Review: Combinatorial biosynthesis of medicinal plant secondary metabolites

Mattijs K. Julsing, Albert Koulman, Herman J. Woerdenbag, Wim J. Quax, Oliver Kayser

Biomolecular Engineering, in press
Abstract

Combinatorial biosynthesis is a new tool in the generation of novel natural products and for the production of rare and expensive natural products. The basic concept is combining metabolic pathways in different organisms on a genetic level. As a consequence heterologous organisms provide precursors from their own primary and secondary metabolism that are metabolized to the desired secondary product due to the expression of foreign genes. In this review we discuss the possibilities and limitations of combining genes from different organisms and the expression of heterologous genes. Major focuses are fundamentals of the genetic work, used expression systems and latest progress in this field. Combinatorial biosynthesis is discussed for important classes of natural products, including alkaloids (vinblastine, vincristine), terpenoids (artemisinin, paclitaxel) and flavonoids. The role and importance of today’s used host organisms is critically described, and the latest approaches discussed to give an outlook for future trends and possibilities.
Introduction

The approach to combine genes from different microorganisms for the production of new and interesting metabolites has become known as combinatorial biosynthesis. Recent achievements with the polyketide biosynthesis from microorganisms, especially in *Streptomyces*, prove the potential of combinatorial biosynthesis [3-6]. It also showed that this approach can be used to improve the biosynthesis capacity of known producing microorganisms like *Escherichia coli*, *Bacillus subtilis* or *Saccharomyces cerevisiae*. The heterologous expression of human genes in microorganisms is well known for more than 30 years now. Fundamental work on the expression of plant genes from biosynthetic pathways, performed since the 1980s, opens a way to similar research that may even be extended in the future by directed evolution. It is now possible to combine these genes and extend the realm of combinatorial biosynthesis far beyond the polyketide biosynthesis. The diversification of products will increase dramatically when genes of very different origins are used. However there is no need to concentrate on new compounds only; there are many interesting natural products, of which the application (e.g., as a drug or fine chemical) is hampered by its availability. This problem might be solved by using alternative production systems yet to be discovered, that are based on enzymes from other biosynthetic pathways. Nature and its huge biodiversity harbours an endless source of compounds containing unique chemical structures. Even on a species level a given biosynthetic pathway adapts through the continuous selection pressure of its surrounding. Only those compounds that are highly favorable for the producing organism are accumulated, which is a delicate balance between energy cost and physiological/ecological benefit. There are many speculations about how evolution diverges biosynthetic pathways [7]. Often the result is that specific compounds are produced by specific organisms. There are certainly products that will not be produced because they cost too much energy to synthesize, their activity is not beneficial enough or the organism lacks the enzyme machinery to perform a specific chemical reaction. In other words, the biodiversity is endless and there are still possibilities to enlarge the diversity from a chemical point of view, by combining genes and products from different sources that in nature would never meet. This strategy will deliver compounds that are not influenced by selection pressures, by a habitat, or the biochemical limitation of an organism (such as compartmentalization or storage). These compounds can be selected for a specific pharmaceutical mode of action or an activity can be adjusted to a more specific pharmaceutical demand.

There are several pharmaceuticals on the market that are highly expensive, due to the fact that these compounds are only found in rare plants and often in extreme low concentrations. Podophyllotoxin and paclitaxel (Fig.1) are clear examples of pharmaceuticals that can only be produced through the isolation from plants. To achieve a sustainable source of such compounds scientists all over the world have been experimenting with biotechnological approaches aiming at the development of an alternative production system. With this aim in mind, combinatorial biosynthetic strategies are expected to yield interesting alternatives in the near future. With regard to the production of podophyllotoxin it has been shown that plant cell cultures of *Linum flavum* L. can be used to convert deoxypodophyllotoxin, a major lignan of *Anthriscus sylvestris* L. into 6-methoxypodophyllotoxin [8, 9]. The combination of the product of one species and the enzymes of another species to yield a desired product is a good example of combinatorial biosynthesis. This topic will be extensively discussed in the following subchapters.
Fig. 1: Important plant natural products subject to combinatorial biosynthesis studies
Not only can the expression of a single gene be of interest. The reconstruction of complete biosynthetic pathways by combining genes of the desired pathway in host organisms is the current aim of actual research projects. There are many papers describing the functional heterologous expression of single genes from biosynthetic pathways. Still in contrast the coupling of more genes and the controlled expression of genes encoding biosynthetic enzymes for metabolizing precursors is a challenging approach. Thus far, the biosynthesis of flavonoids in *E. coli* is the only total heterologous biosynthesis of a plant compound that has been described [10, 11], but promising results have been reported already for the biosynthesis of artemisinin [12-14], paclitaxel [15] and strictosidine [16]. We will discuss the biosynthesis of specific natural products in detail, and we want to give insight in the basic understanding of the concept of combinatorial biosynthesis of other natural products, which is gaining more and more interest.

