid conjugate of jasmonic acid with isoleucine is needed as signal. Binding of the jasmonate conjugate together with the repressor protein to a receptor (COI1), which is an F-box protein, targets the repressor to proteolytic degradation in the 26S proteasome. For this, the target protein must be

ubiquitinated as only proteins with a polyubiquitin tail are recognized by the proteasome. Ubiquitination is achieved in the SCF (SKP, CUL, F-box) complex, which is composed of the receptor itself (the F-box protein), the ubiquitin ligase, and other adapter proteins. Once the repressor is degraded, the MYC-type transcription factor can activate the transcription of the jasmonate-inducible gene.

The biosynthesis of glucosinolates is regulated by many different environmental factors (see Figure 1.19), and consequently many transcriptional events need to be induced under these conditions. Specific transcription factor families were isolated from *Arabidopsis*, which can control preferentially either the indole or aliphatic glucosinolate pathway (for details on the biosynthetic enzymes, see Section 1.1.2).

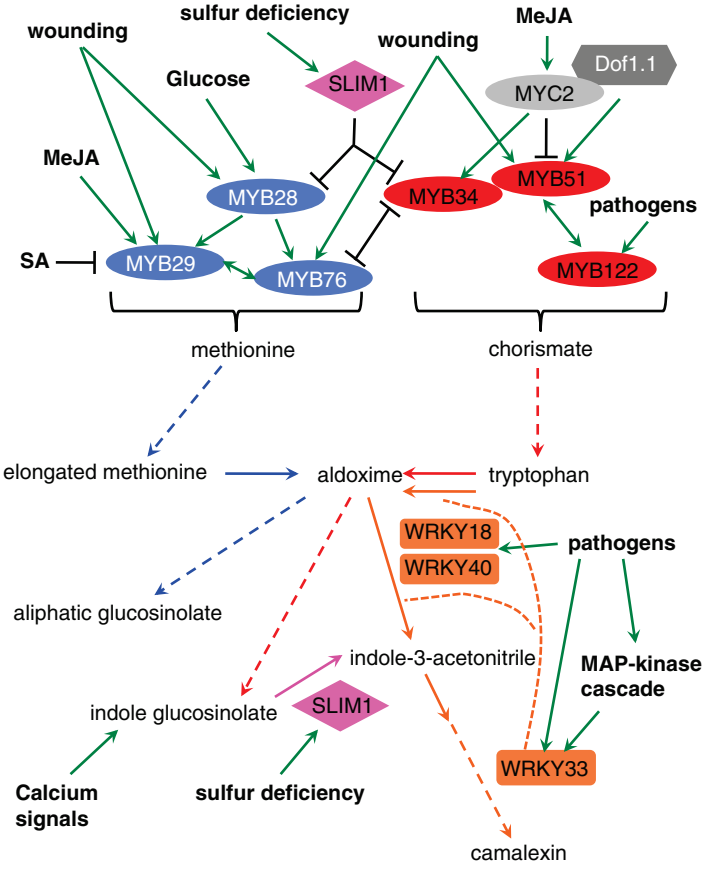


Figure 1.19 Transcriptional control of glucosinolate biosynthesis. Transcription factors involved in the control of aliphatic glucosinolate synthesis are shown in blue; those controlling indole glucosinolate synthesis are shown in red. Camalexin synthesis (orange) is regulated by WRKY transcription factors. WRKY33 itself is regulated by activation of a MAP kinase cascade. Green arrows point to an induction by environmental factors. A solid arrow symbolizes one reaction step, and a dashed arrow several reaction steps. Dof1.1/SLIM1/MYB/MYC/WRKY = transcription factors; MAP = mitogen-associated protein; Me-JA = methyl-jasmonate.

Environmental factors such as pathogens, sulfur deficiency and wounding, or nutrients such as glucose, can each induce transcription factors which, in turn, control genes involved in glucosinolate synthesis. This transcriptional control also involves the synthesis of precursors such as tryptophan. In addition, pathways connected to indole glucosinolates, such as the synthesis of the phytoalexin camalexin and the plant growth hormone IAA (Section 2.2.1), are controlled by members of these transcription factor families.

Some genes are regulated only by one transcription factor, and others by two or even three different factors, depending on the signal(s) to which they respond. Plant signals involved in biotic stress responses, such as salicylic acid or jasmonic acid, can also induce these transcription factors and in turn increase the levels of the glucosinolate defense compounds. Since glucosinolates are sulfur-containing molecules, it is feasible that their synthesis is transcriptionally regulated by sulfur. Sulfate deficiency induces the transcription factor SLIM1, which acts in fact as a repressor of the MYB-transcription factor family. The latter are positive regulators of aliphatic and indole glucosinolate biosynthesis, so that the upregulation of SLIM1 represses the biosynthesis of glucosinolates under unfavorable conditions. On the other hand, the metabolism of glucosinolates is induced by SLIM1, so that the sulfur is released for reuse in other metabolic reactions. Thus, the glucosinolates can be also viewed as S-storage compounds.

The biosynthetic pathway to the phytoalexin camalexin is closely connected to indole glucosinolate and IAA biosynthesis (Figure 2.4). However, the genes encoding the enzymes for the respective biosynthetic pathways are controlled by different transcription factors. While the glucosinolates are controlled by the MYB transcription factor family, WRKY transcription factors regulate camalexin synthesis (Figure 1.19). WRKY transcription factors are not only important for camalexin synthesis but also play a general role in the plant defense activation. One of these, WRKY33, is post-translationally regulated by a MAP kinase cascade, which results in the phosphorylation of the transcription factor. A negative control is achieved by dephosphorylation via a protein phosphatase.

Recently discovered proteins with signatures that associate them with the calcium response of the cell are also involved in the control of indole glucosinolate accumulation (Figure 1.19). These possess a calmodulin-binding motif. Calmodulin then interacts with calcium to activate signaling cascades. The calmodulin-binding protein from *Arabidopsis* has a nuclear-targeting sequence, which implies an effect on gene activation. This protein is especially induced after pathogen colonization. In addition, plants that overexpress the gene for this protein are more resistant to herbivore attack. This is in accordance with the high level of gene expression in the vasculature, especially in the phloem.

1.1.4
Transport of Secondary Metabolites

The molecular transport mechanisms of secondary metabolites (Figure 1.20) are less well understood than the biosynthesis, and very few details concerning transport

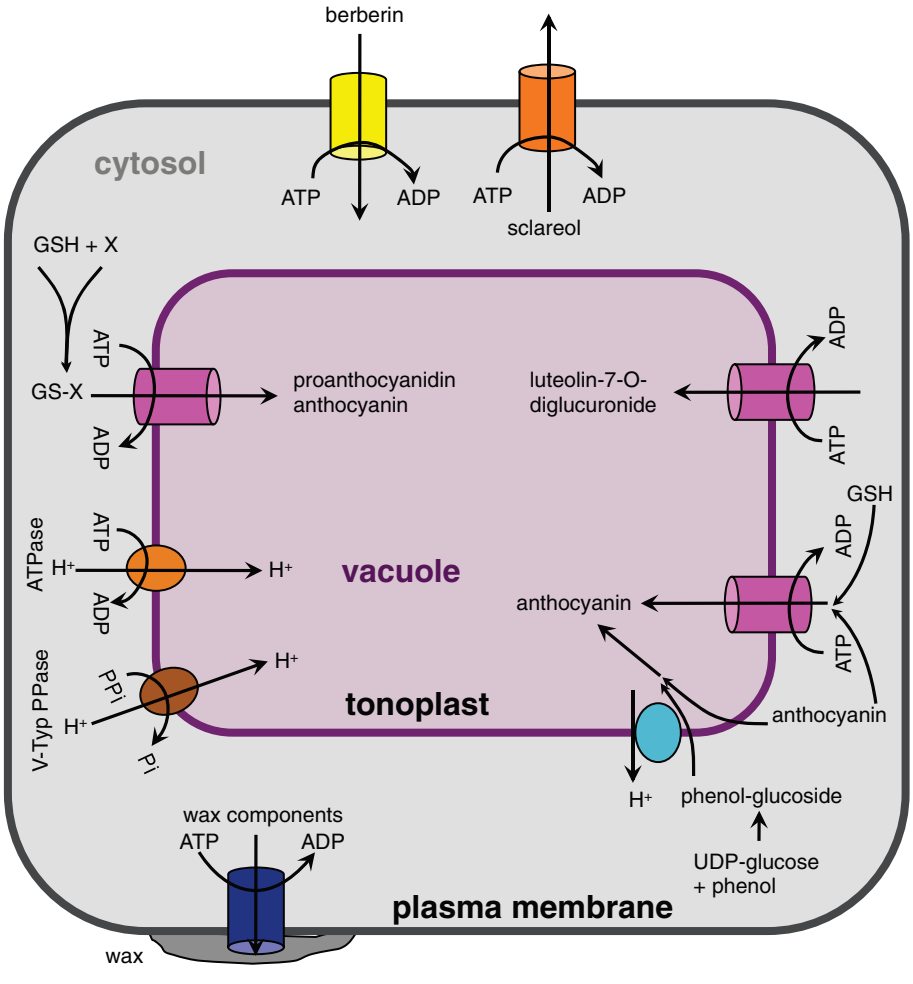


Figure 1.20 Transport of secondary metabolites in different cellular compartments. The ABC-transporters occur in high numbers and transport many different molecules. Representative secondary metabolites, which are transported by one class of these transporter types are shown: ABC-transporter of the MDR-family (yellow), of the PDR-type (orange), of the WBC-family (dark blue) and of the MRP-family (magenta) (for abbreviations see text). The ATP/H^+-dependent transporters (cyan) use a different mechanism. These are mainly dependent on the proton gradient and only indirectly on ATP. ATPase and PPase proteins (brown and dark orange) are involved in the formation of the proton gradient.

proteins and mechanisms are available. The compounds or precursors need to be transported into either the vacuole or other cellular compartments, where they are stored or further metabolized. One major transporter family which is thought to be involved in the transport of secondary metabolites is the ABC-transporters (ABC = ATP binding cassette). These are subdivided into three classes, which occur at the plasma membrane and are ATP-dependent: (i) the MDR (multidrug resistance)-type

family; (ii) the PDR (pleiotropic drug resistance)-type; and (iii) the WBC (white-brown complex)-type. The MRP-type-related proteins represent another form of ABC-transporter which is located at the tonoplast. These transporters differ, for example, in the number of transmembrane domains and the position of the nucleotide-binding site(s). The transport of some compounds is coupled to a proton antiport, and in this case the proton gradient must be maintained by either H^+-dependent ATPases or pyrophosphatases (PPases). Some flavonoids can be transported as glutathione derivatives while other compounds, such as simple phenols, are transported as glycosides.

Flavonoid glycosides, including anthocyanins, are usually transported into the vacuole. The transport mechanisms can depend on the plant species and organs, but the most likely mechanisms involve the transport of anthocyanins via a glutathione *S*-transferase (GST)-like protein or a MDR-like protein (a type of ABC transporter). The involvement of the former has been shown in maize, petunia and *Arabidopsis*, while the latter has been identified in maize. The molecular mechanism of how these protein conjugates (especially GSTs) are transported has not yet been clarified.

A second mechanism involves the vesicle-mediated mass transport of anthocyanins to vacuoles, as has been observed in lisianthus (*Eustoma grandiflorum*). Anthocyanins are also targeted directly to the protein storage vacuole via ER-derived vesicles in *Arabidopsis* seedlings, and this process does not depend on either GST activity or an ATP-dependent transport mechanism. For flavonoids, it has been suggested that their transport can be coupled to their synthesis by an enzyme complex located at the ER, where the final product is directly included in the ER vesicles and transported to the vacuole, where the vesicle is fused with the tonoplast. This type of transport is suitable only for compounds which cannot diffuse through membranes.

A third mechanism may involve an *Arabidopsis* multidrug and toxic compound extrusion (MATE) transporter (TT12). This vacuolar flavonoid/H^+-antiporter is necessary for the vacuolar accumulation of proanthocyanidins, and has been shown to mediate anthocyanin transport *in vitro*. The application of flavonoids to *Arabidopsis* roots can be used to determine their fate after uptake due to their fluorescence after a specific staining procedure *in planta*, using fluorescence microscopy (see also Figure 1.29). *Arabidopsis* roots grown in complete darkness do not accumulate flavonoids, as the expression of genes encoding enzymes of flavonoid biosynthesis is light-dependent. However, flavonoids do accumulate in the root tips of plants with light-grown shoots and light-shielded roots, consistent with shoot-to-root flavonoid movement. In addition, flavonoids were accumulated in the tissues at some distance from the site of application, which is indicative of uptake and movement systems. Flavonoid applications to the root tips resulted in a basipetal movement in the epidermal layers, with their detection 1 cm from the application sites after some incubation time. In addition, the movement of flavonoids towards the root tip occurred mainly in vascular tissues. The results of experiments with inhibitors of ABC-transporters mediating flavonoid transport suggested that a MDR-associated protein facilitates long-distance movement of the flavonoids.

Digitalis glycosides do not require a transporter, and have been observed to pass the plasma membrane simply by diffusion along a concentration gradient. In the cell, they are modified so that they cannot pass the plasma membrane by diffusion and so are trapped within the cell. At this point, a transporter is most likely needed to relocate these compounds to the vacuoles, where they are stored.

If secondary metabolites have to be transported from the origin of synthesis to the tissue where they are accumulating, then long-distance transport is essential. Very little is known about these processes, as most information is based on correlative experimental data. For example, if the site of biosynthesis and the accumulation patterns of an individual compound differ, then the existence of long-distance transport must be assumed. In tobacco (*Nicotiana tabacum*), pyridine alkaloids are synthesized in the roots and transported to the leaves, where their bioactivity is needed (see Sections 2.5.3 and 3.1). The transport within cells and from cell to cell is most likely possible for nicotine via a nonspecific transporter. Although long-distance transport can be achieved in the xylem vessels and the phloem, this type of transport is dependent on the direction and velocity of water and assimilated transport, and so is not very specific for individual compounds. Differences in nicotine distribution have been observed in two closely related tobacco species, *Nicotiana alata* and *N. langsdorffii*; whereas *N. langsdorffii* can translocate nicotine into the shoot, *N. alata* is not capable of this transport and consequently stores nicotine in the roots. Genetic analyses of these two species have shown that the characteristic of "no transport" is dominant over the ability to transport nicotine. The underlying genetic mechanisms for this property will eventually allow the genes responsible for long-distance nicotine transport in *Nicotiana* species to be identified. These findings also indicate that the trait of "transport" or "no transport" has evolved independently, and can be attributed to different environmental conditions where the accumulation of nicotine in leaves provides the plant with a defensive advantage against herbivores.

