18 Vanillin Biosynthesis – Not as Simple as it Seems?

Richard A. Dixon

18.1 INTRODUCTION

Vanillin is the world's most popular flavor, and as such is probably the world's most popular plant natural product. It is also an extremely simple molecule. Why then, at a time when the biosynthesis of increasingly complex plant secondary metabolites is being elucidated at both the chemical and molecular genetic levels, should vanillin biosynthesis still be so controversial? Why do we know most of the steps involved in taxol biosynthesis (Heinig and Jennewein 2009), all of the steps involved in lignin (monolignol) biosynthesis (a pathway that share similarities to the vanillin pathway(s) (Humphreys and Chapple 2002), many of the steps involved in the formation of complex nitrogen-containing alkaloids (Kutchan 2002; Zeigler *et al*. 2006), but not how plants make 3-methoxy, 4-hydroxy-benzaldehyde? To be fair to the small body of researchers who have investigated vanillin biosynthesis, this question should probably be re-phrased to ask why we are still confused about the biosynthesis of most C_6–C_3 benzenoid derivatives in plants.

Vanillin is made in the "pods" of an orchid, *Vanilla planifolia*, a species that lacks genetic or genomic resources, and is stored as its 4-*O*-glucoside, glucovanillin. It is made in specialized cells within the pod (Joel *et al*. 2003), although there is still some disagreement as to exactly which cell types do or do not produce vanillin (Joel *et al*. 2003; Odoux and Brillouet 2009). The nature of the plant species and the restricted cellular location of its famous product should not present insurmountable problems for understanding vanillin biosynthesis, however, since many studies have addressed biosynthetic routes to more complex natural products through the application of molecular genetic approaches to specialized tissues in genetically recalcitrant plant species. Some of the best examples concern the biosynthesis of defensive compounds in glandular trichomes (Gang *et al*. 2002; Weathers *et al*. 2006; Nagel *et al*. 2008). My contention is that the simplicity of vanillin itself poses the major problem, because the structure lends itself to multiple theoretical biosynthetic pathways (Figure 18.1) and, because of a general promiscuity of many enzymes of plant phenolic metabolism, it is possible to find evidence to support any of these pathways from *in vitro* biochemical approaches. This certainly seems to be the case from a brief overview of the history of studies on the biosynthesis of vanillin and related compounds (Table 18.1), from which it is clear that our "understanding" of vanillin biosynthesis has not proceeded in a sequential manner. Rather, each new "advance" has provided an alternative model without effectively disproving existing models.

Handbook of Vanilla Science and Technology, *First Edition*, edited by Daphna Havkin-Frenkel and Faith C. Belanger

