Published as: *Plant Physiol Biochem*. 2008 March ; 46(3): 356–370.

Structure and function of enzymes involved in the biosynthesis of phenylpropanoids

J.-L. Ferrer[a,*], **M.B. Austin**[b], **C. Stewart Jr.**[b], and **J.P. Noel**[b]

[a] Institut de Biologie Structurale, Commissariat à l'Energie Atomique, Centre National de la Recherche Scientifique, Université Joseph Fourier, Groupe Synchrotron, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France

[b] Howard Hughes Medical Institute, Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

Abstract

As a major component of plant specialized metabolism, phenylpropanoid biosynthetic pathways provide anthocyanins for pigmentation, flavonoids such as flavones for protection against UV photodamage, various flavonoid and isoflavonoid inducers of *Rhizobium* nodulation genes, polymeric lignin for structural support and assorted antimicrobial phytoalexins. As constituents of plant-rich diets and an assortment of herbal medicinal agents, the phenylpropanoids exhibit measurable cancer chemopreventive, antimitotic, estrogenic, antimalarial, antioxidant and antiasthmatic activities. The health benefits of consuming red wine, which contains significant amounts of 3,4′,5-trihydroxystilbene (resveratrol) and other phenylpropanoids, highlight the increasing awareness in the medical community and the public at large as to the potential dietary importance of these plant derived compounds. As recently as a decade ago, little was known about the three-dimensional structure of the enzymes involved in these highly branched biosynthetic pathways. Ten years ago, we initiated X-ray crystallographic analyses of key enzymes of this pathway, complemented by biochemical and enzyme engineering studies. We first investigated chalcone synthase (CHS), the entry point of the flavonoid pathway, and its close relative stilbene synthase (STS). Work soon followed on the *O*-methyl transferases (OMTs) involved in modifications of chalcone, isoflavonoids and metabolic precursors of lignin. More recently, our groups and others have extended the range of phenylpropanoid pathway structural investigations to include the upstream enzymes responsible for the initial recruitment of phenylalanine and tyrosine, as well as a number of reductases, acyltransferases and ancillary tailoring enzymes of phenylpropanoid-derived metabolites. These structure–function studies collectively provide a comprehensive view of an important aspect of phenylpropanoid metabolism. More specifically, these atomic resolution insights into the architecture and mechanistic underpinnings of phenylpropanoid metabolizing enzymes contribute to our understanding of the emergence and on-going evolution of specialized phenylpropanoid products, and underscore the molecular basis of metabolic biodiversity at the chemical level. Finally, the detailed knowledge of the structure, function and evolution of these enzymes of specialized metabolism provide a set of experimental templates for the enzyme and metabolic engineering of production platforms for diverse novel compounds with desirable dietary and medicinal properties.

*Corresponding author. jean-luc.ferrer@ibs.fr (J.-L. Ferrer).

Keywords

Phenylpropanoid pathway; Secondary metabolism; Lignin; Flavonoids; Anthocyanins; Plant enzymes; Structure

1. Introduction

Around 450 million years ago, pioneering plant species spread out from the periphery of their watery environments to occupy a challenging new environmental niche: dry land. This important transformation was accompanied by several physiological adaptations, including the evolutionary emergence of entirely new specialized metabolic pathways [66]. Many of the resulting compounds were derived from phenyl-alanine via the general phenylpropanoid pathway, so named because of the C6 + C3 scaffold resulting from the pathway's initial deamination of the primary metabolite phenylalanine. A variety of downstream derivatives of this initial phenylpropanoid scaffold, ubiquitous in modern plants, now serve vital developmental roles in plant structural integrity, UV photoprotection, reproduction and internal regulation of plant cell physiology and signaling. Phenylpropanoids also serve as key chemical modulators of plant communication with insects and microbes, playing contrasting attractive and repellent roles in defensive phytoalexin responses to infection and herbivory, attraction of insect pollinators via flower color, and induction of root nodulation by symbiotic nitrogen-fixing rhizobial colonies.

While some phenylpropanoid-mediated interspecies communication roles are clearly more recent environmental adaptations, two biological applications are especially relevant to the early emergence and adaptation of plants to dry land, and are thus likely to have played key roles in this 'watershed' event in the explosion of terrestrial plant and animal diversity. Lignin is an abundant structural polymer formed from monolignol derivatives of the general phenylpropanoid pathway, and, along with cellulose, provides the structural integrity necessary for the evolutionary emergence of self-supporting structures. While lignin biosynthetic pathways facilitated the upward vertical expansion of land plants into the sky, another biochemical offshoot of the general phenylpropanoid pathway known as flavonoids provided protection against the damaging ultraviolet rays of the sun during the transition of aquatic plants from UV-absorbing water onto land and the unshielded atmosphere. Flavonoids and related stilbene and styrylpyrone polyphenolic natural products all stem from polyketide extension and subsequent cyclization of general phenylpropanoid substrates by chalcone synthase (CHS) or related plant polyketide synthases (PKSs). Although the UV-protective properties of flavonoids undoubtedly facilitated the emergence of land plants, Stafford has proposed that the complex functional requirements for effective UV-protection, including abundant flavonoid biosynthesis and surface localization, make this function unlikely to be the original physiological role of the flavonoid pathway [63]. While some internal signaling role, such as that suggested in auxin regulation [5], would indeed have required less robust initial flavonoid biosynthesis, this argument neglects that plant movement onto land would seem most likely to have been preceded by gradually increasing fitness at the surface or in dynamically fluctuating edges of the water, perhaps even within the shade of geological structures.