In contrast, more information is available for the glucosinolates, which exist as organic anions under the pH conditions of the cell and are therefore unable to cross membranes by simple diffusion. Similar to other ionic compounds, once the glucosinolates are inside the cell they become trapped and must be actively transported. As glucosinolates are found in high concentrations in S-cells, and in vacuoles within those cells, there is a need for an active transport mechanism for these compounds (Figure 1.21). The same has to be anticipated for the long-distance transport of glucosinolates, as they are transported through either the phloem or xylem; however, in order for them to be loaded into the vascular system they must also be able to pass membranes. It is possible that their synthesis in companion cells simplifies such uptake, without transporters, into the phloem system, from where they are further distributed (Figure 1.21).

Currently, biochemical evidence exists for the active transport of aliphatic glucosinolates. The application of radioactively labeled glucosinolates to rosette leaves of *Brassica napus* resulted in an accumulation of radioactivity within the maturing seeds, especially the embryo. Moreover, the phloem was also labeled, which points to a long-distance transport mechanism. Two closely related members of the nitrate/peptide transporter family were identified as high-affinity, proton-dependent,

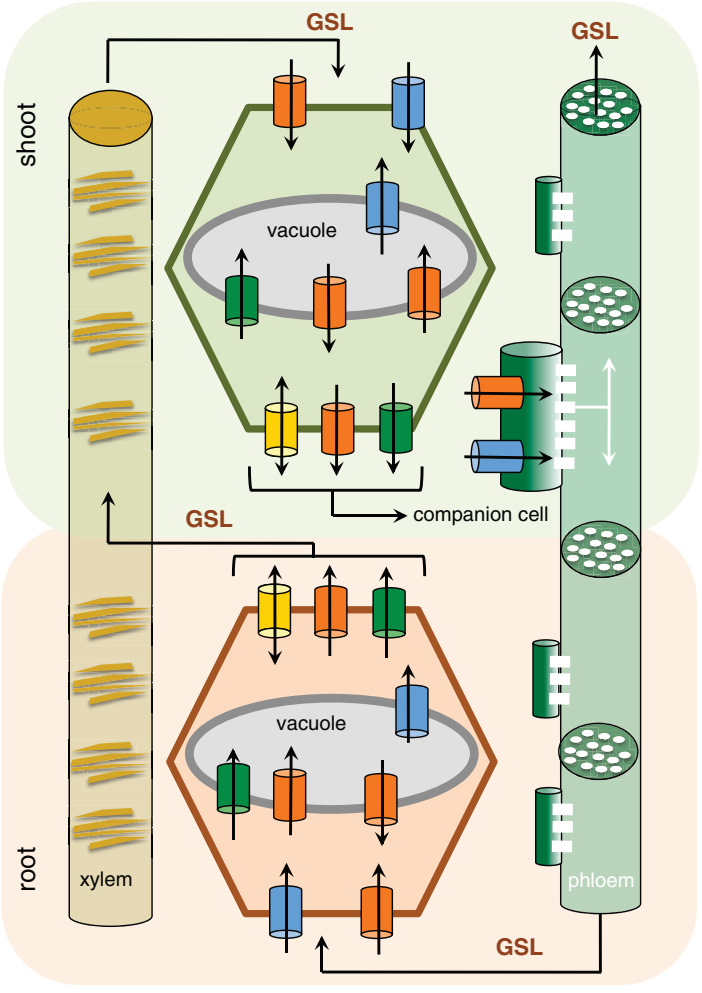


Figure 1.21 Long distance transport and loading/unloading at source and sink tissues of glucosinolates. Glucosinolates are synthesized close to the vasculature and possibly also in companion cells. Part of the unloading for glucosinolates is symplastic, so no transporter is needed, but some part is also apoplastic, and for this a transporter is needed. Glucosinolates are also transported into the vacuole by similar mechanisms, but maybe only by three different transport molecules, which mediate import and export. Blue = proton symporters PUP, NRT/PTR; orange = primary transporters ABC type, green = proton antiporters MATE type, yellow = not further specified permeases (no genes known). For simplification, only the direction of the transport of glucosinolates is shown and not dependence on H^+ or ATP. Redrawn from Nour-Eldin, H.H. and Halkier, B.A. (2013) *Curr. Opin. Biotechnol.*, **24**, 263.

glucosinolate-specific transporters that were most likely needed for the translocation of glucosinolates from the apoplast into the phloem. A double-mutant failed to accumulate glucosinolates in the seeds, but showed a more than tenfold higher accumulation of glucosinolates in source tissues such as leaves and silique walls. This

indicated that both plasma membrane-localized transporters are essential for the long-distance transport of glucosinolates. Finally, the direct transport of aliphatic glucosinolates across the plasma membrane of *B. napus* was demonstrated. Two components were identified for transport: one that was saturable, indicating a transporter; and one that was nonsaturable. Subsequent competition experiments showed the saturable transport to be specific for glucosinolates, as the uptake could be inhibited by other glucosinolates but not by glycosides or amino acids. Further experiments revealed that the transport was dependent on a pH gradient across the membrane, which indicates that ABC-transporters are not responsible for this transport.

1.2 Evolution of Natural Compounds in Plants

It is generally believed that secondary metabolites have evolved independently in different systematic groups of plants. Many plant species are polymorphic for secondary metabolites, which means that they show a natural variation between species or cultivars. This variation in secondary metabolites offers a wealth of information for studies on the evolution of the respective biosynthetic pathways. It is not possible, however, to use plant metabolites for a classification based on "chemical evolution" or chemical taxonomy. Only a few metabolites would fit into such a category that could be assigned to a specific systematic group of plants; for example, the betalains in the Caryophyllales and the glucosinolates mainly in the Brassicales. However, there is one exception to the latter grouping, namely the glucosinolate-containing genus *Drypetes*, which belongs to the Euphorbiaceae.

The betalains are composed of the yellow betaxanthins and the magenta betacyanins (Figure 1.9), and occur as aglycones or, as in the case of anthocyanins, also in glycosylated forms. Anthocyanins and betalains are not found within the same plant species as they are mutually exclusive; however, where no anthocyanins are present the betalains take over their respective functions. The reason for this phenomenon is unknown. Although, interestingly, the precursors of anthocyanin biosynthesis can be converted to leucoanthocyanins in species containing betalains, the later steps in the pathway are absent, which suggests that this part of anthocyanin synthesis has been lost in the Caryophyllales.

Tannins, lignins and lignans are among the "oldest" secondary metabolites (Figure 1.22), and for their synthesis the development of the shikimate pathway was necessary. In this case, aromatic amino acids serve as the precursors and are converted to cinnamic acid derivatives. These secondary metabolites are important for the structure and robustness of land plants, and are found in cell wall components; they are also vital for the development of complex plant architecture. Tannins have antimicrobial activities.

Benzylisoquinones evolved at a rather early stage, whereas terpenes and indoles developed only later. Isoflavonoids have evolved on several occasions, independently of one another. For many of these compounds, two different pathways for the delivery of precursor molecules have evolved in parallel; examples of this are

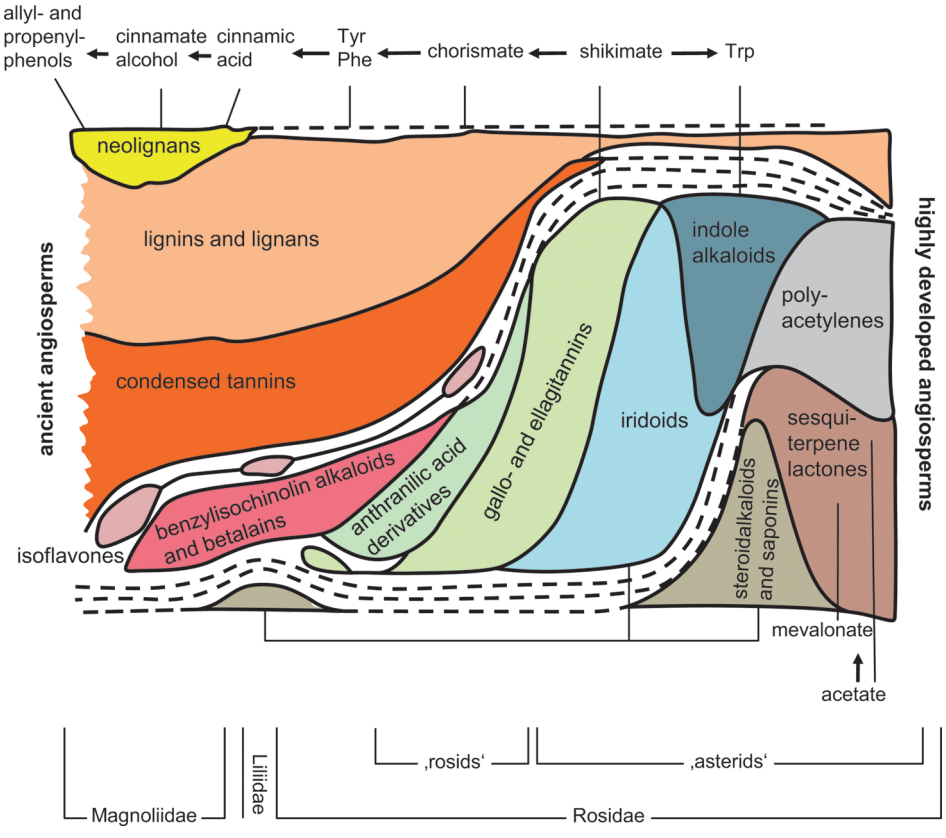


Figure 1.22 Development of different classes of natural compounds in the evolution of plants. The different colors symbolize large groups of secondary metabolites. The horizontal axis as a "time scale in evolution" shows the different plant groups. Above the figure the respective pathways for individual classes of compounds are indicated, which are necessary for their production. The vertical axis shows the compounds, which have been evolved simultaneously.

the shikimate and MVA pathways for the synthesis of terpene and indole alkaloids. Some aspects of the evolution of selected examples are presented in the following subsections. The genes involved in major pathways are known to have been present in the early land plants (e.g., mosses), which had a simpler body plan and where the number of homologous genes within one species would be much lower than in higher plants (Table 1.6). The greatest numbers of genes within a given gene family are found in angiosperms.

The size of the gene families can be affected by several molecular mechanisms:

1) Gene duplication or whole-genome duplication is an ongoing contributor to genome evolution, and occurs within the same order of magnitude as the mutation rate per nucleotide site.

2) The acquisition of new functions for existing genes, followed by gene duplication, is the cause of new gene families.
3) A reduction in gene families can also occur by the deletion of a single gene, or even of several connected genes.

The biosynthetic pathways for secondary metabolites are often modified pathways of the primary metabolism by way of gene duplication. However, gene duplications may also occur in secondary metabolite pathways, resulting in two "new" pathways (Figure 1.23). Although the "old" pathways can be retained, some parts might be lost or altered by new enzyme activities, thereby complicating investigations into the evolutionary descendance of metabolic pathways. In the case of a common evolutionary ancestor protein, the enzymes are recruited from a single precursor. However, closely related compounds can be synthesized by pathways that have evolved separately in different plant families, such that a protein, which catalyzes the same reactions in the two different species may have developed from completely separate precursor proteins. In this case, the whole biochemical pathway has evolved separately in the different plant groups (Figure 1.23).

1.2.1
Parallel Evolution of Similar Pathways

The biosynthetic pathways of two large groups of glycosides – the glucosinolates and cyanogenic glycosides (see Section 1.1.2) – possess obvious similarities in their biosynthetic pathway precursors (amino acids) and intermediates, and metabolites such as aldoximes and nitriles. In addition, nitriles are intermediates in the biosynthesis of the indole phytoalexin camalexin and the plant hormone IAA. While cyanogenic glycosides have been identified in many plant species within the pteridophytes, gymnosperms and angiosperms, the glucosinolates are found predominantly in the order Brassicales. However, a few species exist that contain both classes of compounds, such as *Carica papaya* (Caricaceae) and *Alliaria petiolata* (Brassicaceae).

A current hypothesis implies that the glucosinolate pathway has recruited the genes from the biosynthesis of cyanogenic glycosides, because glucosinolates are evolutionarily younger. The acquisition of such a pathway could be important for detoxification reactions. A mutated cyanogenic glycoside biosynthesis could have produced toxic intermediates that were then metabolized by enzymes; this later evolved into the glucosinolate pathway. Other genes of the glucosinolate pathway may have been taken from basic biosynthetic reactions, for example the C–S lyase and glucosyltransferases (Figure 1.12). Evidence for a common evolution of both metabolic pathways has been found in *A. petiola*, where both the hydroxynitrile glucoside alliarinoside and the glucosinolate 2-propenylglucosinolate are synthesized from the precursor homomethionine. The first step to alliarinoside is catalyzed by a cytochrome P450-dependent monooxygenase, which was recruited for this pathway to metabolize the oxime intermediate. The glucosinolate and cyanogenic glycoside biosynthesis also involves cytochrome P450-dependent reactions.