**Definition of Combinatorial Biosynthesis**

The definition of combinatorial biosynthesis has been changed and is still changing because of the rapid developments in molecular biological techniques and innovative strategies applied in this research area. From the past, combinatorial biosynthesis is defined on the metabolic level, using different precursors or further modification of a structural scaffold. The concept of combinatorial biosynthesis has been introduced from the work with polyketides and oligopeptides. These natural products were model compounds showing that repeated use of the same type of reaction with different precursors like acetyl-CoA units or amino acids can lead to a combined biosynthetic product. The finished peptide or polyketide scaffold can be posttranslational structurally modified. Also this step has been accepted as part of combinatorial biosynthesis. An important example of combinatorial biosynthesis on the metabolic level is the development erythromycin analogues [17, 18], which are impossibly obtained by synthetic organic chemistry. The scope of combinatorial biosynthesis and the number of structural variants, which can be generated by manipulation of biosynthetic modules, is limited by the specificity of different domains and modules for initiating, extending and terminating the growing chain of the polyketide or the nonribosomal peptide, or even by combinations thereof. To date, combinatorial biosynthesis of natural products has to be defined wider, not focussing the metabolic level only. With the current knowledge of molecular biology, it has become possible to combine genes (thus also the resulting enzymes) and products of different organisms. This can yield a further diversification of both chemical and natural product libraries. Because these strategies have also become known as combinatorial biosynthesis, we define combinatorial biosynthesis as the approach to combine genes from different organisms to produce bioactive compounds.

Current research in this field still focuses mainly on the polyketide biosynthesis in microorganisms. But a careful examination of the literature on plant biotechnology reveals that several studies have already been carried out in the past twenty years that can now be called combinatorial biosynthesis as we use the new definition. Due to the strategy of combinatorial chemistry at the beginning of the eighties, which uses a random approach to synthesize novel polymeric or oligomeric chemical entities from uniform monomers (e.g. amino acids), the term combinatorial biosynthesis since the 1990s suggested a random approach and combination of genes in the polyketide or terpenoid biosynthetic pathways using also biosynthetic monomers (e.g. isoprenes, acetyl and propionyl units) from natural
origin. Today, we would like to add to this definition the possibility to have directed and controlled combination of genes to produce a desired single compound. At the moment combinatorial biosynthesis of plant secondary metabolites focuses on the reconstruction of the basic pathways into microbial hosts. This review gives a survey of the use of genes and products from plants in combination with genes and products from other organisms. It emphasizes the potential of plant combinatorial biosynthesis for drug discovery and its future importance for pharmaceutical sciences.

Bioconversion capacity of plant cells

Bioconversion can be defined as the transformation of one chemical into another using a biocatalyst. The biocatalyst can be a cell (e.g. microorganism, plant or animal cell), a vital extract from such cells, or a (partly) purified enzyme. The biocatalyst may exist free, in solution, immobilized on a solid support or entrapped in a matrix. In bioconversions by whole cells or extracts one single enzyme or several enzymes may be involved.

In the past twenty years the use of enzymes as catalysts for the preparation of novel organic molecules has received increasing attention. Enzymes can catalyze a wide range of reactions; it is likely that nearly all existing compounds can react with an appropriate enzyme. Even persistent environmental pollutants such as pesticides, raw oil, and halogenated hydrocarbons can be degraded by certain bacterium species [19].

Each individual cell contains a plethora of enzymes that can display different catalytic properties depending on the conditions to which they are exposed. From genome studies on different organisms it is well known that approximately 30,000 up to 35,000 genes are present in a plant organism [20]. Not all genes are involved in the biosynthesis of secondary metabolites, but a number of proteins/enzymes is capable of catalyzing more than one reaction [21], which will compensate the above assumption. This means that in a higher plant around 30,000 low molecular weight products are biosynthesized and should be present. Today, from no plant species the metabolic profile is completely known, and because of the insufficient analytical techniques and the broad range of polarities and quantities of natural products it seems to be an unrealistic endeavour. Some of the best studied plants are Arabidopsis thaliana and Nicotiana tabacum (tobacco), but here the number of known constituents does not exceed 3,000 [22, 23].

Since each enzyme has its own specific active site, enzymes usually show selectivity towards their precursor. However, a number of enzymes with broad specificity are known. There are two specific properties of enzymes that most chemical catalysts do not possess, making them especially interesting for combinatorial biosynthesis: stereo- and regio-specificity. Stereospecificity: enzymes are chiral catalysts and often able to produce molecules of high optical purity, which characteristic can be of influence on the biological activity. This makes the production of such compounds by enzymes especially interesting as pharmaceuticals. For example, the active stereoisomer of the antimalarial artemisinin is biosynthesized in Artemisia annua L. plants while 128 (27) stereoisomeric forms are theoretically possible (Fig. 2). Two areas of bioconversions that have great relevance for organic synthesis involve the use of hydrolase enzymes on the one hand and oxidoreductase enzymes on the other, e.g. in cells of Papaver somniferum [24, 25]. The stereocontrolled formation of carbon-carbon bonds is the heart of organic synthesis and this reaction is performed by aldolase enzymes. The search for and the employment of lipases, esterases
and amidases for the preparation of chiral compounds of high optical purity will also be continued in the future.

General reaction types catalyzed by enzymes like oxidation, reduction, hydroxylation, methylation, demethylation, acetylation, isomerization, glycosylation, esterification, epoxidation and saponification can all be used in the diversification of compounds and are therefore relevant in combinatorial biosynthesis [26].