Although the first advantageous physiological role of phenylpropanoids remains obscure, the recent explosion of crystal structures of phenylpropanoid biosynthetic enzymes illuminates another interesting aspect of the evolutionary emergence of new biosynthetic pathways: the recruitment of enzymes from existing primary or other specialized metabolic pathways. Although primary sequence similarity provides the structural fold or enzyme class of most phenylpropanoid-metabolizing enzymes, elucidation of actual three-dimensional

structures often illuminates evolutionary relationships more precisely, as secondary and tertiary structural similarity typically persist much longer than primary sequence similarity. For example, comparison of the first phenylpropanoid pathway enzyme crystal structure, that of alfalfa CHS [15], to several distantly related fatty acid synthase (FAS) structures revealed a significantly closer similarity to a particular primary metabolic enzyme of plant and bacterial fatty acid synthesis, in turn shedding light upon the specific functional transformations of the synthase relevant to the evolutionary emergence of plant polyketide biosynthesis (reviewed in [2]).

In this review, we consolidate insights from the many subsequent phenylpropanoid enzyme crystal structures that have been determined in the last decade (Table 1), and their structural similarities with other enzymes both within and outside phenylpropanoid metabolism, in order to arrive at a more expansive perspective into the dynamic evolutionary processes of enzyme recruitment and rapid mutational refinement and selection to form new (or more elaborate) specialized biosynthetic pathways.

2. Branching out from primary metabolism: the general phenylpropanoid pathway

In most species that possess the general phenylpropanoid pathway, three enzymatic transformations redirect the carbon flow from primary metabolism, transforming phenylalanine into the Coenzyme A (CoA)-activated hydroxycinnamoyl (phenylpropanoid) thioester capable of entering the two major downstream pathways, monolignol and flavonoid biosynthesis (Fig. 1A). These plant-ubiquitous initial steps constitute the general phenylpropanoid pathway. Deamination by phenylalanine ammonia-lyase (PAL, Fig. 2A) forms the phenylpropanoid skeleton, producing cinnamic acid. Cinnamic acid 4-hydroxylase (C4H) catalyzes the introduction of a hydroxyl group at the *para* position of the phenyl ring of cinnamic acid, producing *p*-coumaric acid. The carboxyl group of *p*-coumaric acid is then activated by formation of a thioester bond with CoA, a process catalyzed by *p*-coumaroyl:CoA ligase (4CL). Notably, grasses and some species of fungi and bacteria possess a tyrosine ammonia lyase (TAL) that directly uses tyrosine as a substrate, thus reducing the number of enzymes that are necessary for the production of *p*-coumaroyl-CoA from three in the general phenylpropanoid pathway to two [25]. Although not discussed in this review, various general phenylpropanoid pathway intermediates are also diverted into biosynthetic pathways for benzoic acid, salicylic acid, and coumarins.

Because of its key role in secondary phenylpropanoid metabolism, PAL has been extensively studied. It is also the only general phenylpropanoid pathway enzyme for which detailed structure–function information is available. The first structures of bacterial and plant PALs were solved in 2004 [10,55]. Other bacterial and fungal PAL/TAL structures were elucidated more recently [45,47]. These structures revealed that the PAL/TAL catalytic sites contain an unusual electrophilic 4-methylidene-imidazole-5-one prosthetic group, a covalent cofactor derived in two autocatalytic dehydration reactions from an internal tripeptide sequence. The enzyme structurally resembles histidine ammonia-lyase (HAL) from the more widely distributed general His degradation pathway [57]. The plant PAL contains 207 additional residues with respect to HAL, mainly restricted to an N-terminal extension rigidifying a domain interface and an inserted alpha-helical domain restricting access to the active site. Presumably, PAL developed from HAL when fungi and plants diverged from the other kingdoms. Despite the current evolutionary distance between these related ammonia-lyases, only minor active site differences are necessary to modulate aromatic amino acid substrate specificity [45,67].

The second enzyme in the general phenylpropanoid pathway is a member of a ubiquitous enzyme family of oxygenases known as the Cytochrome P450 hydroxylases (P450s), so named because of the characteristic absorption (at 450 nm) of their catalytic iron-containing heme cofactors [48]. As discussed below, P450 monooxygenases play important tailoring roles in both major downstream pathways of phenylpropanoid metabolism, but due to their extreme insolubility when purified or overexpressed in heterologous systems, no plant P450 monooxygenase has thus far been structurally characterized. Notably, these phenylpropanoid-metabolizing P450s are hypothesized to serve as membrane-associated anchor sites for the putative multi-protein complexes formed between co-localizing phenylpropanoid biosynthetic enzymes [72]. Structural elucidation of similarly membrane-associated mammalian hepatic P450s was facilitated by mutation and deletions in primary sequence guided by homology to an earlier crystal structure of a soluble P450 from a bacterium [70]. While homology modeling of plant P450s based upon these related structures may provide clues as to the functional diversity of plant P450s in general, a more detailed and predictive understanding of their roles in phenylpropanoid biosynthetic pathways awaits similar biochemical treatment and subsequent structural elucidation. While substitution of TAL for PAL allows metabolic engineers to bypass this initial inconveniently insoluble P450 pathway enzyme [25], subsequent P450-catalyzed steps in downstream monolignol and flavonoid pathways are not so easily circumvented (see below).