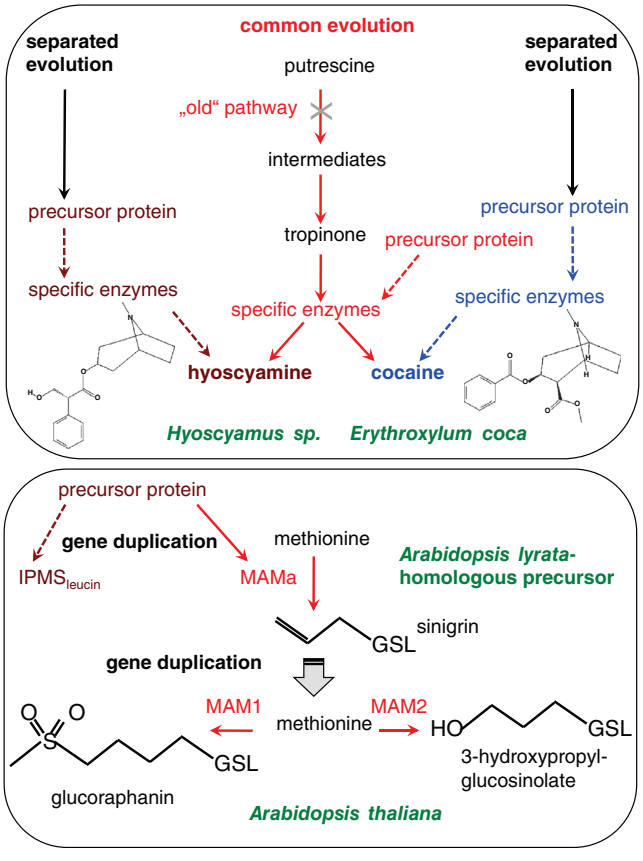


Figure 1.23 Several hypotheses for the evolution of natural substances biosynthetic pathways, using tropane alkaloids (upper panel) and aliphatic glucosinolates (lower panel) as examples. For tropane alkaloids two different possibilities for the evolution of the biosynthetic pathways are shown. The common pathway descends from one precursor, which is converted by homologous proteins to the next intermediate. The enzymes were recruited from one precursor protein. Pathway marked with a light gray cross symbolizes that the "old" pathway could have been lost. For later steps, separate evolutionary events recruit specific enzymes from different precursor proteins. For glucosinolates a hypothesis for the evolution of the biosynthesis of aliphatic glucosinolates in *Arabidopsis* species is shown. The MAM-enzymes catalyze the chain elongation reaction of methionine, whereas the other enzymes are needed for the modification of side chains. They could be derived by gene duplication of *MAMa*. In the glucosinolate structures only the respective side chains are shown. GSL = glucosinolate moiety; IPMS$_{Leucin}$ = alpha-isopropylmalate synthase gene; MAMa = methylthioalkylmalate synthase gene. Chemical structures in the upper panel are from The PubChem Project.

The further metabolic steps to yield alliarinoside share enzymatic activities with the 2-propenylglucosinolate synthesis. The hydroxynitrile may also function as a precursor for the cyanogenic glycosides in *A. petiola*, which are responsible for the liberation of cyanide. By changing the sequences of different genes encoding

for cytochrome P450-dependent monooxygenases during evolution, this complex network has been altered.

The synthetic pathways for alkaloids have some evolutionarily conserved features, and some special modifications in certain groups of plants. In those pathways with a common evolution, the "old" parts may be lost whereas the "new" pathways are present and eventually develop further (Figure 1.23). In contrast to the examples explained above, a separate evolution for tropane biosynthesis in Solanaceae and Erythroxylaceae has been established, based on an analysis of the reduction of tropinone. Importantly, the enzyme occurs in the leaves of *Erythroxylum coca*, but not in the roots, whereas the root system is the major location of tropane alkaloid biosynthesis in Solanaceae. Moreover, the enzymes have evolved from different protein precursors, even though they catalyze the same reaction in both plant families. This is most likely also true for the other biosynthetic steps leading to tropane alkaloids. In consequence, this biosynthetic pathway has been invented at least twice independently during evolution, but its evolution has occurred most likely more often because other, unrelated, plant families also contain these compounds.

1.2.2 Gene Duplication

Gene duplication is one of the most important hallmarks for the evolution of secondary metabolite pathways. In many cases, a gene from primary metabolism has been duplicated and then evolved into an enzyme which is involved in secondary metabolism (see Figure 1.25). In the case of gene duplication events, the two genes are quite often found in a tandem array on the chromosome, indicating rather recent events. However, the gene can also be translocated to different parts of the genome.

Pyrrolizidine alkaloids are mainly found in plant families such as Asteraceae, Boraginaceae and Fabaceae, where they occur in only a few genera. The biosynthesis of pyrrolizidine alkaloids has most likely evolved from the polyamine biosynthetic pathway. The enzyme homospermidine synthase, which catalyzes the first committed step to pyrrolizidine alkaloids, has most likely been developed by gene duplication from deoxyhypusin synthase, a protein that is essential to the plant. This enzyme is involved in the post-translational activation of the eukaryotic initiation factor 5A. This activation is, in turn, essential for cell growth and division, and was also related to seed germination, senescence, and apoptosis.

The duplication of a gene from primary metabolism is most likely also the basis of the biosynthesis of the indole derivative DIMBOA in maize (see Section 1.1.2). The alpha-subunit of tryptophan synthase was recruited into this other pathway to catalyze the synthesis of a different indole derivative (see Figure 1.25). Similarly, genes described for the synthesis of secondary metabolites from *Arabidopsis* and rice (Section 1.2.3) have been recruited by gene duplication from primary metabolism; in one case this was essential sterols, and in the other case the growth-promoting plant hormones, gibberellins. Genes for chain elongation reactions in aliphatic glucosinolate synthesis are possibly derived from the precursor protein methylthioalkylmalate synthase (MAMa) by gene duplication, after the gene had

been recruited from primary metabolism. The *MAMa* gene evolved from primary metabolism itself, by the duplication of a gene that was important for biosynthesis of the amino acid leucine, alpha-isopropylmalate synthase (Figure 1.23). The genes derived from *MAMa* are now termed *MAM1* and *MAM2*. Originally, these might have evolved in *Arabidopsis lyrata*, which is a close relative of *Arabidopsis thaliana*, and where a *MAMa* homologous sequence is still present.

The metabolism of glucosinolates and cyanogenic glycosides depends on the deglycosylation by glucosidases and subsequently, in some cases, further detoxification by nitrilases. In the case of glucosinolates, the product of the nitrilase reaction is a less-toxic carboxylic acid. In addition, indole glucosinolates may serve as precursors for biosynthesis of the growth-promoting hormone IAA (see Section 2.2.1). Other carboxylic acids could be further metabolized to yield different products. In *Arabidopsis*, a nitrilase activity is also involved in the synthesis of the phytoalexin camalexin (Section 1.1.2).

The metabolism of cyanogenic glycosides requires the presence of nitrilases for two distinct steps. A hydroxynitrile lyase is involved in the formation of cyanide, whereas a cyanoalanine nitrilase is necessary for the detoxification (Figure 1.12). In addition, a third group of enzymes – the nitrile hydratases – forms an amide group from the nitrile, catalyzing an alternative reaction.

In Brassicaceae, a large group of nitrilases can be found with different substrate preferences, which are dependent on the number and characteristics of the glucosinolates present in a particular species. In *Arabidopsis*, four nitrilase genes are known (Figure 1.24); three of these genes (nitrilases 1–3) are found in a tandem array on one chromosome, indicating a close relationship among these sequences and most likely gene duplication events, whereas the fourth gene (nitrilase 4) is located on a different chromosome. The latter is characterized by significant sequence differences, and the enzyme is more distributed ubiquitously among plant species, whereas the nitrilases 1 to 3 are specific for specific *Brassica* species. Nitrilases 1 to 3 are involved in the metabolism of glucosinolates, whereas nitrilase 4 catalyzes the detoxification of cyanide by converting the intermediate β-cyanoalanine to asparagine, aspartic acid, and ammonium.

Another type of nitrilase 4 was identified in *Sorghum bicolor*, and shown to catalyze the metabolism of the cyanogenic glycoside dhurrin, without the synthesis of toxic cyanide. However, other reactions exist in plants that also produce cyanide as a toxic byproduct; an example of this reaction is the biosynthesis of the plant hormone ethylene. Therefore, the detoxification of cyanide via nitrilase 4 is widespread among different systematically but unrelated plant groups such as mosses, moss ferns, gymnosperms and, of course, angiosperms, this being proof of their importance in metabolism (Figure 1.24). In grasses, gene duplication events have yielded another group of nitrilase genes (nitrilase 4B), which have then further evolved in species producing cyanogenic glycoside to groups B1 and B2. Interestingly, some of the encoded proteins can only show enzymatic activities as heterodimers (B group), whereas others are also active as homomers (nitrilase A). This results in a larger number of combinations for the resulting enzymatic complexes, which can in turn yield different substrate specificities.

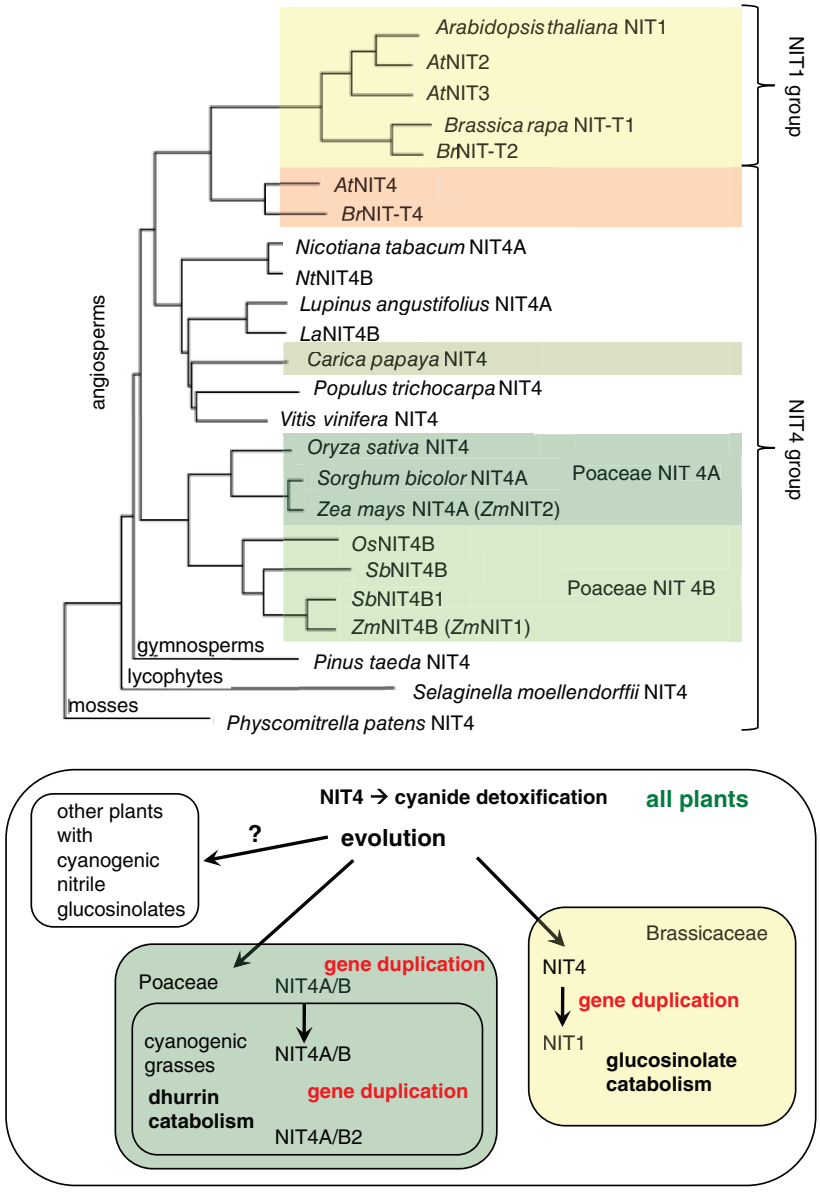


Figure 1.24 Evolution of nitrilase genes. Top: Part of a phylogenetic tree for nitrilase sequences. The nitrilase 1 family is typical for Brassicaceae and related plant species, whereas nitrilase 4 sequences can be also found in lower land plants. In Brassicaceae the nitrilase 1 family is most likely a descendant from nitrilase 4. The nitrilase 4 genes in grasses have also evolved from other nitrilase 4 genes. Yellow = nitrilase 1–3 from Brassicaceae; orange = nitrilase 4 from Brassicaceae; green = nitrilase 4 from grasses and from *Carica papaya*, a glucosinolate and cyanogenic glycoside containing plant species. Bottom: Evolutionary gene duplication events yielding the different nitrilase groups. Modified from Piotrowski, M. (2008) *Phytochemistry*, **69**, 2655–2667.

Gene duplications can also be detected in other phylogenetic trees for other synthetic pathways. An example of this is terpene synthase (Table 1.6), which catalyzes the synthesis of terpenes from different precursor moieties (see Figure 1.25). As each plant species can only synthesize a typical range of compounds, a specialization of the enzymes must be postulated. The sequence information of whole genomes allows the comparison of synthetic pathways in selected species. Whereas only one gene for a terpene synthase was identified in the moss *Physcomitrella patens* (this is actually a bifunctional copalyl synthase/kaurene synthase capable of forming *ent*-kaurene), the moss fern *Selaginella moellendorffii* has already 14 genes, which most likely encode full-length proteins based on their length. In other genomes the number is much higher, with between 40 and 152 terpene synthase genes having been identified, though not all are functional (Table 1.6). Many proteins have only one function – either copalyl synthase or kaurene synthase – and so can catalyze only one reaction step. In gymnosperms, bifunctional enzymes can still be found among the diterpene synthases, whereas this trait is lost in angiosperms. In contrast, monoterpene synthases contain only one functional domain in all of the plant species analyzed to date. Although terpene synthase genes can be classified into seven subgroups according to their sequence homologies, most plant species contain only one or two subgroups. An evolutionary plasticity within the terpene synthase families is shown by closely related enzymes, though these differ in their substrate specificities and also their intracellular localization. Differences in substrate preferences are also due to gene duplication and further evolution of the respective enzymes.

Table 1.6 Number of terpene synthase genes in different organisms and assignment to seven subgroups according to their sequence homologies. The number of individual genes represent those which are most likely full length and thus functional.