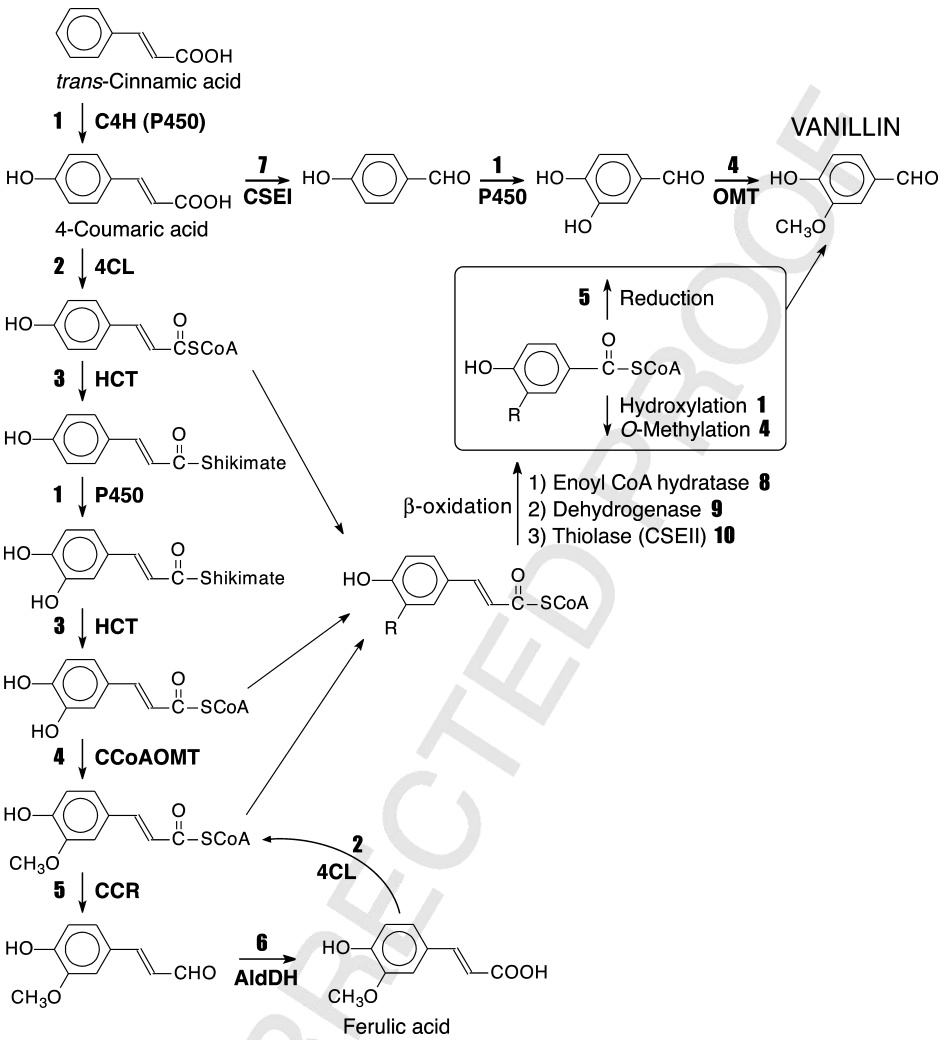

Fig. 18.1 Scheme of potential pathways to vanillin, in comparison to monolignol and ferulic acid formation. The pathway on the left-hand side of the figure shows the formation of ferulic acid from *trans*-cinnamic acid, according to recent studies on monolignol biosynthesis and the formation of ferulate in Arabidopsis. Vanillin is shown arising from two mechanistically different routes: directly from coumaric acid by non-oxidative chain shortening, or via any one of three Coenzyme A esters by β-oxidation. The numbers in circles represent different enzyme types that should be recognizable in EST datasets. Those involved in the formation of ferulate from cinnamate have all be functionally identified; it is assumed that similar types of enzymes (or even possibly the same enzymes) could be involved in the hydroxylation, O-methylation and reduction of benzoyl CoA or benzaldehyde intermediates.

18.2 MULTIPLE PATHWAYS TO VANILLIN?

Past work on the vanillin pathway, and pathways leading to related benzenoids, has been reviewed in more detail elsewhere (Dignum *et al*. 2001; Walton *et al*. 2003; Wildermuth 2006). It is generally agreed that vanillin is a product of the phenylpropanoid pathway from

Table 18.1 A timeline for the development of concepts related to vanillin biosynthesis

System and approach	Concept	Reference
Radiolabeling of *V. planifola* pods	Vanillin is formed directly from ferulic acid	Zenk 1965
Radiolabeling of *V. planifolia* tissue cultures	Intermediacy of isoferulic acid (which is subsequently demethylated)	Funk and Brodelius 1990a,b
Enzyme assay in cell free extracts from *Lithospermum erythrorhizon*	Non-oxidative chain-shortening of coumaric acid to 4-hydroxybenzaldehyde	Yazaki *et al*. 1991
Measuring metabolite levels in *V. planifolia* pods	Intermediacy of tartrate esters	Kanisawa *et al*. 1994
Enzyme isolation and assay from cell cultures of *Hypericum androaemum*	Involvement of a cinnamoyl CoA hydratase/lyase in non-oxidative chain shortening	El-Mawla *et al*. 2002
Enzyme isolation from *V. planifolia* cell cultures	Thiol-dependent non-oxidative conversion of 4-coumarate to benzaldehyde	Podstolski *et al*. 2002

L-phenylalanine, and that the hydroxyl group at the 4-position of the aromatic ring (*para* to the side chain) therefore originates through the action of the cytochrome P450 enzyme cinnamate 4-hydroxylase (C4H, Figure 18.1). This model is supported by labeling studies (Zenk 1965). The conversion of coumarate (4-hydroxycinnamate) to vanillin then "simply" requires four steps:

I shortening of the side chain by two carbons, catalyzed by a "chain shortening" enzyme or enzyme complex (CSE);
II introduction of the aldehyde function to the side chain (in some models this may occur as an integral part of chain shortening;
III introduction of the 3-hydroxyl group; and
IV 3-*O*-methylation (Figure 18.1).