The final thioester-activation step of the general phenylpropanoid pathway (Fig. 1A) is catalyzed by a member of the ubiquitous AMP-producing adenylating superfamily of enzymes (AAE), each of which utilizes an adenylated intermediate formed by condensation of AMP with a carboxylate-bearing small molecule substrate via loss of pyrophosphate from ATP. AMP is subsequently released during covalent transfer of the adenylated substrate to an attacking nucleophilic third substrate, typically CoA or an equivalent phospho-pantetheine moiety. Although no 4CL crystal structure has been published, several related AAE superfamily structures are known, among them the fatty acyl [24] and acetyl:CoA ligases [21] of primary metabolism from which plant phenylpropanoid:CoA ligases presumably evolved. The ‘4’ in 4CL refers to the lone hydroxyl at the *para* (or four) position of the *p*-coumaroyl phenylpropanoid ring, but some 4CL isoenzymes prefer more bulky phenylpropanoid substrates with additional ring substituents, as we shall discuss in the next section on monolignol biosynthesis.

3. Monolignol biosynthesis and associated downstream metabolites: Lignin, lignans and phenylpropenes

Lignin is a principal structural component of cell walls in higher terrestrial plants and, after cellulose, the second most abundant plant polymer. Lignin is composed of phenylpropanol units (monolignols) oxidatively coupled through ether and carbon-carbon linkages. This natural polymer can function as a genetically inducible physical barrier in response to microbial attack [4,23,27,39,49]. In addition to structural support and pathogen defense, lignin functions in water transport as a hydrophobic constituent of vascular phloem and xylem cells [26,56]. In angiosperms, lignin is composed of two major units, the guaiacyl (G) and syringyl (S) units, derived from their corresponding monolignol precursors, the coniferyl and sinapyl alcohols. G units are singly methylated on the 3-hydroxyl group, whereas S subunits are methylated on both the 3- and 5-hydroxyl moieties (Fig. 1B). The ratio of S-to G subunits in lignin dictates the degree and nature of polymeric cross-linking. Increased G content leads to more highly crosslinked lignin featuring a greater proportion of biphenyl and other carbon-carbon bonds, whereas S subunits are typically linked through more labile ether bonds at the 4-hydroxyl position [14,22,42]. The type and extent of lignin polymerization has profound industrial, environmental, and agricultural consequences [26,35,38,53,54]. Lignin formation is generally believed to involve essentially random

polymerization of activated monolignol radicals. Although not fully understood, some evidence suggests that individual monolignol activation can occur either through the catalytic action of laccases, peroxidases, or both (discussed in [37]).

Monolignols are also joined using 8-8′, 8-5′ and 8-*O*-4′ linkages to produce lignans, a large class (several thousand known members) of typically dimeric phenylpropanoid oligomers. Built mostly from coniferyl alcohol, lignan subunits are first activated by a laccase or peroxidase to produce free radical intermediates that are sometimes then pre-oriented by 'dirigent' proteins to facilitate stereospecific coupling [18]. While plants biosynthesize lignan for defense, dietary lignans also play important nutritional roles in maintaining animal and human health. During digestion, intestinal bacteria convert the plant lignans secoisolariciresinol and matairesinol to entero-diol and enterolactone, respectively. Enterodiol and enterolactone are believed to reduce the onset and rate of prostate and breast cancers. This protection accrues to individuals consuming diets rich in grains, vegetables and berries that contain high concentrations of secoisolariciresinol and matairesinol. Unfortunately, typical Western diets tend to poorly represent these foods and do not afford comparable protection [40].

Finally, monolignols also serve as intermediates en route to non-polymeric, aromatic and volatile phenylpropenes such as eugenol, a major contributor to the scent of basil. Eugenol is biosynthesized from coniferyl alcohol in two enzymatic steps, via an intermediate ester linkage with acetate [36].

Despite intensive research, physiological regulation of the relative rates of biosynthesis and the spatial and temporal availability of these monolignol units (*p*-coumaryl, coniferyl and sinapyl alcohols) remains somewhat poorly understood, in part due to the existence of a complex set of related enzymes and a highly intersecting metabolic grid leading to the various monolignol building blocks. While a few details of monolignol ratio regulation remain obscure, the general biosynthetic scheme of the major contributors to plant lignin now appears largely complete (Fig. 1B). Until recently, this pathway was thought to consist of a two-dimensional metabolic grid proceeding from *p*-coumarate with increasing (optional) phenyl ring substitution via hydroxylation and subsequent o-methylation in one dimension, and (mandatory) stepwise activation and reduction of the propanyl tail to a primary alcohol, via CoA-thioester and aldehyde intermediates in the other dimension. Alternative paths through this metabolic grid might be temporally modulated by selective gene expression, or by expression of isoenzymes with alternative substrate specificities sometimes in a species specific manner. For example, this model predicts that increased expression of 4CL isoenzymes with higher relative affinity for sinapate (Fig. 1B) should increase the S-content of resulting lignin. Testing and refinement of this grid has proven difficult, as in vitro substrate preferences of isolated enzymes do not necessarily reflect physiological turnover in the context of a metabolic grid. Moreover, genetic analysis is complicated in most plants by multiple genes encoding similar biosynthetic enzymes, often including some relatives of unknown function [60].