Plant species	Systematic classification	No. of terpene synthase genes	No. of related terpene synthase families in one organism
Physcomitrella patens	Moss	1	1
Selaginella moellendorffiii	Moss fern	14	3
Vitis vinifera	Dicotyledonous angiosperms	69	5
Populus trichocarpa	Dicotyledonous angiosperms	32	5
Arabidopsis thaliana	Dicotyledonous angiosperms	32	5
Oryza sativa	Monocotyledonous angiosperms	34	4
Sorghum bicolor	Monocotyledonous angiosperms	24	5

1.2.3 Gene Clusters

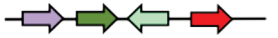
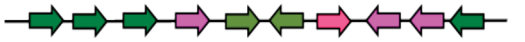
While gene clusters are common among bacteria (operons) and fungi for the genes encoding secondary metabolism enzymes, they have recently also been discovered in plants. Normally, many genes that mediate metabolic pathways in plants do not cluster (i.e., they occur in the form of continuous stretches on the DNA), but several examples have now been identified with either full or partial clusters for a specific metabolic pathway (Table 1.7). In the latter case, a cluster consists only of a few genes which are sometimes disrupted by genes not belonging to the cluster. These partial gene clusters can provide information on their formation during evolution, because they could have been formed by the transposition of large DNA pieces. Most likely, however, these events are preceded by gene duplication.

Several examples of gene clusters can be found for genes coding for enzymes in the synthetic pathways of secondary metabolites, which are active against phytopathogens (Figure 1.25). Among these is the gene cluster directing the synthesis of avenacin, a bioactive compound that is produced by some grasses against a phytopathogenic fungus (Section 2.5). For this gene cluster the rearrangement of genes has been postulated, because each gene in the cluster still has its own promoter and hence is not of the operon-type structure. To date, exactly how genes can rearrange is not clear, but possibilities include gene duplication, the relocation of individual genes, or the rearrangement of larger parts of the genome by defragmentation. These mechanisms could possibly lead to novel functions of relocated genes in the new cluster. Gene clusters need also to be simultaneously transcriptionally controlled, so that genes within one pathway are coexpressed.

Other examples of gene clusters in secondary metabolite synthesis include the clusters in *Arabidopsis* for thalianol, in maize for DIMBOA, and in rice for phytocassane and momilactone (Figure 1.25). A gene cluster normally contains genes that encode enzymes for synthesis of the core structure as well as side-chain modifications, which introduce specificity to the molecule (Table 1.7). The genes in a cluster are not necessarily in the same order in which they function in the pathway. For example, the cluster for thalianol in *A. thaliana* starts with a gene for a modifying enzyme, while the terpene synthases are located at the end. In the close relative *A. lyrata*, the gene cluster already has a different structure, such that the acyltransferase is strongly modified and a 10 kilobase insertion can be found between the acyltransferase gene and the *THAD* gene. The reading frames can also be changed for individual genes within the cluster.

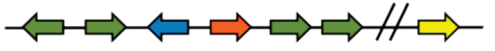
Gene clusters can be rather large, as for example the cluster for phytocassane in rice, which consists of ten genes. While most gene clusters contain arrays of genes without any interruption, the cluster for DIMBOA is discontinuous and contains insertions of many other genes. A gene for DIMBOA synthesis is even located on a different chromosome, which means that not all genes for a given pathway are present in the cluster. The gene clusters for the metabolites DIMBOA, thalianol and avenacin are expressed during development of the healthy plant, while synthesis of the diterpenes momilactone and phytocassane can be induced by elicitors

Table 1.7 Gene clusters in the biosynthesis of secondary metabolites in plants. Six secondary metabolite gene clusters, which have been described in detail for plants, are presented. The size of the cluster (in kilobases), the number and function of genes and the role within core structure or side chain modifications are given. The genes are shown in their order (indicated by the small numbers) on the chromosome. The cluster sizes are not to scale. The asterisks mark genes that are not transcriptionally regulated. Modified from: Chu, H.Y. *et al.* (2011) *Plant J.*, **66**, 66–79; and Winzer, T. *et al.* (2012) *Science*, **336**, 1704–1708.

Plant	Substance	Reading frame and size of clusters (kb)	Enzymes for core structure	Enzymes for modifications
Arabidopsis thaliana	Thalianol	 35	Triterpene synthase[4]	BAHD Acyltransferase[1] CYP76[2] CYP85[3]
Avena strigosa	Avenacin	 202	Triterpene synthase[4]	Methyl-transferase[1] Sugar transferase[2] Serine carboxy-peptidase like acyltransferase[3] CYP51[5]
Oryza sativa	Momilactone	 170	Class II Diterpene synthase[1] Class I Diterpene synthase[4]	CYP99[2] Dehydrogenase[3] CYP99[5]
Oryza sativa	Phytocassane	 245 * * * *	Class I Diterpene synthase[4] Class II Diterpene synthase[7] Class I Diterpene synthase[8] Class I Diterpene synthase[9]	CYP76[1] CYP76[2] CYP76[3] CYP71[5] CYP71[6] CYP76[10]

(*continued*)

Table 1.7 (*Continued*)

Plant	Substance	Reading frame and size of clusters (kb)	Enzymes for core structure	Enzymes for modifications
Papaver somniferum	Noscapine	 214		Dehydrogenase[1] Carboxyl-esterase[2] CYP82[3] CYP82[4] Acetyl-transferase[5] Methyl-transferase[6] CYP82[7] Methyl-transferase[8] CYP719[9] Methyl-transferase[10]
Zea mays	DIMBOA	 15 618	Tryptophan-synthase alpha homolog[4]	CYP71[1] CYP71[2] Sugar transferase[3] CYP71[5] CYP71[6] Methyl-transferase[7]

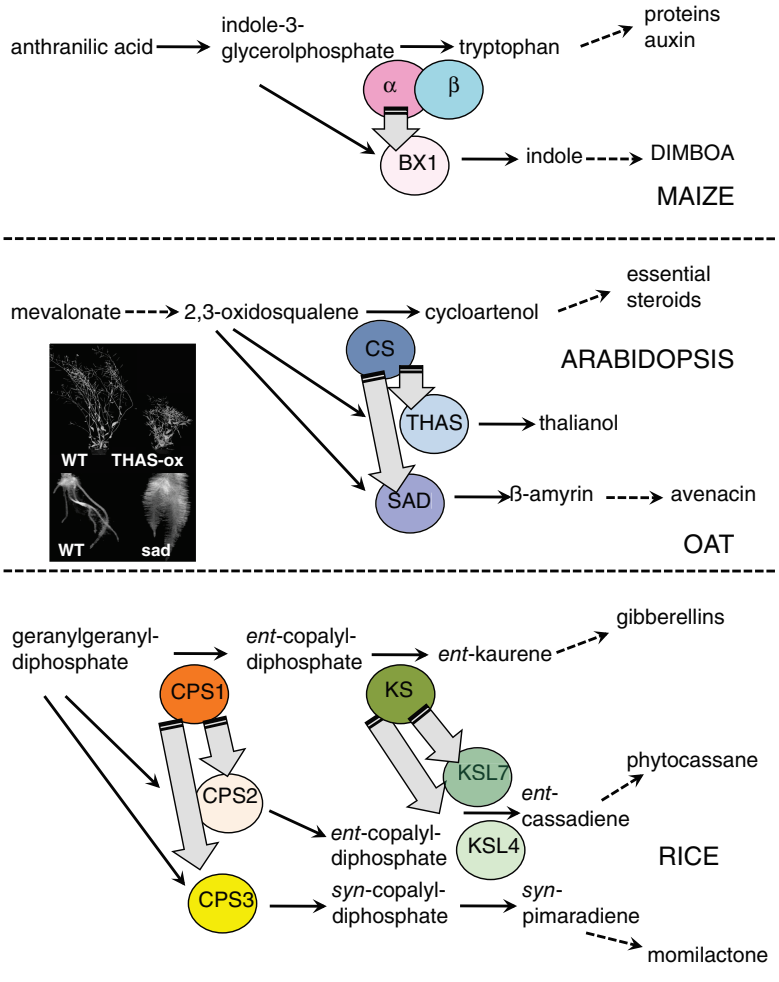


Figure 1.25 A hypothesis for the evolution of core genes within gene clusters involved in the formation of some secondary metabolites. The genes were recruited into the cluster from genes involved in primary metabolism or plant hormone synthesis genes. Top: Evolution of DIMBOA (= 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one) biosynthesis from the shikimate pathway in maize (*Zea mays*). Middle: Evolution of thalianol biosynthesis in *Arabidopsis thaliana* and of avenacin in oat (*Avena sativa*) from sterol biosynthesis. Bottom: Evolution of diterpene biosynthesis from the pathway to gibberellins in rice (*Oryza sativa*). BX = benzoxazin synthase; THAS = thalianol synthase; SAD = β-amyrin synthase; CS = cycloartenol synthase; CPS = class II- and KSL = class I-diterpene synthase. The pictures show *Arabidopsis thaliana* overexpressing the *THAD* gene (THAS-ox) and a mutant of oat for avenacin synthesis (*sad*), which accumulates a toxic intermediate, compared to the respective wild-type tissues. Figure based on Chu, H.Y. *et al.* (2011) *Plant J.*, **66**, 66–79. Photographs from: Osbourn, A.E. (2010) *Plant Physiol.*, **154**, 531–535; see Further Reading). Reprinted with permission from American Society of Plant Biologists.

and pathogens. Although many genes within the phytocassane cluster are coregulated, four of them (indicated by asterisks below the arrow representing the respective gene) are not regulated together with the others (Table 1.7).

A cluster containing only genes for modifying enzymes has been isolated from the opium poppy (*Papaver somniferum*), whereas the genes that encode the proteins involved in the core structure of benzylisoquinolines have been found outside the cluster. The genes in the cluster are involved in synthesis of the alkaloid noscapin; this cluster was identified by transcriptome analyses of noscapin-producing and nonproducing plant lines with ten genes that showed a high genetic coupling in a crossing population and were present only in the noscapin-containing plants. The gene cluster was verified by using bacterial artificial chromosomes containing the whole cluster. The compound noscapin is of great medical interest due to its anticancer potential that results most likely from its interaction with tubulin. Noscapin can also arrest metaphase and induce apoptosis, and also has a cough-suppressant effect (similar to codeine). However, as with thalianol, the biological function of noscapin in plants is currently unknown.

Partial gene clusters can be found encoding for proteins involved in the biosynthesis of the cyanogenic glycosides linamarin and dhurrin. For this pathway, the phenomenon of metabolic channeling has been described, where products are not released but rather are directly reused as substrates for the next reaction step. This phenomenon might also be associated with clustering of the respective genes.

What, then, could be the advantage of gene clusters? One possible explanation is that within gene clusters the chances of joint inheritance of the trait are higher. Also, when gene clusters result in metabolic channeling, toxic intermediates can be avoided. In addition, coordinated gene expression could avoid the creation of toxic metabolites, because the genes encoding metabolic enzymes for detoxification are expressed at the same time. Finally, during plant defense gene clustering may be advantageous when the genes are transcriptionally regulated at the same time, and by the same signals. An example of such a coordinated gene regulation is the synthesis of momilactone, where the genes in the cluster can all be induced by a chitosan elicitor. Only one transcription factor, from the family of leucine zippers, is responsible for this transcriptional activation. However, it is not known whether this transcription factor is responsible only for this gene cluster, or whether it is also involved in the regulation of other genes. In oat (*Avena sativa*), the gene cluster for avenacin synthesis was shown to be localized specifically in the outer epidermal cells of the root tips by using high-resolution DNA fluorescence *in situ* hybridization (FISH). A tissue-specific chromatin decondensation was also shown to be associated with the expression of the gene cluster.

Clustered genes which encode proteins involved in the synthesis of the core structure of a substance are often derived from genes in primary metabolism, and which have evolved after gene duplication (Figure 1.25). The protein BX1, which is involved in the synthesis of DIMBOA, was recruited from tryptophan synthesis, by duplication of the gene for the alpha-subunit of tryptophan synthase. Tryptophan is a proteinogenic amino acid but is also a precursor for many other secondary metabolites, such as terpene indole alkaloids and the plant hormone IAA.

Genes for thalianol synthase in *Arabidopsis* and β-amyrin synthase in oat evolved independently from the pathways to essential steroids in both plant species by recruiting cycloartenol synthase. The proteins for the synthesis of momilactone and phytocassane in rice have evolved from genes encoding diterpene synthases involved in the synthesis of the plant hormone gibberellin. In *Arabidopsis*, overexpression of the *THAD* gene, which encodes thalianol synthase THAS, results in a dwarfed growth and demonstrates that thalianol itself is also toxic to the plant producing it. Mutation of the *SAD* gene in oat results in a phenotype with fewer lateral roots, due to the accumulation of a toxic intermediate.

Whilst the question regarding the evolution of gene clusters in plants remains only partially solved, novel information can be acquired by using fungi as a model. Following a gene duplication event, the gene can relocate in the genome into a cluster. At the point of origin, and if the gene in the cluster is a single copy, the gene can be either retained or lost. Otherwise, it should be possible to find copies elsewhere in the genome. The relocation of a gene into a cluster is an advantage in meiosis, where closely linked genes are most likely inherited together in the progeny. This implies that these events should occur more often in regions with low recombination rates on the chromosome, which would stabilize such a cluster. These events may be advantageous for an organism during changing environmental conditions, when a greater evolutionary pressure would be exerted on the development of novel pathways. Such conditions might include nutrition as well as abiotic and biotic stress factors. This hypothesis is supported by the observation that some of the metabolites synthesized by enzymes encoded in gene clusters (i.e., DIMBOA, avenacin, momilactone and phytocassane) are involved in plant defense reactions against pathogens and insects.

Among the different plant species, gene clusters can either develop further or be retained in the original stages. For example, the biosynthesis cluster for DIMBOA in maize is quite well developed in the sense that it contains all of the necessary genes for the complete biosynthesis of DIMBOA, whereas the cluster in wheat (*Triticum*) does not contain all genes necessary for a complete biosynthesis. These observations can also help to provide an understanding of the evolution of a single pathway in different organisms within one order, for example, grasses. In maize, a noncomplete cluster could have been the precursor, and this was then further developed to the complete cluster. Alternatively, a complete cluster could have been the precursor, so that secondary event genes could have been lost in one species. In this latter case, the biosynthetic pathway would need to recruit other gene loci for the full biosynthesis of the respective metabolite or, as in rye (*Secale*), the biosynthetic pathway will be completely lost, as indicated by the synthesis of an alternative indole compound instead of DIMBOA.