The yield of a plant biosynthetic pathway leading to important or expensive products can be improved by feeding precursors or intermediates. This may be an economically interesting strategy when the precursors are cheap and easily obtainable. The enzyme-catalyzed modification of added precursors, into more valuable products has been performed with plant cells, either freely suspended or in an entrapped state, or with enzyme preparations, sometimes from a heterologous host organism. These biocatalytic systems are mostly able to perform stereo- and regiospecific reactions on a sometimes surprisingly broad precursor range, even including cell-foreign, chemically prepared compounds. From a pharmaceutical point of view, hydroxylations and glycosylations [27, 28] are considered to be particularly useful bioconversions. They can yield new drugs and existing drugs can be improved as to increased activity and decreased toxicity. The biological availability, meaning the overall blood concentration a drug reaches after administration and resulting in a therapeutic action, can be enhanced by the introduction of hydrophilic moieties in the pharmacologically active molecule. The therapeutic action can either be prolonged by the introduction of protecting groups resulting in so-called prodrugs, or be increased when the new moieties enhance the affinity for target cells or receptors involved. Furthermore, side-effects can be reduced and the stability increased by modification of the parent drug. In the following sections of this chapter we concentrate on the role of plant cells and their enzymes as biocatalysts for the production of plant secondary metabolites and related compounds with special attention to (potential) pharmaceuticals. The sections are ordered by their biosynthetic pathways.

**Combinatorial biosynthesis of terpenoids**

Terpenoids represent a large and important class of natural products with more than 30,000 different structures. Terpenoids (consisting of C5 (isoprene) “building blocks”) are known for their wide commercial applications, such as flavour and fragrance additives. Essential oil constituents are predominantly monoterpenoids (C10) and sesquiterpenoids (C15). From a pharmaceutical point of few the sesquiterpenoids are of high relevance. In this group artemisinin, gossypol, and zingiberene (Fig. 1) are of high medicinal and economic interest. Furthermore diterpenoids (C20) are of high interest and paclitaxel as a major representative of this group is a blockbuster drug. Carotenoids with a C40 backbone show important functions in photosynthesis, pigmentation and as antioxidants. From a pharmaceutical point of view the most important class of terpenoids is without doubt formed by the sterols, which are derived from a C30 backbone and are used as starting material in the organic synthesis of synthetic drugs like steroid hormones and contraceptives.

Terpenoids are biosynthesized via the mevalonate (MVA) pathway or the deoxyxylulose phosphate (DOXP) pathway. Both pathways are described elsewhere in an excellent way [29-31]. It is important to mention that the upstream biosynthetic steps are genetically well mapped out. In contrast to other pathways, such as alkaloids or phenolics, the current knowledge allows transfer of the steps from the terpenoid pathways into microbial hosts.
and to hook on with extended pathways for higher terpenoids. The MVA pathway has recently been established in *E. coli*, harbouring the DOXP pathway itself, and an efficient production was shown for the terpenoids amorpha-4,11-diene [12] and taxadiene [32]. For the production of artemisinin or paclitaxel the presence of a terpene cyclase is a prerequisite. The conversion of a linear isoprenoid precursor (e.g. farnesyl diphosphate or geranylgeranyl diphosphate) to a cyclic terpene such as amorphadiene or taxadiene for artemisinin or paclitaxel, respectively, is still considered as a rate limiting step in the overall biosynthesis [33]. Terpenoid cyclization involves the generation of a reactive carbocation and moving of the ions over the isoprene backbone and correct dimensional coupling to the desired cyclic form [34]. Terpene cyclases have been described for microorganisms and plants, which often contain multiple cyclase genes as has been shown for *Arabidopsis thaliana* with 40 different genes [35].

**Artemisinin**

Artemisinin (Fig. 1) is an antimalarial drug isolated from *Artemisia annua*, Asteraceae. The drug is especially used in those areas where resistance of *Plasmodium falciparum* against the commonly used antimalarials is often found. Until now the costs for artemisinin treatment are much too high for most people in low income countries suffering from (life threatening) malaria. Several initiatives have been undertaken to lower the costs for this treatment. The selection of plants yielded varieties containing 0.5-0.8 \% of artemisinin in the aerial parts based on dry weight [36-38]. Alternatives could be the production via transgenic plants or engineering the biosynthetic pathway into less complex host cells. This implies that the full elucidation of the biosynthetic pathway is required. Although several biosynthetic pathways have been postulated, until now only the genes encoding the enzymes for the synthesis of the first specific intermediate amorphadiene by amorphadiene synthase [39, 40] and for artemisinic acid by the cytochrome P450 enzyme, CYP71AV1 [41] have been isolated and identified. The cDNA encoding amorphadiene synthase has been expressed in *E. coli* and characterized [42]. The recently discovered enzyme, CYP71AV1, has been shown to be able to catalyze the regioselective oxidation of amorphadiene into artemisinic alcohol. Besides this metabolic action the enzyme has been shown to be able to oxidize the precursors artemisinic alcohol and artemisinic aldehyde yielding artemisinic acid [41]. Nevertheless, more effort should be invested in the elucidation of the subsequent steps of the pathway leading to artemisinin (Fig. 2).
Fig. 2: Proposed biosynthetic pathway to artemisinin starting from farnesyl diphosphate (FDP) via the intermediates amorpha-4,11-diene, artemisinic alcohol (AAOH), artemisinic aldehyde (AAA), dihydroartemisinic alcohol (DHAAOH), dihydroartemisinic aldehyde (DHAAD), artemisinic acid (AA), and dihydroartemisinic acid (DHAA) [41, 43].
Despite the lack of knowledge of the entire biosynthetic pathway, research already achieved some progress in the metabolic engineering of host cells for the production of amorpha-4,11-diene. An artificial fusion of the proteins farnesyl diphosphate synthase from *A. annua* and 5-epi-aristolochene synthase from *Nicotiana tabacum* yielded a bifunctional enzyme producing the sesquiterpenoid structure from isopentenyl diphosphate (IDP) and geranyl diphosphate (GDP) [44]. The same technique could be applicable for amorphadiene synthase.