Notwithstanding these difficulties, magnified in some cases by the lack of quantitative biochemical characterization, it has become increasingly clear that only portions of this metabolic grid are relevant to most cases of plant physiology. Quite recently an entirely unanticipated third critical dimension of the metabolic grid was identified and characterized by multiple research groups (reviewed in [8], Fig. 1B), revealing that P450-catalyzed hydroxylation of the phenylpropanoid C3 position occurs after transfer, catalyzed by a hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase (HCT), of *p*-coumarate from a thioester linkage with CoA to an ester linkage with shikimate or quinate. This physiologically predominant route to C3 hydroxylation remained obscure due to the HCT-

catalyzed complementary reverse transfer of the C3-hydroxylated caffeoyl-ester product back onto CoA for subsequent transformations. In a further complication, some sinapyl derivatives are nonetheless biosynthesized in a genetic knockout of the shikimate or quinate ester-specific P450 hydroxylase [1].

While the recent discovery of this 'hidden' dimension of monolignol biosynthesis is a cautionary note against over-interpretation of isolated in vitro data, structural elucidation of several monolignol pathway enzymes has contributed to our understanding of enzyme-mediated substrate flux through this metabolic grid (see below). We will address recent structure elucidation of enzymes catalyzing steps after the generation of monolignol units, namely eugenol synthase (EGS) and two lignan reductases (Fig. 1C), but not the poorly understood formation of monolignol radicals leading to lignin and lignans. We will also not discuss in any more detail two additional structurally uncharacterized P450 enzymes (C3H and F5H in Fig. 1B), reminiscent of C4H in the general phenylpropanoid pathway and similar processes in the flavonoid biosynthetic branches (see below and Fig. 1D), that catalyze important hydroxylation reactions constituting half of the ring-substitution reactions in one dimension of the monolignol metabolic grid. Likewise, we do not further consider cinnamoyl-CoA:NADPH oxidoreductase (CCR), a short-chain dehydrogenase/reductase (SDR) superfamily member (discussed below) whose structure is not yet available (Fig. 1B).

Two *S*-adenosyl-L-methionine (SAM)-dependent *O*-methyl-transferases (OMTs) methylate the phenyl ring of monolignol precursors: caffeic acid 3-OMTs (COMTs) and caffeoyl-CoA OMTs (CCoAOMTs). COMT methylates caffeoyl- and 5-hydroxyferuloyl-containing acids, aldehydes and alcohols in vitro while displaying a kinetic preference for the alcohols and aldehydes over the free acids. The 2.2 Å crystal structure of COMT in complex with *S*-adenosyl-L-homocysteine (SAH) and ferulic acid (ferulate form), as well as the 2.4 Å crystal structure of COMT in complex with SAH and 5-hydroxyconi-feraldehyde, were reported in 2002 (Fig. 2B), providing a structural understanding of the previously reported in vitro substrate preferences [78].

In contrast, plant CCoAOMTs accept only the CoA-ester as a substrate. CCoAOMTs belong to a distinct family of plant small molecule OMTs, more closely related to the mammalian catechol OMTs than to other plant OMTs (Fig. 3A). The crystal structures of alfalfa (*Medicago sativa*) CCoAOMT in complex with the reaction products *S*-adenosine-L-homocysteine and feruloyl/sinapoyl-CoAs (Fig. 2C), were first reported in 2005 [16]. These structures provided a new understanding of the substrate preferences and the role of divalent cations in the catalytic mechanism accompanying CCoAOMT-mediated O-methylation of CoA-linked phenylpropanoid substrates.

Another family of structurally characterized enzymes involved in the lignin biosynthetic pathway includes the cin-namyl/sinapyl alcohol dehydrogenases (CAD/SAD). These reductases are NADP(H)-dependent dehydrogenases that catalyze the last reductive step in the formation of monolignol alcohols.

The first, and up to now unique, three-dimensional structure of sinapyl alcohol dehydrogenase (SAD, Fig. 2D), from *Populus tremuloides* (aspen), was described in 2005 [7]. The active site topology revealed by the crystal structure substantiates kinetic results indicating that SAD maintains highest specificity for the substrate sinapaldehyde, as well as substantial substrate inhibition kinetics for the SAD-catalyzed reduction of hydroxycinnamaldehydes. However, kinetic analyses of wild-type SAD and several active site mutants demonstrate the complexity of defining determinants of substrate specificity in these enzymes. These results, along with a phylogenetic analysis, support the inclusion of SAD in a plant alcohol dehydrogenase subfamily that includes cinnamaldehyde and

benzaldehyde dehydrogenases. The SAD three-dimensional structure was used to model several of these SAD-like enzymes, and although their active site topologies largely mirror that of SAD, it was possible to draw correlations between substrate specificity and amino acid substitution patterns in their active sites [7].

In *Arabidopsis thaliana*, two members of the cinnamyl alcohol dehydrogenase (CAD) multigene family, AtCAD4 and AtCAD5, display the highest level of in vitro catalytic activity and appear to be largely involved in the formation of guaiacyl/syringyl lignins. The crystal structures of AtCAD5 in the apo-form and as a binary complex with NADP+, were reported in 2006 (Fig. 2E), and used to model the structure of AtCAD4. Both AtCAD5 and AtCAD4 behave as dimers with two Zn^{2+} ions per subunit and belong to the Zn^{2+}-dependent medium chain dehydrogenase/reductase (MDR) superfamily [76].

Although SAD and CAD catalyze the same reaction and share a reasonable level of sequence identity, the active site topology is strikingly different. Indeed, the substrate-binding pockets of both AtCAD4 and AtCAD5 were found to be significantly smaller compared to that of SAD from aspen (Fig. 3D). However, not unlike many specialized metabolic enzymes in plants belonging to extended families of highly similar genes, SADs and CADs possess a high degree of similarity in their overall 3D structures (Fig. 3C). Analysis of CAD sequences from various species revealed that nine out of the twelve residues which constitute the proposed substrate-binding pocket are conserved. This observation is indicative of a characteristic active site fingerprint for the CAD family and suggests that enzymes such as SAD, while evolutionarily related, may constitute more specialized variants in particular plant species [76].