Clearly the *O*-methylation reaction has to occur after the 3-hydroxylation, but these reactions could theoretically occur in any order. However, the number of possible theoretical pathways to vanillin is increased beyond three factorial by the fact that there is more than one mechanism for chain shortening of hydroxcycinnamic acids, and these lead to products with different oxidation states of the terminal group of the side-chain. Furthermore, if the model assumes a shared pathway to that involved in monolignol biosynthesis in which the first reactions are the ring modifications, additional reactions associated with formation of different types of ester intermediates could likely also be involved (Figure 18.1).

Similar complexities have been encountered in studies on related molecules. For example, salicylic acid (SA, 2-hydroxy-benzoic acid) was long thought to be synthesized through the phenylpropanoid pathway via *L*- phenylalanine, followed by chain shortening to a benzoic acid followed by subsequent 2-hydroxylation (Yalpani *et al*. 1993; León *et al*. 1995), and this was supported by genetic studies in which modification of expression of *L*-phenylalanine ammonia-lyase (the first enzyme of the phenylpropanoid pathway) gave disease response phenotypes predictably associated with modification of SA levels (Pallas *et al*. 1996). The subsequent demonstration that, at least in Arabidopsis, defense-associated SA formation occurs directly from the shikimate pathway via isochorismate (Wildermuth *et al*. 2001) came

as a total surprise. Similarly, recent labeling and genetic studies have demonstrated that the formation of benzoic acids in Petunia flowers occurs by multiple pathways involving both oxidative and non-oxidative chain shortening (Boatright *et al*. 2004; Orlova *et al*. 2006). This complexity makes it difficult to interpret labeling studies, particularly if (as in the case of most studies on vanillin to date) multiple tissue types are being labeled and the labeling is only carried out over a short period relative to the period of biosynthesis and accumulation. It has been argued that the existence of multiple pathways to benzenoid natural products within one plant might reflect a biological need for flexible responses to different environmental conditions (Wildermuth 2006). This is quite plausible, but it seems to the present author that constitutive vanillin biosynthesis during the development of the vanilla pod is more likely to occur via a single major pathway. The question is how to elucidate that pathway when enzyme promiscuity can mislead in vitro studies.

Early labeling experiments suggested that vanillin biosynthesis in plants occurs via ferulic acid, a molecule known to be synthesized via the phenylpropanoid/monolignol pathway (Zenk 1965). Although subsequent studies have suggested other alternatives (Table 18.1), it is instructive to consider this model for the formation of vanillin because it allows discussion of the types of enzymes that may be involved in the ring modification reactions, and their identification through functional genomics approaches.

At least in Arabidopsis, ferulate is formed from 4-coumarate by six enzymatic steps in a pathway, shared with monolignol biosynthesis, that is considerably more complex than envisaged at the time that the first labeling studies on vanillin biosynthesis were performed. The first step is the formation of a Coenzyme A ester through the action of 4-coumarate: CoA ligase (4CL), an enzyme generally encoded by multiple genes in plants (Ehlting *et al*. 1999) (Figure 18.1). The subsequent coumaroyl CoA ester is potentially a substrate for β-oxidative chain shortening (Figure 18.1) but, in the monolignol pathway, is directly converted to the corresponding shikimate ester by the action of hydroxycinnamoyl CoA: hydroxycinnamoyl transferase (HCT) (Hoffmann *et al*. 2003); it is this shikimate ester that undergoes hydroxylation of the aromatic ring at the 3-position by a second cytochrome P450 monooxygenase (Schoch *et al*. 2001). However, the subsequent 3-*O*-methylation does not happen at the shikimate ester stage; rather, the shikimate ester is converted back to the CoA ester through HCT acting in the reverse direction, and the resulting caffeoyl CoA is then methylated via caffeoyl CoA 3-*O*-methyltransferase (CCoAOMT) to yield feruloyl CoA. This compound is reduced to coniferaldehyde by the action of a cinnamoyl CoA reductase (CCR), another enzyme that is encoded by multiple genes in plants (Escamilla-Treviño *et al*. 2009). Finally, coniferaldehyde is converted to ferulic acid by the action of an aldehyde dehydrogenase (Nair *et al*. 2004) (Figure 18.1). It is important to note that detailed biochemical and genetic studies support the operation of this complex pathway over the simple mechanism whereby coumarate is converted to ferulate in two steps by 3-hydroxylation followed by 3-*O*-methylation, at least in dicotyledonous plants. However, early enzymatic work with crude and partially purified plant extracts did indeed suggest that this simpler pathway might operate.