The concept of *E. coli* as a host cell producing sesquiterpenoids out of the endogenous pool of farnesyl diphosphate (FDP) has been investigated [45]. This work resulted in the production of 10.3 µg of (+)-δ-cadinene, 0.24 µg of 5-epi-aristolochene, or 6.4 µg vetispiradiene per liter of bacterial culture. Furthermore the authors concluded that the poor expression of the plant terpene cyclases was limiting for the synthesis of sesquiterpenes and not the endogenous supply of FDP. This has been confirmed in their further work by coexpressing the *E. coli ddx* gene, which did not result in an increase of sesquiterpenoids produced where it did result in an increase of lycopene production in *E. coli* [46, 47].

To overcome the low enzyme levels, the expression of amorphadiene synthase has been optimized by constructing a synthetic amorphadiene synthase gene completely optimized for the expression in the bacterial host. This strategy has been combined with engineering of genes from the mevalonate dependent isoprenoid pathway, which resulted in an *E. coli* strain producing 24 µg/ml amorpha-4,11-diene (calculated as caryophyllene equivalent) from acetyl-CoA after supplementation of 0.8% glycerol [12]. Recently, attempts to use *S. cerevisiae* for the production of artemisinin precursors have been described. The expression of the amorphadiene synthase gene in yeast using plasmids and chromosomal integration led to the production of respectively 600 and 100 µg/ amorpha-4,11-diene after 16 days batch cultivation [13]. Using a *S. cerevisiae* strain containing an engineered MVA pathway coupled with the genes encoding amorphadiene synthase and CYP71AV1 the production of artemisinic acid up to 100 mg/l has been reported [14]. This strain transported the artemisinin precursor outside the yeast cell, which makes purification of the product less complex. Artemisininic acid can be used for the semi-synthesis of artemisinin, but to lower the costs for production of the drug bioprocessing must be optimized [48].

**Paclitaxel**

Paclitaxel (Fig. 1), mostly described by the tradename Taxol®, is a diterpenoid that can be found in the bark and needles of different *Taxus* trees. Paclitaxel was first isolated from *Taxus brevifolia* (Pacific yew tree) Taxaceae in the sixties from last century [49], and its derivative Taxotere® was clinically introduced 30 years later for the treatment of mainly ovarian and breast cancers. Isolation from the *Taxus brevifolia* bark is a problem, because of the low yield (500 mg kg⁻¹). Facing the high demand, various *Taxus* species are endangered in China and India. Paclitaxel has a complex chemical structure. Its total synthesis has been established [50], but the complexity and low yield of this alternative for natural sources made it commercially inapplicable. Semisynthetic approaches have been more successful [51], since the more easily available intermediate taxoids, like 10-deacetylbaccatin III, can be isolated from the green needles of various *Taxus* species and used as starting material.
Fig. 3: Biosynthetic pathway of paclitaxel starting with the formation of geranylgeranyl diphosphate (GGDP) from isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP) catalyzed by geranylgeranyl diphosphate synthase (GGDPS), followed by several biosynthetic steps catalyzed by taxadiene synthase (TS), cytochrome P450 taxadiene 5α-hydroxylase (THY5a), taxadiene-5α-ol-O-acetyltransferase (TAT) and cytochrome P450 taxane 10β-hydroxylase (THY10b) [52].
Nevertheless, the production of paclitaxel still relies on the yew species or on cell culture systems derived from these plants. *Taxus* cultures elicited by methyl jasmonate showed an increased biosynthesis of paclitaxel [53]. The biosynthesis of paclitaxel starts with the cyclization step from geranylgeranyl diphosphate (GGDP) to taxadiene (Fig. 3). Most of the 19 known enzymatic steps in the biosynthesis are related to hydroxylation and other oxygenation reactions of the taxadiene skeleton [52, 54]. Croteau and coworkers isolated and identified several genes from different *Taxus* species that are responsible for steps in the biosynthesis and built a basis for today’s combinatorial biosynthesis in a heterologous microorganism. Today, all genes have been cloned into *E. coli* and activity screening confirmed the function of isolated enzymes [52-67]. The first intermediate, taxadiene can now be produced in *E. coli*. Co-expression of the taxadiene synthase from *Taxus brevifolia* [68] with a geranylgeranyl diphosphate synthase isolated from *Erwinia herbicola* [69], isopentenyl diphosphate synthase from *Schizosaccharomyces pombe* [70], and the endogenous deoxyxylulose 5-phosphate synthase from *E. coli* resulted in a production of 1.3 mg taxadiene per liter of cell culture [32]. This non-optimized system proved the principle of genetically engineering *E. coli* for the heterologous production of taxanes by combining enzymatic biosynthetic steps derived from several different organisms.