Recent investigation into the biosynthetic pathways to the lignans [19,74] resulted in the isolation and characterization of several lignan reductases, including pinoresinol–lariciresinol reductases (PLRs) [13,17] and phenylcoumaran benzylic ether reductases (PCBERs) [19] (Fig. 1C). The three-dimensional X-ray crystal structures of both PLR from *Thuja plicata* (Fig. 2G) and PCBER from *Pinus taeda* (Fig. 2H) were solved and refined to 2.5 and 2.2 Å resolutions, respectively. Both enzymes share a high degree of amino acid sequence similarity and correspondingly similar 3D structures (Fig. 3B), possessing continuous α/β NADPH-binding domains and smaller substrate-binding domains [46]. Through careful comparison of the two reductases, the basis for enantiospecific and regiospecific hydride transfer was deduced.

As mentioned above, the monolignol coniferyl alcohol central to lignan biosynthesis is also subject to acetyl-CoA dependent acetylation in some plants, resulting in at least two recently characterized cases in the biosynthesis of the phenylpropenes eugenol and isoeugenol. The 3D structure of basil EGS was recently elucidated as a ternary complex with NADP+ and a substrate analog bound [44] (Fig. 2F). Structurally, EGS is a member of the short-chain dehydrogenase/reductase (SDR) family shared with PLR, PCBER and isoflavone reductase (IFR; discussed below). EGS consist of a series of parallel β-sheets flanked on each side by a layer of α-helices forming a Rossman fold domain. Although this Rossman fold, which is involved in binding NADPH, is well conserved in most NADPH-dependent reductases, the C-terminal domain of EGS contains divergent structural elements which may account for the emergence of novel substrate binding and mechanistic activity.

4. Flavonoid biosynthesis

Flavonoids are a large (≈9000) structurally diverse class of phenolic compounds found in all higher plants. In addition to their physiological roles in the biochemical ecology of plants, flavonoids played a key role in the discovery of fundamental biological phenomena including Gregor Mendel’s discovery of the laws of heredity, Barbara McClintock’s

discovery of transposable elements and the independent discovery by several groups of co-suppression/RNAi [73]. All flavonoids are derived from the chalcone scaffold, which is biosynthesized by the ubiquitous plant enzyme chalcone synthase (CHS). CHS catalyzes the iterative condensation and subsequent intramolecular cyclization of three acetate units onto the *p*-coumaroyl-CoA end product of the general phenylpropanoid pathway. Following CHS, an assemblage of isomerases, reductases, hydroxylases, glycosyltransferases and acyltransferases emerged to decorate the basic flavonoid skeleton leading to a variety of flavonoid chemical subclasses (Fig. 1D). These tailoring enzymes result in the production of a vast array of compounds that can generally be classified into one of the following groups: flavones, isoflavanoids, flavonols, flavandiols, proanthocyanidins and anthocyanins.

The presence of a C2–C3 double bond significantly alters the conformation and reactivity of flavonoid molecules and differentiates a flavanone (double bond absent) from a flavone (double bond present). Isoflavonoids constitute a structurally distinct class of flavonoids found mainly in legumes and are defined by a dramatic C2–C3 aryl ring migration and concomitant double bond formation catalyzed by isoflavone synthase. The addition of hydroxyl groups to core flavonoid rings leads to the formation of the flavonols and flavandiols, which serve as entry points into the formation of proanthocyanidins and anthocyanins. Proanthocyanidins arise by catalytic cross-linking of flavandiols to form oligomeric units that vary in the number of linked flavonoid units. Alternatively, the flavandiols can undergo glycosylation and methylation to form various anthocyanins.

The CHS condensation mechanism is similar to that of fatty-acid synthases. After condensing three acetate units derived from decarboxylative condensation of malonyl-CoA onto *p*-coumaroyl-CoA, CHS forms a chalcone via an intramolecular cyclization of the resulting covalently linked linear tet-raketide intermediate. A closely related enzyme, stilbene synthase (STS), catalyzes a similar set of extension reactions but performs a divergent cyclization reaction to produce stilbenes from the same starting materials. The phenylpropanoid-utilizing STS and CHS enzymes are central members of a larger superfamily of enzymes (type III PKSs), most of which catalyze similar reactions, but with evolutionarily optimized specificities for various other (non-phenylpropanoid) substrates [3]. The crystal structure of a CHS from alfalfa (*Medicago sativa*) revealed that CHS maintains an αβαβα protein fold conserved among all β-ketoacyl synthases and thiolases including primary metabolic enzymes of fatty acid biosynthesis and degradation [15] (Fig. 2K). The structure of STS (Fig. 2L) reveals a similar thiolase fold as well as a cryptic thioesterase activity which switches the cyclization reaction from a Claisen condensation (CHS) to an aldol condensation (STS) [2]. The structural and biochemical evidence clearly indicates that type III PKSs such as CHS/STS evolved from a non-iterative β-ketoacyl ACP synthase (KASIII) involved in fatty acid biosynthesis (reviewed in [3]).