The alternative and much simpler pathway to vanillin involves non-oxidative chain shortening. At least *in vitro*, 4-coumarate can be converted to 4-hydroxybenzaldehyde through a non-oxidative process requiring the presence of a thiol reagent but no other cofactor (Podstolski *et al*. 2002) (Figure 18.1), although no gene has yet been identified to encode this type of enzyme. Conversion to vanillin then simply requires 3-hydroxylation and *O*-methylation. Classical COMT enzymes are able to catalyze this methylation at the level of the benzaldehyde (Kota *et al*. 2004).

18.3 THE WAY FORWARD?

What did it take to establish the above pathway for ferulate formation? Interestingly, labeling experiments played only a small part. Rather, the major paradigm shifts came about through the application of genetic approaches in Arabidopsis, coupled with substrate specificity studies with recombinant enzymes. Unfortunately, *V. planifolia* does not appear to be a genetically tractable system at this stage. However, various tools of functional genomics that are now commonly applied to other systems might help throw light on the biosynthetic pathway, and some progress has already been made in this area (Pak *et al*. 2004; Havkin-Frenkel and Belanger 2007). The potentially applicable approaches center on gene expression profiling, both temporally and spatially.

The so-called "next generation" sequencing techniques (454 and Solexa/Illumina; www.454.com; www.solexa.com) have made it relatively simple to obtain massive expressed sequence tag datasets from relatively small amounts of tissue. Such EST datasets could easily be obtained from dissected tissues from vanilla pods throughout their period of development. As a control, similar datasets should be obtained from tissues shown not to accumulate significant amounts of vanillin, such as stems, roots, and leaves. After assembly and initial annotation of the sequences, the data can be mined for sequences matching the enzyme types predicted for involvement in vanillin biosynthesis based on all potential pathway models in Figure 18.1. Apart from the side chain shortening reaction (the most problematical part, as it is not immediately clear what the chain-shortening enzyme might look like), these will include aromatic hydroxylation and subsequent *O*-methylation, and possibly CoA ester reduction (analogous to CCR). It is more than likely that the hydroxylation reaction will be catalyzed by a cytochrome P450 enzyme, and that this will exhibit a significant degree of substrate specificity (Chapple 1998). Plant phenolic *O*-methyltransferases fall into two major classes, the type members being the so-called caffeic acid 3-*O*-methyltransferase (COMT, type I), which should properly be referred to as 5-hydroxyconiferaldehyde 3-*O*-methyltransferase based on its preferred substrate in the lignin pathway, and the type II CCoAOMT that is also involved in monolignol biosynthesis (Noel *et al*. 2003). Either type could potentially be involved in vanillin biosynthesis.

In contrast to most plant biosynthetic P450 enzymes, COMT is relatively promiscuous. In fact, the enzyme from alfalfa shows high activity against 3,4-dihydroxybenzaldehyde to form vanillin (Kota *et al*. 2004), although this is unlikely to be a function for the enzyme in alfalfa. Because vanillin accumulation occurs over a long time period, high activity may not be critical for candidate enzymes. For example, the formation of a major strawberry aroma compound involves the activity of a COMT, even though this enzyme is much more active with monolignol precursors than it is with the percursor of the 2,5-dimethyl-4-methoxy-3 (2H)-furanone flavor compound (Wein *et al*. 2002). Thus, *in vitro* biochemistry will ultimately need to be confirmed by either genetic approaches or detailed flux analysis measurements. Rapid techniques for reverse genetics based on virus-induced gene silencing are now being developed, and work well in some monocot systems (Lu *et al*. 2003; Ding *et al*. 2006). Likewise, techniques for precursor labeling and metabolic flux analysis are becoming increasingly sophisticated (Boatright *et al*. 2004).

Two factors are currently limiting the final assault on the vanillin pathway; the lack of a good experimental system (e.g. a highly inducible cell or tissue cultures) to simplify labeling experiments, and the lack of economic drivers to stimulate funding for this type of work. Pure vanillin is very cheap to produce synthetically but, at the same time, high value natural vanilla

flavor has to be extracted from the pods and is a complex mixture of natural products, among which vanillin predominates. There is currently no clear economic benefit from understanding how the vanillin molecule is assembled, since the idea of using such information to engineer the pathway, at least in *V. planifolia*, goes against the concept of natural vanilla, and synthetic vanillin is so cheap that introducing this molecule alone into other plants as a flavor component also does not make much economic sense. These factors should not, however, be used to argue against supporting research on vanillin biosynthesis. The pathways and mechanisms uncovered could in the future prove critical for the development of more complex bioactives in plants with applications in agriculture, food science, and biomedicine.