Recently, Dejong described the genetic engineering of *S. cerevisiae* for the production of taxadien-5α-acetoxy-10β-ol and the implementation of 8 of the 19 genes in two plasmids [15]. Figure 3 provides an overview of the biosynthetic pathway of paclitaxel, including the genes that have been transferred to *S. cerevisiae* by Dejong. The use of *S. cerevisiae* seems to tackle a physiological problem in the combinatorial production. *E. coli* does not have an efficient isoprenoid biosynthetic pathway. This is the main reason to clone this pathway into *E. coli* as discussed before for artemisinin. A second problem with *E. coli* is the limited supply of complementary NADPH:cytochrome P450 reductase that is also essential for the correct function of reconstituted plant cytochrome P450 enzymes. Yeast does have endogenous microsomal cytochrome P450 enzymes and energy supporting systems, which is a major advantage for this host system.

The taxadiene synthase encoding gene has also been expressed in *Arabidopsis thaliana* [71]. Constitutive expression of the gene led to taxadiene accumulation, but the *A. thaliana* plants showed growth retardation and decreased levels of photosynthetic pigment. The negative effects may have been caused by the toxicity of taxadiene, but more likely they are a result of the disturbance of the endogenous geranylgeranyl diphasphate pool. The use of an inducible expression system resulted in an increase of taxadiene accumulation. These findings clearly show that only the expression of heterologous genes result in the production of the desired compound, but the influence on the metabolic network has to be taken into account as well.

**Carotenoids**

Carotenoids are tetraterpenoids (C40 compounds) and produced in many plants and microorganisms. Their main biological function is the protection against oxidative damage and some are used as warning colors in plant defense system. The commercial interest for carotenoids can be explained mainly by their use as colorant, nutraceutical, or antioxidant in food and cosmetics. Next to that, it has been suggested that carotenoids could possibly play an important role as anticarcinogenic drug and in the prevention of chronic diseases.
Combinatorial biosynthesis

The carotenoid β-carotene is the primary source of vitamin A in the human diet. The biosynthesis of carotenoids starts with the tail-to-tail coupling of two molecules of the general precursor GGDP by phytoene synthase (CrtB) resulting in the colorless carotenoid phytoene. Desaturation reactions inserting four additional double bonds in the molecule give eventually lycopene, the main carotenoid in tomato fruit, from which different cyclic and acyclic structures can be synthesized depending on the producing organism. Lycopene cyclase (CrtY) catalyzes the cyclization at both ends of the lycopene molecule, resulting in two β-rings at the molecule β-carotene. Several other enzymes involved in the carotenoid biosyntheses have been identified, responsible not only for cyclization, but for glycosylation and diverse oxygenations as well.

More than 600 different naturally occurring carotenoids have been identified so far. The three main carotenoids β-carotene, astaxanthin, and lycopene are produced by chemical synthesis [76] and fermentation [77] for commercial purposes. However, for carotenoids combinatorial biosynthesis in microorganisms is also described [78]. Several carotenoid producing plants have been genetically modified to increase the production of the desired compounds. This review does not describe this research topic in detail, but the use of transgenic medicinal plants of Lycopersicum esculentum, Daucus carota, Solanum tuberosum, and Brassica napus has been reported (reviewed by Fraser and Bramley, 2004) [79]. To overcome the problems with vitamin A deficiencies in the third world, the biosynthetic pathway to β-carotene engineered in rice (Oryza sativa) has led to the production of Golden Rice providing β-carotene, also referred to as pro-vitamin A [80, 81]. Here we focus on the use of microorganisms for the production of carotenoids.

The production of carotenoids by fermentation of carotenoid producing microorganisms such as Xanthophyllomyces dendrorhous, Haematococcus pluvialis, and Blakeslea trispora has been investigated [82-85]. X. dendrorhous produces 200-400 µg g⁻¹ astaxanthin (85% of total carotenoid content). Engineering of X. dendrorhous by random mutagenesis led to an increase of 1.5-9 fold of the astaxanthin production in mutant strains. As a disadvantage of this approach growth inhibition and a decrease of biomass have been observed. More sophisticated recombinant DNA techniques introducing multiple copies of genes encoding a bifunctional phytoene synthase / lycopene cyclase and a phytoene desaturase also showed an increase in carotenoid production, but unexpectedly mostly other carotenoid structures than the desired astaxanthin (reviewed by Visser et al, 2003) [82]. Apparently, the hydroxylating enzyme became limited by overexpressing the mentioned enzymes. Several groups used gene clusters of Erwinia sp. for the expression in other hosts. In the last years several non-carotenoid producing organisms have been explored for the production of carotenoids. This heterologous production is dependent on efficient expression systems for the carotenoid gene clusters, but increasing the supply of precursors in the host organisms is of importance as well. The yeasts Candida utilis and S. cerevisae have been engineered for the production of lycopene, β-carotene, and astaxanthin [86, 87]. The prokaryote E. coli is most elaborated as a heterologous host, because most of the genes were already expressed in the strain for functional analysis. An overview of the heterologous expression of carotenoid gene clusters in the three mentioned non-carotenogenic hosts is described by [88].