Chalcone isomerase (CHI) stereospecifically directs and greatly accelerates the spontaneous additional cyclization of chalcones to form the flavonoid core (flavones). Structurally, CHI possesses a unique open-faced β-sandwich fold (Fig. 2I) and was originally thought to exist only in plants [29]. The assumed exclusivity of CHI to plants has been challenged by more recent sequence and structural comparisons of genes in diverse genomes which indicate that structural homologs of CHI likely exist in fungi, slime molds and gammapro-teobacteria [20]. While in some of these cases the catalytic residues are highly conserved, many other of these diverse ‘CHIs’, including some in plants, possess highly divergent predicted active sites suggesting a multitude of currently unrecognized roles in plants, fungi and bacteria. Interestingly, many species containing these CHI homologs do not contain CHS; thus the physiological substrate that CHI would act upon is unclear, and in many cases the lack of key catalytic residues suggests some may not even act as enzymes [29]. Because land plants are thought to have originated from green algae, further research on CHI activity in

chlorophytes and charophytes is needed to shed light on the origin of flavonoid biosynthesis that is set in motion by the combined catalytic actions of CHS and CHI.

One class of enzymes that plays an important role in decorating the flavonoid backbone is the NADPH-dependent reductases. Chalcone reductase (CHR) has only been identified in legumes, and its activity forms a metabolic branch point that allows these related plant species to synthesize isoflavonoids. In addition to acting as precursors for various defense chemicals, isoflavonoids also serve as signaling molecules to induce nitrogen fixation via root nodulation by *Rhizobia* bacteria. CHR, a ketoreductase, acts on an intermediate of the CHS reaction resulting in the creation of a 6′-deoxychalcone. X-ray crystallographic analysis of CHR revealed an overall structure consisting of alternating α-helix and β-sheet secondary structure elements to form a well-characterized TIM-barrel protein fold [6] (Fig. 2J). As suggested by earlier sequence comparisons [69], CHR belongs to the aldo-keto reductase superfamily [28], and thus is more closely related to carbohydrate reductases than to analogous ketoreductase domains of fatty-acid and polyketide synthases.

In contrast to CHR, several other flavonoid NADPH-dependent reductases are members of the short-chain dehydrogenase/reductase (SDR) superfamily. Isoflavone reductase (IFR, Fig. 2P) and vestitone reductase (VR, Fig. 2Q) are specific to legumes and contain a typical Rossman fold in their N-terminal domain [58,65]. Both IFR and VR are key enzymes deployed during the biosynthesis of the phytoalexin medicarpin. IFR reduces the 2,3 double bond of the central ring of isoflavonoids creating an isoflavanone intermediate. VR is key to the subsequent formation of enantiomeric pterocarpans such as medicarpin. Dihydroflavonol 4-reductase (DFR) catalyzes a central flavonoid pathway transformation in the biosynthesis of proanthocyanidins and anthocyanins which are found broadly throughout the plant kingdom. DFR converts dihydroflavonols into flavandiols by reducing the ketone group on the central ring. Structurally similar to IFR and VR, DFR also contains an N-terminal domain Rossman fold and a more variable substrate binding region in the C-terminal domain [52] (Fig. 2M). Anthocyanidin reductase (ANR), a newly discovered enzyme [75] of the flavonoid pathway involved in the biosynthesis of condensed tannins, is yet another SDR-superfamily enzyme. Although not yet published, the structure of ANR was recently elucidated by molecular replacement (Fig. 2O), using the structure of DFR as a template (Gargouri, personal communication). It is worth noting that aldo-keto reductases (e.g. CHR) and short-chain dehydrogenase/reductases (e.g. IFR, DFR) utilize the same network of active-site residues for catalysis, which indicates convergent evolution between these otherwise dissimilar protein families [28,34].

In addition to these reductases, several dioxygenases and cytochrome P450s act to oxidize core flavonoid intermediates into one of the many end products found throughout the plant kingdom. Anthocyanidin synthase (ANS), an oxoglutarate-dependent $Fe^{2+/3+}$ dioxygenase, catalyzes the penultimate step in anthocyanin biosynthesis. The crystal structure of ANS reveals an interior region consisting of several parallel β-strands arranged as a double-stranded β-helix or ‘jellyroll’ topology [68,71] (Fig. 2N). Interestingly, two flavone synthases exist in plants with one being a dioxygenase (FNSI) and the other a cytochrome P450 hydroxylase (FNSII). FNSI activity has only been identified within the Apiaceae plant family. Although crystal structures are not available, the FNSI primary sequence is more similar to ANS than to any of the P450s, raising interesting questions about the evolutionary underpinnings of the convergent FNSI and FNSII reactions.

Lastly, flavonoid biosynthesis exploits a suite of enzymes that decorate the flavonoid backbone with sugars, methyl groups and acyl moieties (Fig. 1D). These modifications often modulate the physiological activity of the resulting flavonoid by altering solubility, reactivity and interaction with cellular targets. Chalcone *O*-methyltransferase (ChOMT) and

(iso)flavone *O*-methyltransferases (IOMTs) use the universal methyl donor *S*-adenosyl-L-methionine (SAM) to regiospecifically methylate hydroxyl moieties. Consequently, both enzymes belong to a family of methyltransferases that possess a conserved SAM binding domain α/β Rossman fold consistent with their nucleotide binding activity [43,77]. ChOMT and IOMT (Fig. 2R) differ from other structurally characterized methyl-transferases in that they exist as homodimers, with a novel helical domain forming a dimerization interface that is unique to plant OMTs and which also forms a major portion of the chalcone/(iso)flavonoid substrate binding site.