REFERENCES

Boatright, J., Negre, F., Chen, X. *et al.* (2004) Understanding *in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiol.*, **135**, 1993–2011.

Chapple, C. (1998) Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 311–343.

Dignum, M.J.W., Kerler, J. and Verpoorte, R. (2001) Vanilla production: technological, chemical, and biosynthetic aspects. *Food Rev. International*, **17**, 199–219.

Ding, X.S., Schneider, W.L., Chaluvadi, S.R., Mian, M.A.R. and Nelson, R.S. (2006) Characterization of a *Brome mosaic* virus strain and its use as a vector for gene silencing in monocotyledonous hosts. *Mol. Plant-Microbe Interact.*, **19**, 1229–1239.

Ehlting, J., Buttner, D., Wang, Q., Douglas, C.J., Somssich, I.E. and Kombrink, E. (1999) Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.*, **19**, 9–20.

El-Mawla, A.M.A.A. and Beerhues, L. (2002) Benzoic acid biosynthesis in cell cultures of ~~Hypericum androsaemum~~, *Planta*, **214**, 727–733.

Escamilla-Treviño, L.L., Shen, H., Uppalapati, S.R. *et al.* (2009). Switchgrass (*Panicum virgatum* L.) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties. *New Phytol.*, ~~in press~~.

Funk, C. and Brodelius, P.E. (1990a) Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia*. Andr. II. Effects of precursor feeding and metabolic inhibitors. *Plant Physiol.*, **94**, 95–101.

Funk, C. and Brodelius, P.E. (1990b) Phenylpropanoid metabolism in suspension cultures of *Vanilla planifolia* Andr. III. Conversion of 4-methoxycinnamic acids into 4-hydroxybenzoic acids. *Plant Physiol.*, **94**, 102–108.

Gang, D.R., Beuerle, T., Ullmann, P., Werck-Reichhart, D. and Pichersky, E. (2002) Differential production of *meta* hydroxylated phenylpropanoids in sweet basil peltate glandular trichomes and leaves is controlled by the activities of specific acyltransferases and hydroxylases. *Plant Physiol.*, **130**, 1536–1544.

Havkin-Frenkel, D. and Belanger, F.C. (2007) Application of metabolic engineering to vanillin biosynthetic pathways in *Vanilla planifolia*. In: *Applications of Plant Metabolic Engineering*. Verpoorte, R. (Ed.), Springer, pp. 175–196.

Heinig, U. and Jennewein, S. (2009) Taxol: A complex diterpenoid natural product with an evolutionarily obscure origin. *African J. Biotechnol.*, **8**, 1370–1385.

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. and Legrand, M. (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J. Biol. Chem.*, **278**, 95–103.

Humphreys, J.M. and Chapple, C. (2002) Rewriting the lignin roadmap. *Curr. Opin. Plant Biol.*, **5**, 224–229.

Joel, D.M., French, J.C., Graft, N., Kourteva, G., Dixon, R.A. and Havkin-Frenkel, D. (2003) A hairy tissue produces vanillin. *Israel J. Plant Sci.* **51**, 157–159.

Kanisawa, T., Tokoro, K. and Kawahara, S. (1994). In: *Olfaction Taste XI*, Proceedings of the International Symposium. Kurihara, K., Suzuki, N. and Ogawa, H. (Eds), Springer, Tokyo, p. 268.

Kota, P., Guo, D., Zubieta, C., Noel, J.P. and Dixon, R.A. (2004) *O*-Methylation of benzaldehyde derivatives by "lignin-specific" caffeic acid 3-*O*-methyltransferase. *Phytochemistry*, **65**, 837–846.

Kutchan, T.M. (2002) Sequence-based approaches to alkaloid biosynthesis gene identification. In: *Phytochemistry in the Genomics and Post-Genomics Eras*, Romeo, J.T. and Dixon, R. A. (Eds), Elsevier, Oxford, pp. 163–178.

León, J., Shulaev, V., Yalpani, N., Lawton, M.A. and Raskin, I. (1995) Benzoic acid 2-hydroxylase, a soluble oxygenase from tobacco, catalyzes salicylic acid biosynthesis. *Proc. Natl. Acad. Sci. USA*, **92**, 10,413–10,417.