As complete genome sequences for different organisms become available, bioinformatic methods can be applied to locate further putative gene clusters for secondary metabolite synthesis. In addition to the known gene clusters, five further putative clusters were predicted in *Arabidopsis* by computer-based methods. These have been postulated on the basis of homologous genes encoding cytochrome P450-dependent enzymes in *Brassica*, even though their specific function in *Arabidopsis* is not yet known.

1.2.4 Natural Variation

Many examples of natural variations in secondary metabolites have been described, an example being nicotine in tobacco (see Section 3.1 and Figure 3.12). At this point, the concept will be explained by using the examples of flavonoids and the glucosinolates.

The glucosinolates are a group of compounds which demonstrate a high variability among different plant species within the Brassicaceae, and also within cultivars and ecotypes (these are also termed "accessions" for *A. thaliana*). They can be used as examples to study the genetic variation, the aim being to understand their evolution. *Arabidopsis thaliana* is used as a model plant to investigate natural variations, because of the availability of hundreds of ecotypes. Here, the differences are used to identify quantitative trait loci (QTLs) to locate novel traits associated with differences in glucosinolate patterns, and which can encode either genes for either biosynthetic and metabolic enzymes or regulatory proteins – that is, transcription factors. Evolutionary pathways can also be investigated using this approach. The MAM enzymes (see Figure 1.23) are good examples as they have become diversified in evolution after gene duplication, and could be allocated to specific glucosinolate patterns because they can now use different substrates. However, there are some ecotypes of *Arabidopsis* which do not contain all three known *MAM* genes, and these will be deficient in the synthesis of several aliphatic glucosinolates. These different patterns are termed "natural variations" between accessions. For indole glucosinolates, the variation has been detected in relation to the side-chain modifications rather than to the core structures. This is a result of the differences between genes for modifying enzymes such as hydroxylases and methyltransferases.

One important metabolic reaction is removal of the sugar moiety of glycosinolates by myrosinase (see Figure 1.12), while further metabolic reactions depend on the specifier proteins that determine the outcome of the reaction. Both, epithionitrile and nitrile specifier proteins, as well as thiocyanate-forming proteins, direct the formation to respective products. In the case of nitrile the reaction product is less toxic than the two others, the epithionitriles and thiocyanates. The so-called myrosinase-binding proteins from *Brassica* also belong to this protein group, based on their sequence homology. A lectin-binding motif is necessary in order to bind to myrosinase, and this is common to all of these proteins. Lectins bind to specific sequence motifs via covalently bound sugar moieties of glycosylated proteins, of which myrosinase is a member. Epithionitrile- and nitrile-specifying proteins are found in *Arabidopsis*, while garden cress (*Lepidium sativum*) possesses a thiocyanate-forming protein. Consequently, these different species also exhibit different patterns of glucosinolate metabolites. As these compounds play a major role in defense for these plants, an accumulation of toxic metabolites is essential to provide their resistance against pathogens and insects. Thus, it is also important which of the specifier proteins has been developed in a given plant species. If only nitrile specifier proteins are present, then fewer toxic compounds will accumulate and this could have overall a negative effect on the defensive performance

against pathogens and herbivores. Consequently, the evolution of these proteins has an important implication for defense responses. Among the different *Arabidopsis* ecotypes, the metabolites of the glucosinolates occur in different concentrations and also in different proportions to one another, which indicates that different specifier proteins are present. Interestingly, the evolution of the specifying proteins – and especially of the nitrile specifiers – has also been demonstrated in specialized insects that feed on *Brassica* species, which are essential for their nutrition (Section 3.1).

A structural variation of flavonoids has also been found in different *Arabidopsis* ecotypes, and a comparison of these natural variations led to the identification of novel compounds and regulatory proteins for the synthesis of these flavonoids. Whilst regulatory elements are most likely involved if only quantitative differences are present in the patterns, a novel compound implies a new biosynthetic reaction. Flavonoids can be grouped differently based on their occurrence in different ecotypes that occur all over the world. One flavonoid group was seen to occur only in ecotypes from America and Asia while European ecotypes did not show any common pattern, although the numbers of samples taken was most likely insufficient to corroborate such a classification for flavonoids. Nevertheless, the results of this geographic distribution could point to evolutionary pathways of groups of flavonoids on a worldwide basis.

1.2.5
Modulation of Different Molecules with Similar Functions

During the evolution of secondary metabolites, the functional aspect of molecules with which they interact must also be taken into account. One such example is the inhibition of polar auxin transport by flavonoids (see Section 2.2.1; Figure 2.3). Polar auxin transport from cell to cell is regulated by efflux transporters of two different structural families, the PIN and PGP proteins (Figure 1.26), both of which can be interfered with by flavonoids. The PGP-type transporters belong to

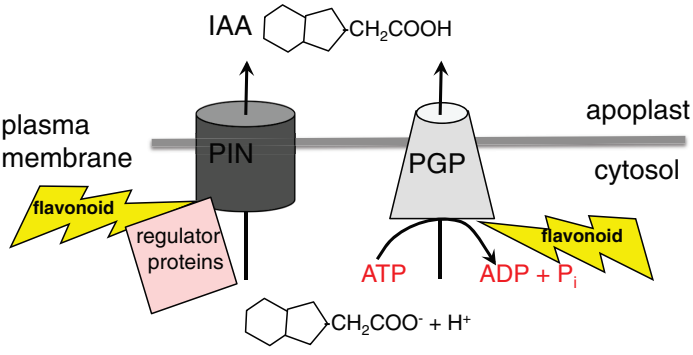


Figure 1.26 Flavonoids modulate two different types of auxin transporters, the evolutionary younger PIN-type and the older ABC-transporter PGP-type, which is dependent on ATP. The flash symbolizes possible interaction sites with flavonoids and transporters.

the ABC-transporter family, where flavonoids inhibit their ATPase activity. The same situation occurs in mammals (Section 4.1.1), where flavonoids have an inhibitory effect on the transporters, but in plants the effect is less pronounced, perhaps because the molecules have been able to adapt to the flavonoids over a longer time period. In contrast, the flavonoids can indirectly influence the PINs by affecting regulatory proteins such as kinases/phosphatases that regulate PIN activity. The evolutionarily younger PIN-transporters are less strongly inhibited by flavonoids than the PGP-family, which is indicative of their better adaptation to a plant environment during the evolution of their specific transport activity towards auxin. The modulation of auxin transport is used by bacterial symbionts (rhizobia) to establish auxin maxima at the site of nodule formation (see Section 2.4). Plants also use flavonoids to modulate their auxin transport capacities during developmental processes. Thus, auxin transport proteins and their regulators may be regarded as cellular components that have been adapted to a certain environment during evolution, rather than to act as specific target molecules for flavonoids.

In the model plant *Arabidopsis*, as a member of the higher plants, the PIN-family consists of eight different genes that are expressed in different tissues and localized intracellularly at the plasma membrane, with one exception. Recently, a PIN protein in the ER was described which was also a site of flavonoid synthesis. These proteins appear not to be involved in intercellular transport, but rather to control intracellular auxin homeostasis. Since in the moss *Physcomitrella patens* this specific PIN member is the only one present, it was assumed that the endoplasmic PIN proteins are evolutionarily older than those found at the plasma membrane. It has been assumed that the earliest function of the PIN proteins was to regulate intracellular IAA levels, and that cell-to-cell transport was evolutionarily younger. Thus, an attractive hypothesis might be that polar auxin transport is a hallmark of the development of higher land plants with more complicated body plans, whereas the control of cellular auxin levels was already necessary in lower land plants.

1.2.6 Evolution by Stress Factors

1.2.6.1 Abiotic Stress

Another possible approach to investigate the evolution of secondary metabolites uses flavonoids as examples. Based on their UV-absorbing properties, flavonoids were described as the "sunprotection screen" of plants. Their absorption maxima are within the wavelengths of hazardous UV light, by which the expression of genes encoding enzymes for biosynthetic steps are also induced (see Section 2.3). Although, in *Arabidopsis*, UV-B light induces quercetin synthesis, there are indications that flavonoids are not the primary protection compounds, despite their high induction under light stress. Rather, they might have an indirect effect as antioxidants against reactive oxygen species (ROS) produced during high light stress. Other candidates providing protection against hazardous light include sinapine derivatives.

In the lower land plants that occurred at the early stages of evolution, a mycosporin-like amino acid metabolism was lost in favor of flavonol metabolism, despite the mycosporins having better effects on leaf surfaces in terms of protection against light stress. Thus, it was hypothesized that flavonoids could have evolved their UV-protective function as a secondary trait from other roles in the plant. Interestingly, most genes encoding flavonoid biosynthetic enzymes are already present in lower land plants, such as liverworts and mosses. One characteristic of these "old" genes is a strong and rapid induction after light stress and other abiotic stress factors, which typically leads to the formation of ROS. In addition, transcription factors of the R2R3-MYB type (Figure 1.17) are present in mosses and can be induced by a high light irradiation. Observations that the flavonoid pathway had been relatively well conserved over many million years indicated a central role for the evolution of these secondary metabolites, which play diverse roles in the stress adaptation of plants. Such a general function may be very important when considering unpredictable changes in environmental conditions, and organisms with a wide range of metabolic compounds might be better protected against such variations in the environment. Indeed, flavonoids may be considered a class of plant metabolites that have conferred the ability to adapt to changing environments over long periods of time.

The role of flavonoids is in accordance with the hypothesis concerning the regulation of auxin transport. Quercetin is a better inhibitor of polar auxin transport than kaempferol, and therefore UV-B induced quercetin synthesis should have an indirect effect on plant growth and development, as the distribution of a growth-promoting hormone would be changed. Likewise, stressed plants could react with a different habitus to changing environmental conditions.

1.2.6.2 **Biotic Factors**

The interaction of plants with their environment, especially with changing pathogen and herbivore populations, necessitates the synthesis of different secondary metabolites (see Section 3.1). The development of novel defense compounds during evolution is dependent on changes in genes, gene function, or a combination of different proteins to result in novel pathways. One important mechanism in this respect is gene duplication (see Section 1.2.2). It is disadvantageous for a plant to focus on only a few metabolite classes, although this would result in a highly adaptive defense behavior in relation to any parasites invading the plant. The energy costs to synthesize only a few metabolites would also be much lower, and this again would be advantageous. During evolution, however, the parasites may also adapt to the chemical defense of the plants, and consequently a plant with only a few different metabolites would be less well protected against pests that were capable of adapting. Clearly, a wider variation of secondary metabolites would be advantageous within the natural environment, and the evolution of a wider spectrum of compounds by one plant species would be necessary if a plant were to be attacked by a range of parasites at different stages of its development. Thus, a broad series of defense reactions would hamper the adaptation of a parasite.

1.3 Biotechnological Applications

1.3.1 Identification and Quantification of Natural Plant Products

In order to identify or quantify a compound of interest, the metabolite must first be extracted from the plant tissues. However, the chemical properties of a material under investigation is of great importance in the development of a relevant purification scheme (Figure 1.27). The most important issues to be taken into account include:

1) It must be defined whether a compound that is already known should be extracted and quantified (targeted approach), or whether a broad range of unknown (bioactive) compounds should be identified (untargeted approach). The latter experimental approach is also coined "metabolomics" (Figure 1.28).
2) For individual compounds, it must be determined which properties are already known, and which solvents can be used for their extraction.
3) The purity of the compound might be important for identification and also for bioactivity assays; in this situation the metabolite must be further purified using chromatographic methods (Figure 1.27).
4) If identification and quantification is the only purpose then the sample will not need to be recycled. If it should be subjected to further assays for bioactivity, a larger amount of the compound(s) will be needed after sample purification.

The first decisions to be made involve the extraction solvent. It is important to know whether polar or unpolar compounds are to be extracted, as different pH values can be used to further differentiate compounds according to their molecular properties. Following the evaporation of organic solvents, the remaining aqueous phases can be further extracted using organic solvents with different polarities. At this point the extract will still contain many molecules, and these must be further enriched by applying chromatographic methods. The complex extract can be subjected to different chromatographies, depending on the molecule(s) to be isolated. Such methods include size fractionation, separation according to charge/polarity, or affinity chromatography. The chromatographic steps can be analytical, using small volumes, or preparative for the purification of larger samples. If the chromatographic methods are used together with a suitable detector (i.e., for high-performance liquid chromatography; HPLC), then a certain amount of information, perhaps about the absorption properties of the molecules, can be obtained at the early purification steps. For affinity chromatography, additional information must be gathered about the compound; in this case, ligands that bind to the molecule in question can be attached to a resin and will retain a small group of compounds.