At various stages of flavonoid biosynthesis sugar residues are attached to the flavonoid core via a uridine diphosphate glycosyltransferase (UDG, Fig. 1D). UDGs transfer an UDP-activated sugar donor molecule onto an acceptor molecule. Three UDGs involved in flavonoid biosynthesis have been structurally characterized. UDP-flavonoid/triterpene glycosyl-transferase (UGT71G1, Fig. 2T) transfers glucose monomers onto triterpenes, flavonols, and isoflavones in in vitro assays [59]. UDP-glucose:flavonoid 3-*O*-glycosyltransferase (VvGT1, Fig. 2U) transfers glucose units onto flavonols [50]. (Iso)flavonoid glycosyltransferase (UGT85H2) is a multi-functional enzyme which can transfer sugar units onto flavonols, isoflavones, and chalcones [41]. All three of the above-mentioned UDGs contain two Rossman-like folds and belong to the GT-B fold superfamily. These UDGs show activity with a range of donor and acceptor molecules in vitro; consequently the cellular context of these enzymes needs to be investigated further to identify in vivo function.

Once attached, these sugar molecules can undergo acylation via the action of acyltransferases. The crystal structure of a malonyl anthocyanin acyltransferase (Dm3MAT3, Fig. 2S), which transfers a malonyl group onto an anthocyanin glucoside, shows that it is most similar to vinorine synthase and belongs in the BAHD acyltransferases superfamily [64]. This superfamily is involved in the biosynthesis of a dazzling array of plant natural products including small volatile esters, modified anthocyanins, as well as constitutive defense compounds and phytoalexins [11,62]. Notably, both vinorine synthase and Dm3MAT3 possess protein folds with 3D similarity to an acyltransferase involved in polyketide biosynthesis in *Mycobacterium tuberculosis* [9,51]. The structural similarity of these divergent polyketide-derived substrate acyltransferases perhaps hints at some ancestral link to fatty acid metabolism for Dm3MAT3 activity, similar to the aforementioned evolution of CHS/STS activity.

The evolutionary model for the emergence of flavonoid biosynthesis involves a process of gene duplication, in some cases originating in enzymes of primary metabolism, followed by genetic variation leading to functional divergence and selection. Following duplication, redundant metabolic enzymes have the freedom to evolve new biochemical activities with little attendant negative selective pressure, giving rise to new natural products and biosynthetic pathways, some of which may confer a selective advantage on the host plant species. Structural analysis of phenylpropanoid and flavonoid biosynthetic enzymes has afforded us an opportunity to not only understand their reaction mechanisms and the structural basis for substrate specificity/promiscuity but also to peer into the evolutionary landscape that these enzymes have traversed in arriving at their modern day functions. Indeed, this process allows us to go ‘back to the future’, engineering protein function based on principles that nature has used for millennia to evolve the dazzling array of natural products that form the basis of ecological interactions, nutrition and health on the terrestrial earth.

5. Concluding remarks

The explosion of structural studies of phenylpropanoid metabolizing enzymes over the past decade has contributed another critical layer of understanding to the considerable chemical, biochemical and genetic body of work that preceded these studies. The atomic resolution insights afforded by each of these 3D structures illuminated or clarified important mechanistic and substrate specificity determinants, thus also informing future bioinformatics, engineering and functional characterizations. While scientifically interesting, details gleaned from the phenylpropanoid pathway crystal structures accumulated over the past decade are also relevant to agricultural and biomedical studies. PAL, the first enzyme in the general phenylpropanoid pathway, is being developed as a detoxification treatment (medicine) for phenylketonuria, a congenital disease of phenylalanine metabolism. The OMTs and other monolignol pathway enzymes are of industrial and agricultural interest for engineering altered lignin content to facilitate wood pulping (paper), fermentation (biofuels) and digestion (herbivory) of commercially important plant species. In the flavonoid pathway, our structurally discovered ability to convert any endogenous and ubiquitous CHS into an STS with beneficial antifungal (agriculture) and human health (nutraceutical) properties is also of interest, as are numerous other applications of downstream flavonoid phytoalexins, phytoestrogens, and anthocyanin pigments.

While a significant amount of structural progress has been made on the phenylpropanoid biosynthetic pathway, many important structures remain unsolved, perhaps most notably the collection of P450 hydroxylases scattered throughout each of the major branches of the overall metabolic grid. As with these plant P450s, conserved primary sequence motifs in other structurally uncharacterized enzymes often allow identification of class or fold, and sometimes even homology models templated upon distant relatives of known structure. However, as was recently illustrated in our structural analysis of the evolutionary conversion of CHSs to resveratrol-producing STSs [2], homology models often fail to accurately predict relevant conformational differences between even closely related proteins. Thus even when sequence conservation allows homology modeling, experimental structural elucidation of individual pathway enzymes remains critical to achieve a more predictive understanding of the molecular basis of enzyme-catalyzed reactions, particularly in specialized (secondary) metabolism where even small variations in sequence can lead to profound alterations in specificities and mechanism.

The atomic resolution precision of currently known structures also contributes to our understanding of the evolutionary emergence of these critical plant pathways. Remaining structural gaps within the phenylpropanoid pathways aside, this greater evolutionary perspective is also dependent upon the availability of related structures outside of plants and specialized metabolism. For example, the close evolutionary relationship between the (first phenylpropanoid) structure of alfalfa CHS [15] and its most similar fatty acid relative was not apparent until subsequent publication of the *Escherichia coli* FabH crystal structure [12].