The production of carotenoids in a host requires the biosynthesis of the intermediate GGDP. E. coli produces the C15 precursor FDP for endogenous terpenoid molecules. The extension of the prenyl chain to C20 has been performed by the expression of the CrtE gene.
encoding geranygeranyl diphosphate synthase from Erwinia sp. [89]. This prenyltransferase catalyses the production of GGDP from FDP. The GGDP synthase encoding gene gps from Archaeoglobus fulgidis has been expressed as well. Expression of this gene is more efficient, because the enzyme catalyzes the three chain elongation reactions starting from the C5 precursors to the C20 molecule [90].

One way to increase the heterologous production is to increase the pool of precursors in the host. Overexpression of several genes upstream in the isoprenoid biosynthesis resulted in the identification and overcome of bottlenecks in this pathway. Where the expression of a carotenoid gene cluster in C. utilis resulted in a lycopene production of 1.1 mg g\(^{-1}\) (dry weight) of cells [91], the overexpression of the catalytic domain of the HMG-CoA enzyme, involved in the isoprenoid biosynthesis via the mevalonate pathway, resulted in a 4-fold increase. Following disruption of the ergosterol biosynthetic gene ERG9 encoding squalene synthase yielded even more lycopene (7.8 mg g\(^{-1}\) (dry weight) of cells) [92].

To increase the isoprenoid flux in E. coli several genes of the DOXP pathway have been overexpressed. This resulted in a maximum increase of 10 times of the total carotenoid production. Overexpression of genes encoding enzymes involved in a biosynthetic pathway is not always the solution for higher production levels, because they often cause an imbalance in the metabolic system of a host cell. Regulation of the supply of precursors and expression levels can contribute to the heterologous biosynthesis systems as well. The negative effects of overexpressing a rate limiting protein have been demonstrated for the deoxyxylulose phosphate synthase gene (dxs). The use of a multicopy plasmid containing a tac promoter resulted in a decrease of growth and lycopene production when expression was induced by IPTG where the dxs gene constructed on a low copy plasmid did not show these negative effects [93]. Instead of plasmids the strong bacteriophage T5 promoter has been used to replace native promoters in E. coli. As a consequence the increased expression of isoprenoid genes led to improved production of lycopene (6 mg g\(^{-1}\) of dry cell weight) in E. coli [94].

The balance of the starting precursors of the DOXP pathway has been investigated by Farmer and Liao, 2001 [96]. Overexpression of several central metabolic genes redirected the flux of pyruvate towards glyceraldehyde 3-phosphate, resulting in an increase of lycopene in the heterologous E. coli strain. The same group also tried to design a controlled expression system for limiting enzymatic steps using an artificial intracellular loop [97]. Since most carotenoid genes of different origin can function together in a host, combining several enzymatic combinations led to the production of new carotenoid structures not isolated from nature before [77].

The use of host cells gives the opportunity to use directed evolution techniques for the modification of enzymes as well. Schmidt-Dannert and coworkers shuffled phytoene synthases of different bacterial species, which has resulted in a fully conjugated carotenoid containing six instead of four double bonds [98]. The combination with shuffled lycopene synthases has shown production of the monocyclic carotenoid torulene. Extension of these pathways with other carotenoid modifying enzymes led to the production of novel structures in E. coli [99]. Directed evolution has been used to create carotenoid-like
molecules with different amounts of carbon atoms (C30, C35, C45, C50) as well [100, 101].

Out of the group of terpenoids, the carotenoids have been most investigated in the production by naturally non-producing microorganisms and the production of new structures by combinatorial biosynthesis strategies. In contrast to the commercial interest, the pharmaceutical relevance of these compounds seems not to be of high importance at the moment. However, the knowledge out of this work can be applied for the heterologous production of other valuable terpenoid drugs like the mentioned artemisinin or paclitaxel. Although the availability of carotenoid gene clusters and promiscuity of the enzymes involved in the carotenoid biosynthesis are not present for structures of other terpenoids, the progress made, especially in engineering the upstream pathway creating a higher flux of general isoprenoid precursors, can be useful for all terpenoid structures as counts for the directed evolution techniques as well.

**Combinatorial biosynthesis of alkaloids**

By definition alkaloids contain nitrogen which is usually derived from amino acids. Because of the presence of a nitrogen atom, alkaloids react mostly alkaline and are able to form soluble salts in aqueous environments. In plants however, they can occur in the free state, as a salt or as an N-oxide and they are accumulated in the plant vacuole as reservoir or often coupled to phenolic acids like chlorogenic acid or caffeic acid. Alkaloids can be classified in terms of their biological activity, their chemical structure, or more accepted according their biosynthetic pathway. In plants over 12,000 alkaloids are known and several are used medicinally with a world market volume of 4 billion US$. Alkaloids are usually divided into five major groups depending on the amino acid of origin in the biosynthesis (amino acid in brackets):

1. Tropane-, Pyrrolidine- and Pyrrolizide-alkaloids (ornithine),
2. Benzylisoquinoline (tyrosine),
3. Indolequinoline (tryptophane),
4. Pyridine (pyridine), and
5. Quinolizidine- and Piperidine-Alkaloids (lysine).