The analytical methods applied to secondary metabolites include paper chromatography (this is barely used any more), thin-layer chromatography (TLC), HPLC, gas

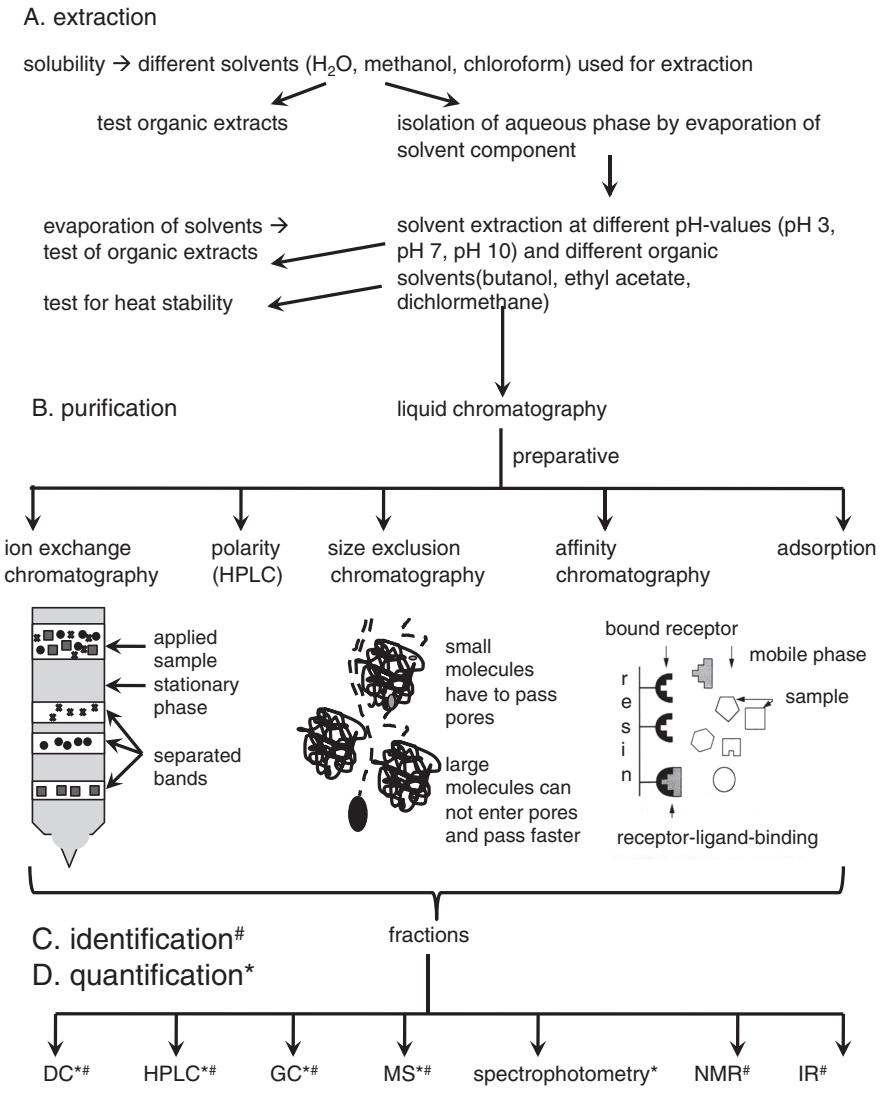


Figure 1.27 Example of a scheme for the extraction, purification and possible identification of complex metabolite mixtures from plants. TLC = thin-layer chromatography; HPLC = high-performance liquid chromatography; GC = gas chromatography; MS = mass spectrometry; NMR = nuclear magnetic resonance; IR = infrared.

chromatography (GC), mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, infrared (IR) spectroscopy, and spectrophotometry (Figure 1.27). Some of these methods can be combined, such as GC or liquid chromatography (LC) with MS as detector. Some methods also allow the quantification of metabolites as well as

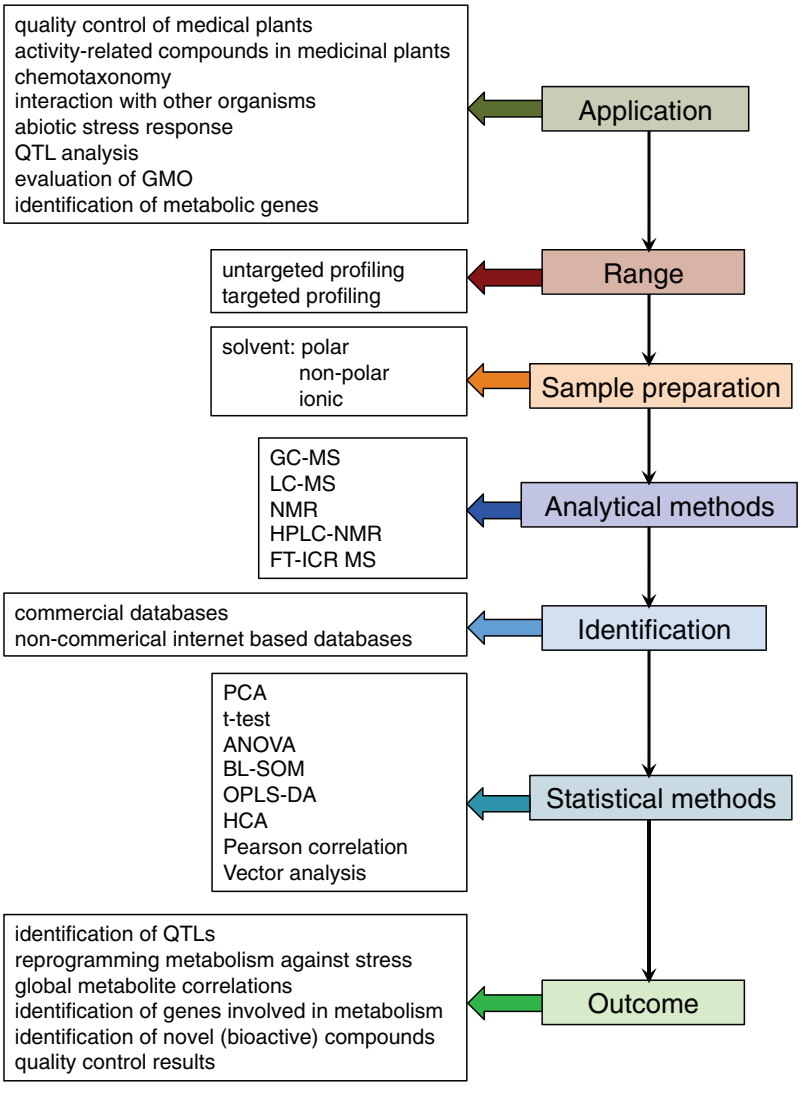


Figure 1.28 Metabolomics approaches in plant research, especially in biotechnology. Plant-specific components for the expectations and outcome are shown in green; possible experimental conditions are shown in brown; analytical conditions are shown in blue.

their identification. In particular, photometric assays with specific reagents can yield rapid and reliable results, and can also be used occasionally in combination with TLC (Figure 1.30). By using reference compounds, a tentative identification is possible with this technique in some cases. The chromatograms obtained can be scanned, and specific spots subsequently quantified. Comparisons with standards when using

TLC involves using the R_f value; this is a factor that indicates the relative migration of a given compound (for HPLC and GC, the retention times are used in similar fashion). If a reference substance is available, then standard calibration curves can be used to quantify a compound. If MS is used as a detection system for either LC or GC, identification via the mass spectrum should be possible. A quantification can also be achieved if reference compounds are available that have been labeled with heavy isotopes; these behave exactly as the natural compounds but, if added prior to the extraction, they can be used to determine the efficiency of the extraction procedure for a specific compound. On the basis of the characteristic masses for standard and endogenous compound, concentrations of compounds in the cells or tissues can be calculated. This method is accurate even for small molecules, which occur in minimal quantities. The accurate identification of a compound is sometimes not possible using MS, but can often be achieved using NMR- or IR-spectroscopy.

Metabolomics – which is defined as the mostly unbiased parallel analysis of as many compounds as possible – has attracted attention relating not only to the identification of novel metabolites but also to biotechnological aspects (see Section 1.3.3). This approach can provide an insight into the changing metabolic pattern under various conditions during the development, as well as the changing environmental conditions under abiotic and biotic stress situations, and thus help to identify novel compounds (Figure 1.28). By using modern biotechnology approaches, plants with different biological properties can be compared to identify additional bioactive substances, for example in medicinal plants (Figure 1.31). Other applications include the quality control of medicinal plants, the identification of QTLs, chemotaxonomy, the evaluation of genetically modified organisms (GMOs), and the identification of metabolic genes.

The choice whether a targeted or an untargeted approach is chosen determines the suitable solvent(s) used for the extraction. It should be noted that the perfect universal solvent for all applications does not exist, and that the solvent could be polar, nonpolar, or ionic. The following step determines the intensity, reliability and potential for the identification of novel compounds, and the analytical techniques to be used for the analysis of the compounds. While the pioneering metabolomics studies were performed using GC-MS, this method has now been complemented or replaced by other techniques such as LC-MS, NMR-spectroscopy, or combinations of these (Figure 1.28).

The identification of individual compounds within chromatograms obtained using either GC or LC relies on deconvolution software, which is needed to separate the peaks from one another, as well as databases that may be either open source internet-based or commercially available. Finally, the data need to be evaluated (Figure 1.28). Various statistical methods used in conventional genetic studies are applicable to metabolomic data by considering the amount of each metabolite as a trait value. For this, principal component analysis (PCA), which is a multivariate analysis method, is commonly used in metabolomic studies. The PCA model can provide an overview of all observations or samples in a data table by projecting and clustering each sample and highlighting any holistic differences

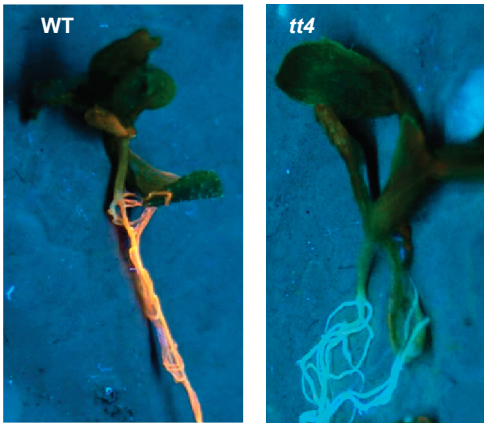


Figure 1.29 ***In situ* staining of flavonoids in *Arabidopsis thaliana* with diphenylboric acid-β-aminoethylester.** Seedlings on the left show orange fluorescence typical of quercetin and its derivatives, while the seedling on the right, mutated in chalcone synthase, *transparent testa 4* (*tt4*) is flavonoid-free and thus shows the cyan fluorescence indicating sinapates. WT = wild-type.

in a given sample set. Many reports have described the application of PCA to metabolomic data. In addition, several statistical analytical methods have been used to analyze metabolomic datasets, including hierarchical cluster analysis (HCA), partial least-squares-discriminant analysis (PLS-DA), and batch-learning self-organizing map analysis (BL-SOM).

Depending on the objective of each study, the most appropriate statistical analytical method should be exploited to evaluate the available metabolomic data. The outcome of such an experiment could then be the identification of QTLs, the reprogramming of metabolism against stress, global metabolite correlations, the identification of genes involved in metabolism, the identification of novel (bioactive) compounds, and quality control of the results.

In order to localize secondary metabolites within a plant tissue, a so-called *in situ* staining method can be used for some compounds (Figure 1.29). Many natural compounds, such as the flavonoids shown here, can be rendered visible under UV-light because they form fluorescing complexes with certain staining reagents. In the case of treatment with diphenylboric acid-β-aminoethylester (Figure 1.29; see also Figure 1.30), an orange fluorescence is typical for quercetin and its glycoside derivatives, while the cyan fluorescence in a flavonoid-free mutant of *Arabidopsis* reveals the presence of sinapates. In leaves, it is difficult to visualize the flavonoid complexes because of the high concentration of chlorophyll.

Laser microdissection can be used to obtain an insight into the more detailed distribution of different secondary metabolites. For this, small tissue sections can be generated or even single cells isolated; these are then collected and further

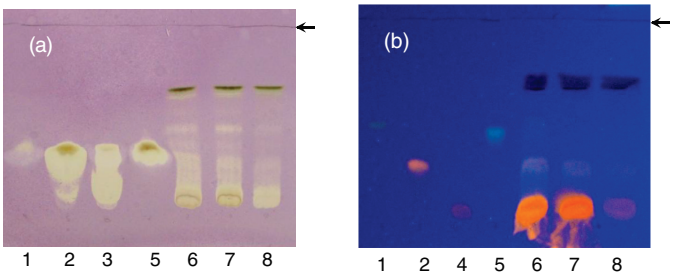


Figure 1.30 An example of "biochromatography" in comparison with fluorescence staining. Two silica gel TLC plates are compared. The upper plate was stained with a reagent to show antioxidative activity (2,2-diphenyl-1-picrylhydrazyl); the lower plate was stained using a reagent which forms fluorescing complexes with flavonoids and sinapates (diphenylboric acid-β-aminoethylester). The lower plate was photographed under UV-light. The solvent front has been marked with an arrow. The individual lanes show 1–5: standards of naringenin (1), quercetin (2), rosmarinic acid (3), rutin (4), and kaempferol (5); lanes 6–8: metabolites from a carnation plant, which were extracted using a methanol–water mixture. After evaporation of the organic solvent, the remaining aqueous phase was adjusted to different pH values and extracted using an organic solvent (lane 6: pH 3; lane 7: pH 7; lane 8: pH 10).

analyzed. Using this method, four different parts of mature seeds of oilseed rape (*Brassica napus*) were sampled to investigate the distribution of major secondary metabolites such as glucosinolates and sinapine, and of two flavonoids in the hypocotyl and radicle, the inner cotyledon, outer cotyledon, seed coat, and endosperm. Subsequently, these compounds were identified and quantified in dissected samples by using HPLC with diode array detection and MS. The differential localization of the secondary metabolites can also provide clues regarding their putative functions. MS can also be used on fresh tissue sections, and this results in an even higher spatial resolution of secondary metabolites. This technique has been used to demonstrate the distribution of glucosinolates in various *Arabidopsis* organs. This high-resolution "imaging" of secondary metabolites can also reveal information about the specific roles of certain tissues. For example, a high concentration of secondary metabolites at the leaf margin, which is among the first contact points between a leaf and an insect, could indicate a possible role against herbivores (see also Figure 2.35).

1.3.2 Biological Activity

Various test systems used for the determination of biological activities, together with the important targets, are described in Chapters 4 to 6. However, the possibility of conducting bioassays either together or immediately after a chromatographic separation should be mentioned at this point. Such test

systems should be simple to apply and cheap to operate, especially if many compounds are to be screened. Consequently, the systems are directed towards the simple properties of secondary metabolites, such as their antioxidative potential, which can easily be determined by using spetrophotometric tests with colored compounds. Alternatively, these staining methods can be used in combination with TLC as a so-called "biochromatography," revealing the chromatographic behavior of a compound together with an indication of its antioxidative activity and other analytical tests (Figure 1.30). In this case, the TLC plate has been stained (after development) with a colored radical, leaving white spots where compounds with antioxidative activities migrate (Figure 1.30a). TLC of the same extracts, after staining with a reagent that forms complexes with flavonoids and sinapates, is shown in Figure 1.30b. In combination, these two methods reveal important biochemical properties of the compounds in the analyzed extract.