Lu, R., Martin, H.A.M., Peart, J.R., Malcuit, I. and Baulcombe, D.C. (2003) Virus-induced gene silencing in plants. *Methods*, **30**, 296–303.

Nagel, J., Culley, L.K., Lu, Y. *et al.* (2008) EST analysis of hop glandular trichomes identifies an *O*-methyltransferase that catalyzes the biosynthesis of xanthohumol. *Plant Cell*, **20**, 186–200.

Nair, R.B., Bastress, K.L., Ruegger, M.O., Denault, J.W. and Chapple, C. (2004) The ~~Arabidopsis thaliana~~ REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell*, **16**, 544–554.

Noel, J.P., Dixon, R.A., Pichersky, E., Zubieta, C. and Ferrer, J.-L. (2003) Structural, functional, and evolutionary basis for methylation of plant small molecules. *Rec. Adv. Phytochem.*, **37**, 37–58.

Odoux, E. and Brillouet, J.-M. (2009) Anatomy, histochemistry and biochemistry of glucovanillin, oleoresin and mucilage accumulation sites in green mature vanilla pod (*Vanilla planifolia*; Orchidaceae): a comprehensive and critical reexamination. *Fruits*, **64**, 1–21.

Orlova, I., Marshall-Colon, A., Schnepp, J. *et al.* (2006) Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *Plant Cell*, **18**, 3458–3495.

Pak, F.E., Gropper, S., Dai, W.D., Havkin-Frenkel, D. and Belanger, F.C. (2004) Characterization of a multifunctional methyltransferase from the orchid *Vanilla planifolia*. *Plant Cell Rep.*, **22**, 959–966.

Pallas, J.A., Paiva, N.L., Lamb, C.J. and Dixon, R.A. (1996) Tobacco plants epigenetically suppressed in phenylalanine ammonia-lyase expression do not develop systemic acquired resistance in response to infection by tobacco mosaic virus. *Plant J.*, **10**, 281–293.

Podstolski, A., Havkin-Frenkel, D., Malinowski, J., Blount, J.W., Kourteva, G. and Dixon, R.A. (2002) Unusual 4-hydroxybenzaldehyde synthase activity from tissue cultures of the vanilla orchid *Vanilla planifolia*. *Phytochemistry*, **61**, 611–620.

Schoch, G., Goepfert, S., Morant, M. *et al.* (2001) CYP98A3 from ~~Arabidopsis thaliana~~ is a 3′-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J. Biol. Chem.*, **276**, 36,566–36,574.

Walton, N.J., Mayer, M.J. and Narbad, A. (2003) Vanillin. *Phytochemistry*, **63**, 505–515.

Weathers, P.J., Slkholy, S. and Wobbe, K.K. (2006) Artemisinin: The biosynthetic pathway and its regulation in *Artemisia annua,* a terpenoid-rich speices. *In Vitro Cell Dev Biol-Plant*, **42**, 309–317.

Wein, M., Lavid, N., Lunkenbein, S., Lewinsohn, E., Schwab, W. and Kaldenhoff, R. (2002) Isolation, cloning and expression of a multifunctional *O*-methyltransferase capable of forming 2,5-dimethyl-4-methoxy-3 (2H)-furanone, one of the key aroma compounds in strawberry fruits. *Plant J.*, **31**, 755–765.

Wildermuth, M.C. (2006) Variations on a theme: synthesis and modification of plant benzoic acids. *Curr. Opin. Plant Biol.*, **9**, 288–296.

Wildermuth, M.C., Dewdney, J., Wu, G. and Ausubel, F.M. (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, **414**, 562–565.

Yalpani, N., Leon, J., Lawton, M.A. and Raskin, I. (1993) Pathway of salicylic-acid biosynthesis in healthy and virus-inoculated tobacco. *Plant Physiol.*, **103**, 315–321.

Yazaki, K., Heide, L. and Tabata, M. (1991) Formation of *p*-hydroxybenzoic acid from *p*-coumaric acid by cell free extract of *Lithospermum erythrorhizon* cell cultures. *Phytochemistry*, **30**, 2233–2236.

Zenk, M.H. (1965) Biosynthese von vanillin in *Vanilla planifolia* Andr. *Z. Pflanzenphysiol.*, **53**, 404–414.

Ziegler, J., Voigtländer, S., Schmidt, J. *et al.* (2006) Comparative transcript and alkaloid profiling in *Papaver* species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. *Plant J.*, **48**, 177–192.