Combinatorial biosynthesis of alkaloids is known for a few examples like vincristine, vinblastine, ajmaline [102] and morphine from plants and rebeccamycin and staurosporine from *Streptomyces albus* [103, 104]. The compounds mentioned have in common that a rather long biosynthetic pathway (30 enzymes for monoterpenoid indole alkaloids like vincristine and more than 17 enzymes for morphine) has to be elucidated and transferred into a heterologous host. In this review we discuss in detail the combinatorial biosynthesis of morphine as benzylisoquinoline alkaloid and *Vinca* alkaloids as monoterpenoid indole alkaloids as examples for recent research strategies.

**Benzylisoquinoline alkaloids**

Morphine (Fig. 1) is the most important member of the group of benzylisoquinoline alkaloids and is a natural product with high medicinal significance. Also other benzylisoquinoline alkaloids are pharmacetically important. Like morphine, codeine is used as an analgetic. Berberine and sanguinarine are used as antimicrobials and others as muscle relaxants like papaverine and (+)-tubocurarine.
The morphine biosynthesis consists of 17 steps in *Papaver somniferum*, Papaveraceae, and has almost completely been elucidated. In the biosynthesis a key intermediate is (S)-norcoclaurine, that is biosynthesized by condensation of dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA). The catalyzing enzyme (S)-norcoclaurine synthase has recently been identified from *Thalictrum flavum*, Ranunculaceae, and cloned in *E. coli* [105, 106]. Further key enzymatic steps towards (S)-reticulin include three NADPH oxidoreductases [24, 25] and cytochrome P450 [107, 108] and an acetyl-CoA dependent acetyltransferase [109, 110]. Recently the last step reducing codeine to morphine by codeinone reductase has been elucidated and the gene expressed in insect cells and / or in *E. coli* [111].

Berberine as a second representative for benzylisoquinoline alkaloids is known from different plants, but is mostly associated with *Chelidonium majus*, Papaveraceae, and responsible for the color and the antimicrobial activity of the yellow latex and the plant extract [112]. With the exception of the oxidase leading to the quartenary nitrogen all enzymes are known, like the key step for bridging from (S)-reticulin to (S)-scoulerin [113], the introduction of a methylgroup to give (S)-tetrahydrocolumbamin [114], and the building of a methylenedioxy ring [115]. Therefore we can expect that in the near future the successful combinatorial biosynthesis of berberine in a heterologous host will be tested.

**Vinca alkaloids**

Vinblastine (Fig. 1) and vincristine are monoterpenoid indole alkaloids from *Catharanthus roseus*, Apocyanaceae, and are used in medicine as antineoplastic drugs. Because of the high importance and the extreme low yield from plants (3 mg kg\(^{-1}\)) [116] they could be considered as trace compounds. For the production of 3 kg of *Vinca* alkaloids, which is the annual need worldwide, around 300 tons of plant material has to be extracted [117]. Production of *Vinca* alkaloids in plant cell cultures did not lead to a significant improvement and today it is accepted that biotechnological approaches in plant cell culturing may not provide an instant solution to this problem [116].

The biosynthesis of vincristine and vinblastine is complex and is shown in Fig. 4 for the early phase starting from geraniol to strictosidine and in Fig. 4 for the late phase leading to the desired compounds [116]. In the early phase tryptophan and secologanin as terpenoid precursors are condensed to form strictosidine as an important branching intermediate for also other alkaloids. In this short part of the entire route already seven enzymes and corresponding genes are involved. From these seven genes four of these have been cloned in *E. coli* [116, 118]. For the whole biosynthesis at least 30 biosynthetic and two known regulatory genes are involved, which encode around 35 intermediates. Furthermore, intracellular trafficking of intermediates between 7 compartments must also be considered, what can be considered as a major challenge in combinatorial biosynthesis [116].

The tryptophan decarboxylase and strictosidine synthase (STR) genes (Fig. 4) were the first two genes from *Catharanthus roseus* cloned from the monoterpenoid indole pathway into *S. cerevisiae*. In the past single genes of the biosynthesis have been expressed in different heterologous organisms [16]. The cDNA coding for STR from *R. serpentina* has previously been expressed in *E. coli* and in insect cells and was found to convert secologanin and tryptamine into strictosidine [118]. Recently, after feeding the precursors tryptamine and secologanin, strictosidine and its aglycon were biosynthesized in *S. cerevisiae* as a new heterologous host. When strictosidine glucosidase was additionally overexpressed in the
Combinatorial biosynthesis

recombinant host *S. cerevisae* carrying the tryptophan decarboxylase and strictosidine synthase gene, a sufficient amount of strictosidine was formed [119].