1.3.3
Biotechnological Production of Natural Compounds

Bioactive compounds have long been used by humankind for many different purposes (Table 1.8), and the requirements for compounds with better or even novel properties is increasing continuously. The exploitation of natural compounds of plants can be traced back generally to their medicinal use, with hallucinogenic, mind-altering properties and analgesic effects having received much of attention (see also Chapters 5 and 6). Other compounds are used in foods as flavor or spice components; examples include the spice capsaicin from peppers (which has also antimicrobial activity), vanillin from vanilla, and also the isothiocyanates released from the glucosinolates upon tissue disruption. Finally, colored substances have received much attention in the food industry and as natural dyes in other applications.

Plants are considered to be the organisms of choice for the production of metabolites that rely for their synthesis on complex pathways. The realization of such complex biosynthetic pathways in microbes is very difficult, because of missing intermediates and potential problems with protein folding and modifications. Yeasts can be seen as an alternative, but in this case reactions that need different plant compartments cannot be imitated. Today, sterile cell or organ cultures in plants are among the methods of choice.

The compounds used by humans can be employed for dyes (e.g., betanidine, indigo, shikonine), as aroma volatiles and spices, as food additives (see Chapter 6), in the cosmetics industry as aroma compounds and antioxidants, or in pharmacology (e.g., pilocarpine, scopolamine) (Table 1.8; see also Chapter 5). In addition, their use in agriculture as biopesticides has received much attention (see Section 2.5.2).

Many plants with beneficial properties are either difficult to cultivate and/or are on the list of endangered species; therefore, their harvest is difficult or

Table 1.8 Applications of secondary metabolites in medicine, pharmacy, as food additives, and in cosmetics industry.

Compound class	Product	Effect
Aroma volatiles	Perfume	
Aroma compounds	Spice	Many
Natural dyes	Indigo	
	Betanidine	
	Shikonine	
Biopesticides[a)]	Glucosinolate	Fungicide
		Herbicide
		Antiviral
		Antimicrobial
		Insecticide
Stimulants		Stimulating, hallucinogen
Pharmaceutically active compounds	Phytotherapy[a)]	Anti-inflammatory
		Antibacterial
		Antidepressant
		Fungicide
		Antimigraine
		Antiparasitic
		Muscle relaxant
		Diuretic
		Cardiotonic
		Dermatologic
		Receptor agonist/antagonist

a) Only examples are given for the listed classes of compounds or for their effects.

sometimes even prohibited. Hence, novel methods for the cultivation of such species are required to ensure the production of beneficial compounds from medicinal plants. Laboratory cultures have also the advantage that gene alterations created by molecular biological methods are possible in order to change pathways in favor of a desired metabolite (see Figure 1.32). Consequently, during the past few decades plant cell or organ cultures have become established for the production of bioactive metabolites. These contribute to a high-quality production of these compounds, based on the controlled environment of the cultures. In addition, large-scale bioreactors for plant cultures are now available whereby, the induction of secondary metabolites by elicitors is a feasible and easy-to-perform method under controlled conditions. The elicitors may be either stress-signaling compounds or changes in culture conditions, such as oxygen depletion. In addition, genetically modified cultures can be used to raise the concentration of a desired compound per biomass. These genetic alterations can lead to an activation of the biosynthetic pathway in a targeted fashion. Some features for the selection of plant materials used in secondary metabolite production are shown in Figure 1.31.

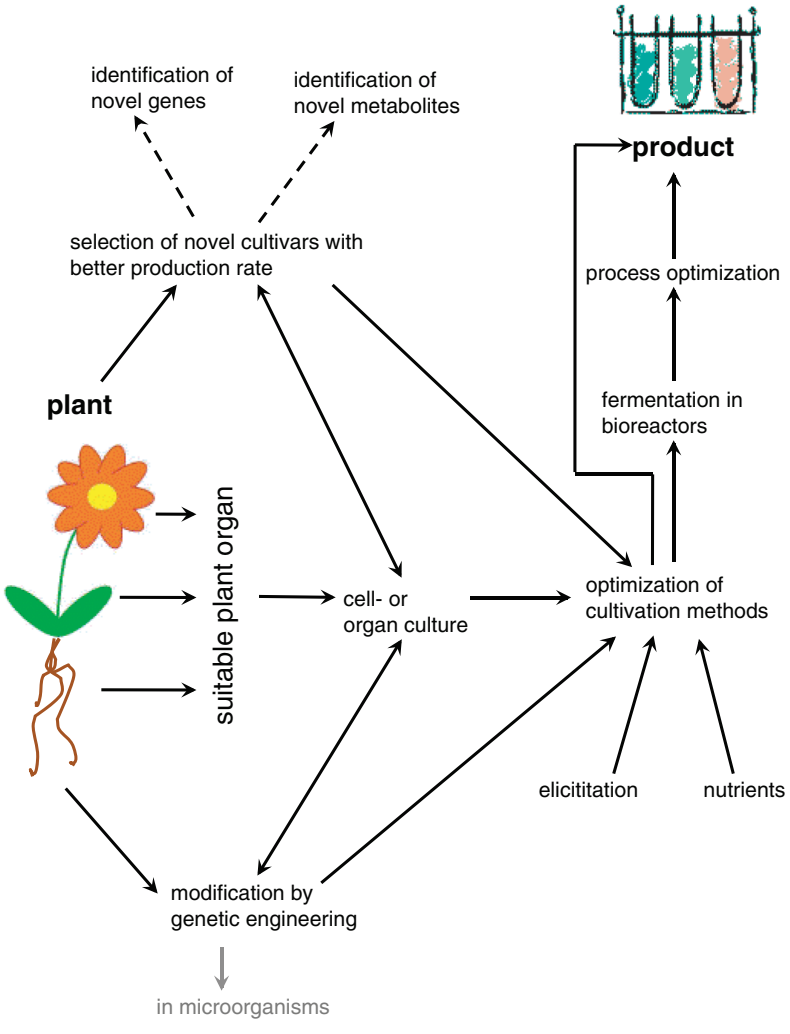


Figure 1.31 Production of secondary metabolites and optimization of plants for these procedures. It is indicated that microorganisms can also be used under certain conditions for the production of secondary metabolites of plants (gray arrows). Screening for novel genes and metabolites (broken arrows) can also be performed in cell or organ cultures and genetically modified organisms.

The first stage is to select the most suitable plant organ to generate the cell or organ culture (Figure 1.31), and to this end the metabolite pattern of the respective organs must be analyzed. The next step is to decide whether cell or organ cultures are to be established; the former are dependent on the continuous addition of plant hormones, while the latter are hormone-autonomous and therefore

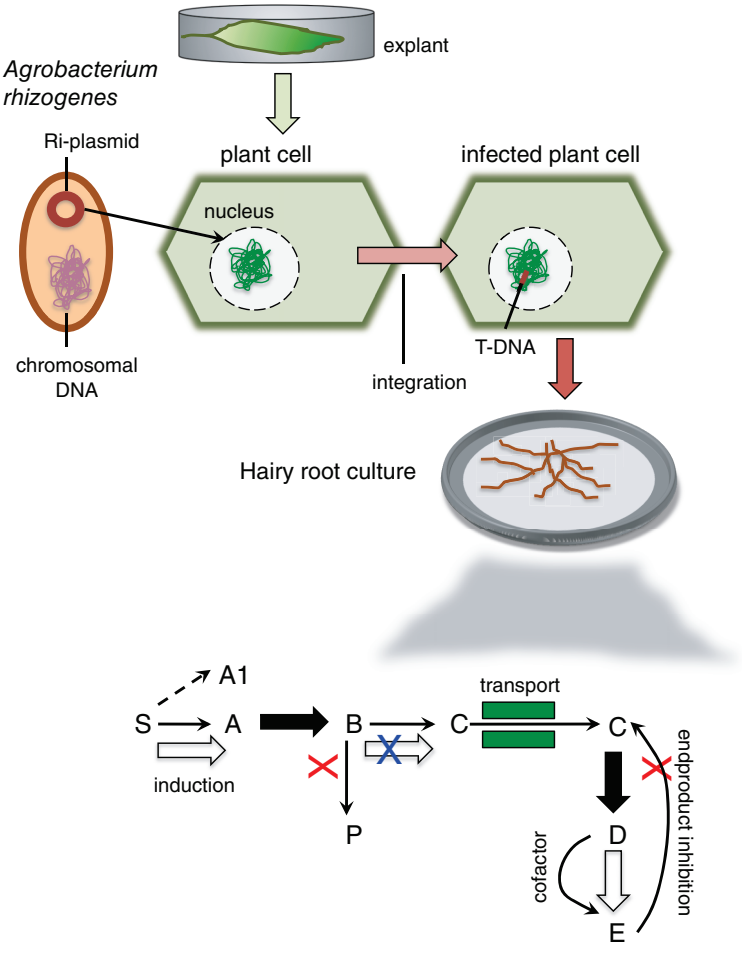


Figure 1.32 Generation of Hairy root cultures from leaf explants using *Agrobacterium rhizogenes*. These are inoculated with either wild-type or transgenic bacteria, which introduce their T-DNA from the Ti-plasmid (tumor-inducing) into the host genome, resulting in stable transformation events and a sterile Hairy root culture. Some factors are shown in the lower part of the figure which could lead to problems in secondary metabolite production after genetic modification. Closed arrows symbolize the situation in a wild-type culture or plant; open arrows and red-marked pathways situations in a transgenic culture or plant, resulting in better fluxes of metabolites. Small arrows show limiting steps. More factors are described in the text. Lower part based on Georgiev, M.I. *et al.* (2010) In: *Medicinal Plant Biotechnology* (Arora, R., ed.), CAB International, Oxon, UK, pp. 156–175.

cheaper to cultivate. The cultures can be induced from a variety of different plant tissues by the soil bacterium *Agrobacterium rhizogenes*, for example from the leaves, flowers, and also the roots (Figure 1.32). This phytopathogenic bacterium transforms plant roots with its tumor-inducing DNA (T-DNA), located on the

Ti-plasmid, that in Nature results in the symptoms of Hairy root disease. The genes transformed into the host genome are termed *rol* genes, but in addition to the *rol* genes the bacterium can be transformed by other genes taken from plants or other organisms, so as to alter the metabolic pathways or profiles in a desired manner. The resulting organ cultures are also termed Hairy roots, similar to the naturally occurring root disease. These cultures are used very often nowadays for secondary metabolite production, transformed with either wild-type or genetically modified bacteria (Table 1.9).

The genetically optimized cell or organ cultures must first be adapted to the cultivation conditions in bioreactors, and this may occasionally lead to stress situations due to the possible depletion of nutrients or shearing forces in the reactor. The cultivation can be also used to introduce stress factors deliberately, as this can increase the synthesis of the product, as well as elicitors (plant signaling molecules) to create biotic stress (Table 1.9). As the plants do not leave the laboratory during cultivation in a bioreactor, there is no (legal) conflict with these genetic modifications. A detailed knowledge of the biosynthetic pathways, transport between compartments, and their (transcriptional) regulation are necessary for targeted alterations. Details of the metabolites produced by cell and organ cultures

Table 1.9 Examples for the production of secondary metabolites in cell or Hairy root cultures. Often, a higher metabolite concentration was found in the cultures compared to the mother plant, even without genetic alterations. In many cultures higher metabolite contents was generated by elicitation using either biotic or abiotic stress factors or signals. It is also indicated by "yes" when the cultures were genetically modified by other genes, either from plants or bacteria (yes in brackets). In addition to the plant metabolites synthesized under normal conditions, after transformation with bacterial genes also other metabolites can be synthesized.

Plant species	Secondary metabolite	Cell culture	Hairy Root	Elicitor	Genetically modified
Ammi majus	Scopoletine		Yes	*Enterobacter*	
Argemone mexicana	Sanguinarine	Yes			
Atropa belladonna	Scopolamine Atropine [a)]Hyoscyamine		Yes	Nitrate	Yes
Beta vulgaris	Betalaines		Yes	*Lactobacillus*	
Catharanthus roseus	Indole alkaloids (chlorinated indole alkaloids)	Yes	Yes	*Trichoderma*	Yes (yes)
Cinchona ledgeriana	Quinine (Strictosidine)		Yes		(Yes)
Datura innoxia	Hyoscyamine Scopolamine		Yes		
Datura quercifolia	Hyoscyamine Scopolamine		Yes		

Datura candida	Hyoscyamine Scopolamine		Yes		
Datura metel	Scopolamine		Yes	*Aspergillus Alternaria Fusarium* Yeast extract	Yes
Duboisia leichhardtii	Scopolamine		Yes		(Yes)
Fagopyrun esculentum	Catechin Epicatechin		Yes		
Hyoscyamus niger	Hyoscyamine Scopolamine		Yes		Yes
Lavandula officinalis	Rosmarinic acid Caffeic acid	Yes		Oxygen depletion	
Lithospermum erythrorhizon	Shikonine	Yes	Yes		(Yes)
Morinda citrifolia	Anthraquinone	Yes			
Mucuna pruriens	L-DOPA	Yes			
Panax ginseng	Ginsenoside		Yes	Chitosan Methyl-jasmonate	
Papaver somniferum	Alkaloids	Yes			
Rauwolfia serpentina	Reserpine	Yes			
Rubia tinctoria		Anthraquinone		yes	
Salvia miltiorrhiza	Rosmarinic acid	Yes		Yeast extract	
	Tanshinone		Yes	Yeast extract	
Solanum khasianum	Solasodine		Yes		
Solanum tuberosum	Sesquiterpenes		Yes	*Rhizoctonia*	
Tagetes patula	Thiophene		Yes	*Fusarium*	
Taxus brevifolia	Paclitaxel	Yes			
Valeriana officinalis	Valpotriate		Yes		

a) racemic mixture from hyoscyamine and scopolamine.

are summarized in Table 1.9, including several important compounds with medicinal uses (see Chapter 5).

The question of which factors can influence the outcome of the production of secondary metabolites by a given biosynthetic pathway remains unclear. The best results can be obtained by using simple changes in pathways, where only

Figure 1.33 Activation of a single MYB transcription factor in a mutant of cauliflower (*Brassica oleracea*) leads to the purple color compared to wild-type. Although this picture shows a cultivated form, similar results can be obtained with a transgenic approach, as described in the text. A green variety is shown, which has chlorophyll in its tissues, for comparison. Fotolia.com, © Lsantilli.

one enzyme or transcription factor is activated. For example, synthesis of the alkaloid scopolamine via hyoscyamine and an intermediate is catalyzed in a two-step reaction by the same enzyme. Another example is the overexpression of a gene encoding a maize transcription factor in other plants; this particular transcription factor controls many steps in anthocyanin biosynthesis, and the various tissues of the resulting transgenic plants are typically colored. A similar effect is shown for a natural mutation in Figure 1.33, where a single MYB-type transcription factor is activated, leading to the purple- headed cauliflower (*Brassica oleracea*).

In order to modify complex biosynthetic pathways, many factors must be taken into account to ensure that the desired metabolite can be produced, including:

- the identification of the step determining the velocity;
- the availability of precursors;
- the limitation of cofactors;
- competing pathways using the same precursor or intermediates;
- inhibition by the final product;
- transport over cellular compartments; and
- transcription factors (Figure 1.32).

First, an attempt is made to determine the critical step for the overall velocity of the pathway. This step is an ideal candidate for metabolic engineering, and changes can be achieved by overexpressing the gene(s) that encode the proteins responsible for the limiting steps. By increasing the flux from substrate S to

metabolite A, the second metabolite A1 is not synthesized, and therefore production of the desired product C should be increased.

However, when homologous genes are used it must be guaranteed that there is no cosuppression effect that will reduce the expression of the transgene and of the original gene itself. The phenomenon of cosuppression was first observed after transforming petunia by a chalcone synthase gene; the effect led to white areas, which represented an indication of reduced gene expression and, consequently, to less pigment formation. In the example shown in Figure 1.32, intermediate C would no longer be generated, and this would lead to an accumulation of intermediate B. This in turn could lead to the undesired accumulation of side compound P.

The availability of precursors is also dependent on the number of biosynthetic pathways competing for a substrate. For example, tryptophan is a precursor for a wealth of different secondary metabolites, and also for protein synthesis and synthesis of the plant hormone IAA. Therefore, the overexpression of tryptophan decarboxylase could lead to a reduction in other indole metabolites, such as indole glucosinolates. In laboratory cultures this is not important, but in the field the plants would be less tolerant to various biotic stress factors. An alternative to the upregulation of the major pathway would be to suppress the side pathways, for example by applying antisense or RNA interference techniques. This would result in the availability of more precursors for the main pathway (Figure 1.32).

One major problem, however, would be the transport of intermediates between organelles, because the levels of transporters would also need to be increased; consequently, metabolic pathways that occur in only one compartment would be easier to manipulate. A modification of compartmented pathways might be possible by adding signal sequences to the heterologously expressed genes. Alternatively, a synthesis of metabolites in the culture medium could be attempted by adding secretion sequences to the protein(s). The secretion of a protein or reaction product ensures an easy purification of the product from the culture supernatant. Finally, the inhibition of an enzymatic reaction by either the product or an intermediate could also constitute a problem. In this case, even when the transgene is highly expressed, no product would accumulate. This is also the case when an essential cofactor is not present, although such a compound could be added to the culture, if not too expensive.

Catharanthus roseus produces the powerful anticancer drugs vinblastine and vincristine by dimerization of the terpenoid indole alkaloids vindoline and catharanthine (see Figure 1.6). Therefore, various modifications of these interesting metabolites have been genetically engineered in this species. The experimental strategy to modify terpene indole alkaloids by the Hairy roots technique is described here as an example. The biosynthesis of these metabolites starts with the generation of indole and terpene precursors via the shikimate and MEP pathways (see Section 1.1.2). The biosynthetic pathway of terpene indole alkaloids is especially highly compartmented, and therefore numerous transporters are also involved.

A key enzyme for the pathway is tryptophan decarboxylase, which has been used on many occasions as a target for genetic engineering, though several upstream enzymes can also be used (Figure 1.34). *C. roseus* Hairy root cultures expressing

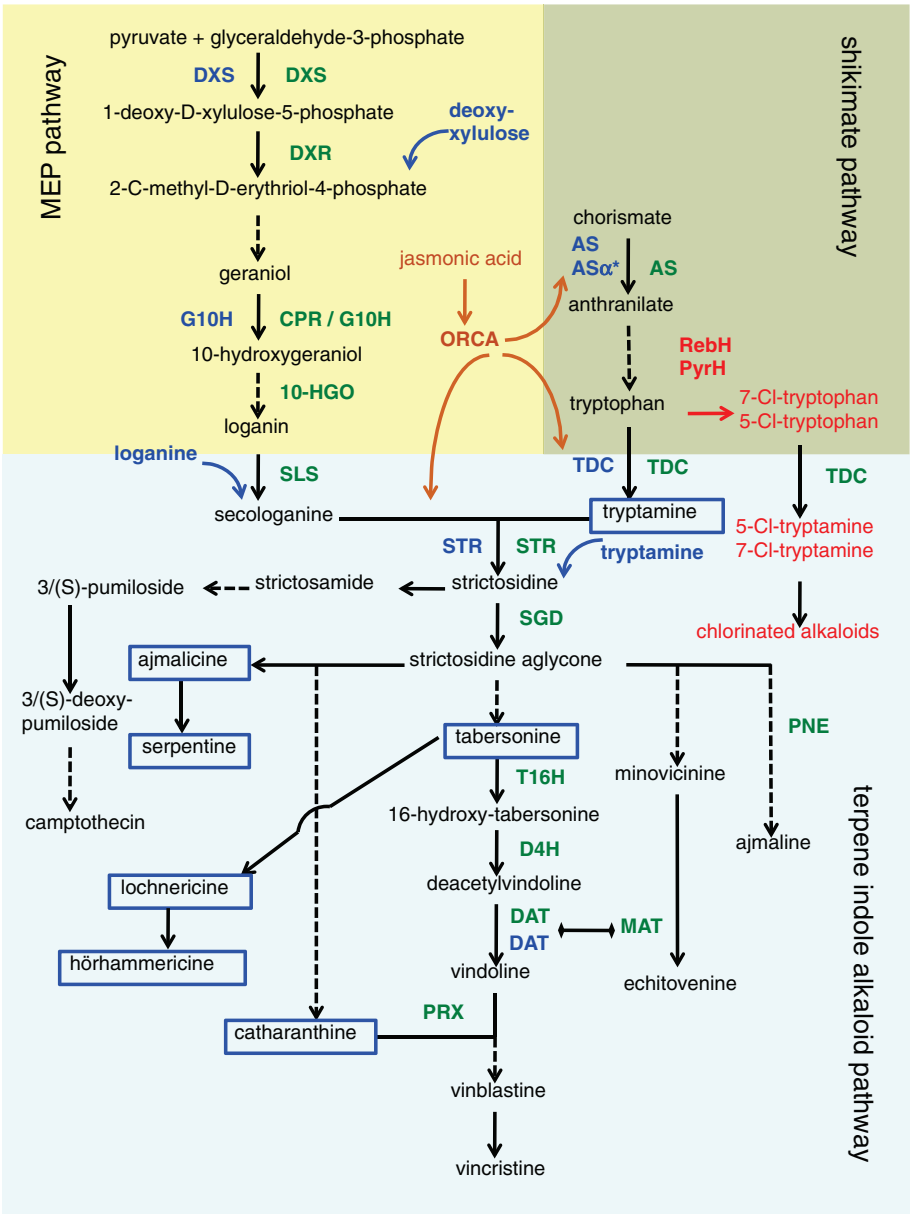


Figure 1.34 Genetically modified terpene indole alkaloid metabolism in *Catharanthus roseus*. Transgenes contain either homologous or heterologous genes or transcription factors. The MEP pathway (shaded in yellow) produces the terpene moiety; the part of the shikimate pathway delivering tryptophan is shaded in green; the core terpene indole alkaloid pathway is shaded in blue. Genes encoding native plant enzymes are shown in green; transgenic enzymes from the same plant in blue; heterologous enzymes from another plant in blue, and

various forms of anthranilate synthase and/or tryptophan decarboxylase have been generated with enhanced fluxes through the tryptophan branch of the terpene indole alkaloid pathway, such that enhanced levels of tryptamine and serpentine result. The additional treatment of anthranilate-overexpressing cultures with terpenoid precursors resulted in an elevation of various compounds, including catharanthine, ajmalicine, lochnericine, and tabersonine. Tabersonine is a direct precursor of the anticancer drugs vinblastine and vincristine, and therefore its increase could be beneficial for their production. In contrast, an increased production of side products such as lochnericine could result in a lower concentration of the precursors for the desired compounds. Increasing the expression of the gene encoding strictosidine synthase – an enzyme that catalyzes another rate-limiting step for terpene indole alkaloid biosynthesis – also enhances the synthesis of downstream metabolites, and their production is further enhanced if a precursor (loganin and/or tryptamine) is added. The overexpression of other genes from the pathway also alters metabolite composition.

In a different approach, the RNA-mediated suppression of tryptamine biosynthesis in *C. roseus* Hairy root culture can totally eliminate the production of terpene indole alkaloids, and be used to generate a spectrum of unscheduled products by introducing tryptamine analogs to the culture media. In addition, bacterial genes can be used to alter the metabolite spectra of Hairy roots. Previously, *C. roseus* has been successfully transformed using bacterial tryptophan halogenase genes, and this resulted in an accumulation of halogenated terpene indole alkaloids (Figure 1.34). Other possible strategies include the modulation of gene expression using transcription factors and/or elicitation by stresses, because the biosynthesis of terpene indole alkaloids is induced by many stress factors, including fungal elicitors, UV-B light, and jasmonate. The role of jasmonate has been confirmed, showing that it increases the levels of the transcription factor family ORCA 2 and 3 (octadecanoid-derivate responsive *Catharanthus* AP2-domain protein), thereby enhancing the transcription of several genes involved in the terpene indole alkaloid pathway.

marked with an asterisk; bacterial transgenes in red. Compounds that were increased in genetically modified lines are indicated by the corresponding genetic modification color (when not native to the plant) or blue boxes (if native). Elicitors such as jasmonic acid also increase flux through the pathway by inducing transcription factors (in orange). Arrows each indicate a single enzymatic step, dashed arrows indicate multiple steps. AS = anthranilate synthase; ASa = anthranilate synthase subunit a; CPR = cytochrome P450 reductase; D4 H = desacetoxyvindoline 4-hydroxylase; DAT = deacetylvindoline acetyltransferase; DXR = 1-deoxy-D-xylulose-5-phosphate reductoisomerase; DXS = 1- deoxy-D-xylulose-5-phosphate synthase; G10 H = geraniol-10-hydrolase; 10-HGO = 10- hydroxygeraniol oxidoreductase; MAT = minovincinine-19-hydroxy-*O*-acetyltransferase; MEP = methylerythriol-4-phosphate; ORCA = octadecanoid-derivate responsive Catharanthus AP2-domain protein; PNE = polyneuridine aldehyde esterase; PRX = peroxidase; SGD = strictosidine β-D-glucosidase; SLS = secologanin synthase; STR = strictosidine synthase; T16 H = tabersonine 16-hydroxylase; TDC = tryptophan decarboxylase. From Georgiev, M.I. *et al.* (2012) *Trends Biotechnol.*, **10**, 528–537; see Further Reading). Reprinted with permission from Elsevier.

Table 1.10 Examples for bioreactor configurations used for Hairy root cultivation in bioactive compound production. The operation mode, bioreactor type, volume of the bioreactor and an example for the plant, where a Hairy root culture was derived from, are shown. Modified from: Georgiev, M.I. *et al.* (2010) In: *Medicinal Plant Biotechnology* (Arora, R., ed.), CAB International, Oxon, UK, pp. 156–175.

Operation mode	Bioreactor type	Volume (liters)	Plant species
Mechanically driven	Stirred-tank reactor	5	*Beta vulgaris*
	Stirred-tank reactor with separate impeller	25	*Atropa belladonna*
	Wave reactor	0.5	*Panax ginseng*
Pneumatically driven	Bubble column reactor	2	*Harpagophytum procumbens*
	Bubble column reactor	2	*Beta vulgaris*
	Airlift reactor	2	*Nicotiana tabacum*
Bed reactors	Mist reactor	1.5	*Artemisia annua*
	Mist reactor	4	*Nicotiana tabacum*
	Trickle bed	14	*Hyoscyamus muticus*

In addition to changes in pathway fluxes, the molecular properties of plant enzymes can be altered to modulate pathways. In this case, the genes are mutagenized in such a way that is likely to modulate the activity of the encoded protein by, for example, altering the substrate-binding domain. This can only be achieved with proteins for which the structure is known in detail. The enzymatic properties can be first tested by gene expression in microorganisms and, if the desired mutant enzyme has been produced, then transformed into the plant. An alternative is "domain swapping," where complete domains with important properties are switched between enzymes, or between proteins. These proteins with novel structural elements could have completely new enzymatic properties; for example, the main biosynthetic pathway may be preferentially catalyzed compared to the side pathways. This has been shown to be effective for three genes encoding all of the dioxygenases of opium poppy (*Papaver somniferum*) morphine biosynthesis where novel mutant proteins were created after combining the various parts of the resulting protein, one of which had the desired property.

The type of bioreactor also influences the growth of the culture and the production of compounds (Table 1.10). A wide variety of bioreactor set-ups were used for the cultivation of transformed root culture, such as mechanically driven reactors (e.g., stirred-tank reactors, wave reactors, rotating drum reactors), pneumatically driven reactors (e.g., bubble column reactors and airlift reactors), and bed reactors (e.g., trickle bed reactors and mist reactors). Experimentally, it is challenging to select the type of bioreactor that will ensure the best media for the cultivation of transformed root cultures, although typical mechanically agitated reactors should not be used, if possible, because of

the high stress-sensitivity of the transformed roots. This is indicated by the high level of so-called plant stress hormones, which could have either positive effects by elicitation (Table 1.9) or negative effects due to the highly variable levels of compounds produced. However, slight changes in bioreactor design could allow the successful cultivation of transformed root cultures. Until now, the volumes used for these Hairy root bioreactors have been rather small (Table 1.10), and generally insufficient for the large-scale production of metabolites, in contrast to what is possible when using cell cultures.