Besides in microbial hosts the mentioned genes of the early biosynthesis have also been cloned into *Nicotiana tabacum*. The major drawback however is the disability to hydrolyze strictosidine glucoside because *Nicotiana tabacum* does not possess specific glucosidases [120]. Later, strictosidine glucosidase has also been successfully inserted and expressed in suspension cultured tobacco cells [121]. The strictosidine glucosidase protein in *Nicotiana tabacum* was present in the same high molecular weight complexes as known before in *C. roseus*.

The late biosynthesis of vindoline and related monoterpenoidal indole alkaloids is only partly known and comprehensively summarized by Van der Heijden, 2004 [116]. Starting point in this phase is the strictosidine aglycon (Fig. 4) and its transformation via an unknown route to cathamine and tabersonine. From this precursor, vindoline is biosynthesized in at least six enzyme reactions and in multiple cellular compartments. All enzymes have been cloned and expressed in *E. coli* [122-125] and in part in *S. cerevisae* [119].
Fig. 4: Early biosynthesis of Vinca alkaloids in *S. cerevisiae* [116, 119]. (Abbreviations in the text).
Combinatorial biosynthesis of phenolic natural products

Flavonoids
Flavonoids represent a very important group of plant natural products. They are considered as health promoting substances in the human diet for their antioxidant, antiasthmatic, anti-blood-clotting, and anticancer activities. Flavonoids are exclusively produced in plants and found in almost all studied species in the plant kingdom. Flavonoids are produced via the so-called phenylpropanoid pathway, in which phenylalanine ammonia lyase (PAL) deaminates phenylalanine or tyrosine yielding cinnamic acid (Fig. 5). The biosynthetic route on the enzymatic and genetic level has been elucidated in the past [126, 127] and can be reconstructed in detail. The biosynthesis starts with L-phenylalanine that is metabolized to cinnamic acid derivatives, which condensates with malonyl-CoA to a chalcone. In the biosynthesis cinnamic acid is hydroxylated by cinnamic-4-hydroxylases (C4H) to para-4-hydroxy-cinnamic acid, activated by 4-coumarate/cinnamate coenzyme A, coupled with 3 malonyl-CoA units and converted by chalcone synthase (CHS) to a chalcone derivate as first committed precursor for the flavonoid biosynthesis. Chalcones are converted to flavonoids by a ring closing step forming the heterocyclic C ring by chalcone isomerases. Naringenin is a chalcone and key intermediate leading to isoflavones, to condensed tannin precursors and, via different hydroxylation, glycosylation, prenylation and alkylation steps, to more than 600 known flavonoids [127].
Recent publications have documented the production of pinocembrin, naringenin and chrysin, apigenin, galangin, kaempferol and dihydrokaempferol in recombinant E. coli BL21 (DE3). Because the main genes for flavonoids are missing in E. coli, recombinant plasmids (pUC, pET) containing the genes of interest have been constructed [128, 129]. These artificial gene clusters contain up to three genes from microorganisms or plant origin (Glycyrrhiza echinata, Petroselinum crispum and Citrus sinensis). Expression of all genes encoding the flavonoid biosynthesis up to the level of naringenin was successful, but only limited amounts of flavonoids were detected. To overcome this problem, the production of the essential precursor malonyl-CoA was increased by overexpression of the acetyl-Coa carboxylase from Corynebacterium glutamicum.
In recent publications from Miyahisa et al., 2005, further biosynthetic genes have been introduced to modify the oxygenation pattern of flavonoids leading to kaempferol and apigenin [10, 11]. The published work is of high interest, because for the first time a nearly complete biosynthetic pathway from plants was established in a heterologous microorganism. In the future it will be of interest to investigate whether further enzymes modifying flavones and flavonols like glycosylation, prenylation or O-methylation, can be integrated.
Fig. 5: Combinatorial biosynthesis of flavonoids and in *E. coli* [10, 11] (ScCCL: 4-coumarate/cinnamate coenzyme A; CHI: chalcone isomerase; other abbreviations in the text).
Conclusion

The concept of expressing genes from biosynthetic pathways in heterologous organisms has dramatically extended the possibilities for combinatorial biosynthesis. The purpose of this review is to highlight these recent developments in genetic engineering of heterologous microorganisms to reconstitute biosynthetic pathways from plants. At the start of this development only single genes were expressed for single enzyme characterization. By the development of genetic techniques multigene expression systems in hosts organisms like E. coli, B. subtilis or S. cerevisiae will be/became realistic to engineer whole pathways.

Today the flavonoid pathway has been fully transferred to E. coli as a host, and the early isoprenoid pathway up to farnesyl diphosphate has been fully reconstructed up to the level of linear precursors for the cyclization by different terpene cyclases. The example of heterologous biosynthesis of flavonoids and taxadien-5α-acetoxy-10β-ol documents clearly the development to construct multigene vectors and to express more than one gene.

The strategy for the future will be to identify a microbial host in which basic primary pathways can be exploited for the production of biosynthetic precursors for further secondary pathways. An advantage is that no transfer of genes and promoter sequences for a primary pathway is necessary – as known for the MVA and DOXP pathway– and that genes and expressed enzymes for the desired secondary pathway can just hook on. These systems can then be used for the production of valuable compounds or for further engineering strategies.