Nonetheless, it is already quite clear that evolution has repeatedly thrown a relatively small toolbox of biosynthetic tailoring enzyme families at existing chemical repertoires, in the process stumbling upon and retaining advantageous modifications. Useful enzymes are recruited randomly and piecemeal from other existing primary or specialized metabolic pathways. Gene duplication and divergence, as well as variation of the timing or location of gene expression, are all relevant variables in this combinatorial approach to specialized chemical divergence. In the case of the phenylpropanoids, the sheer number and diversity of advantageous physiological roles obscures their precise evolutionary history in land plants. In modern plants, retention of specific metabolic pathways often reflects current ecological niches, instead of the evolutionary history revealed by phylogeny.

Likewise, each modern species continues to tinker around the edges of its current chemical repertoire. As science has closely examined only a small percentage of existing species, and rarely all the related genes of any given enzyme family within a species, it would be sheer folly to claim that we have discovered the full extent of phenylpropanoid and related pathways. This point is underscored by the recent elucidation of the eugenol/isoeugenol biosynthetic pathway, as well as the discovery of a previously obscure third dimension of the monolignol biosynthetic grid. Combinatorial tailoring reactions aside, our understanding of the scope of evolutionary functional divergence of even distinctive chemical scaffold-forming enzymes like CHS continues to undergo additions and revisions, as novel type III PKS superfamily members with unanticipated structures and functions continue to be discovered, not only in plants but also fungi, slime molds and bacteria. As structure informs other scientific research, so will these related endeavors continue to identify interesting new enzymes for structural characterization.

Acknowledgments

We wish to thank Bernard Gallois (CBMN, UMR CNRS 5248, Talence, France) for providing the unpublished data regarding the structure of ANR.

Abbreviations

4CL	*p*-coumaroyl:CoA ligase
AAE	AMP-producing adenylating enzymes
ANR	anthocyanidin reductase
ANS	anthocyanidin synthase
C4H	cinnamic acid 4-hydroxylase
CAD	cinnamyl alcohol dehydrogenases
CCoAOMT	caffeoyl-CoA OMT
CCR	cinnamoyl-CoA:NADPH oxidoreductase
CHI	chalcone isomerase
ChOMT	chalcone *O*-methyltransferase
CHR	chalcone reductase
CHS	chalcone synthase
Dm3MAT3	malonyl anthocyanin acyltransferase
CoA	coenzyme A
COMT	caffeic acid 3-OMT
EGS	eugenol synthases
FAS	fatty acid synthase
FNS	flavone synthases
HAL	histidine ammonia-lyase
HCT	hydroxycinnamoyl transferase
IFR	isoflavone reductase
IGS	isoeugenol synthases

IOMT	(iso)flavone *O*-methyltransferase
KASIII	β-ketoacyl ACP synthase
OMT	*O*-methyl transferase
P450s	cytochrome P450 hydroxylases
PAL	phenylalanine ammonia-lyase
PCBER	phenylcoumaran benzylic ether reductase
PKS	polyketide synthase
PLR	pinoresinol-lariciresinol reductase
SAD	sinapyl alcohol dehydrogenases
SAM	*S*-adenosyl-L-methionine
SDR	short-chain dehydrogenase/reductase
STS	stilbene synthase
TAL	tyrosine ammonia lyase
UDG	uridine diphosphate glycosyltransferase
UGT71G1	UDP-flavonoid/triterpene glycosyltransferase
UGT85H2	(iso)flavonoid glycosyltransferase
VR	vestitone reductase
VvGT1	UDP-glucose:flavonoid 3-*O*-glycosyltransferase

References

1. Abdulrazzak N, Pollet B, Ehlting J, Larsen K, Asnaghi C, Ronseau S, Proux C, Erhardt M, Seltzer V, Renou JP, Ullmann P, Pauly M, Lapierre C, Werck-Reichhart D. A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. Plant Physiol 2006;140(1):30–48. [PubMed: 16377748]
2. Austin MB, Bowman ME, Ferrer JL, Schröder J, Noel JP. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem Biol 2004;11(9): 1179–1194. [PubMed: 15380179]
3. Austin MB, Noel JP. The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep 2003;20:79–110. [PubMed: 12636085]
4. Baldridge GD, O'Neill NR, Samac DA. Alfalfa (*Medicago sativa* l.) resistance to the root-lesion nematode, *Pratylenchus penetrans*: defense-response gene mRNA and isoflavonoid phytoalexin levels in roots. Plant Mol Biol 1998;38(6):999–1010. [PubMed: 9869406]
5. Besseau S, Hoffman L, Geoffroy P, Lapierre C, Pollet B, Legrand M. Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. Plant Cell 2007;19:148–162. [PubMed: 17237352]
6. Bomati E, Austin MB, Bowman ME, Dixon RA, Noel JP. Structural elucidation of chalcone reductase and implications for deoxychalcone biosynthesis. J Biol Chem 2005;280(34):30496–30503. [PubMed: 15970585]
7. Bomati EK, Noel JP. Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. Plant Cell 2005;17(5):1598–1611. [PubMed: 15829607]
8. Boudet A-M. Evolution and current status of research in phenolic compounds. Phytochemistry. 2007 in press.

9. Buglino J, Onwueme K, Ferreras J, Quadri LEN, Lima CD. Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from *Mycobacterium tuberculosis*. J Biol Chem 2004;29:30634. [PubMed: 15123643]
10. Calabrese JC, Jordan DB, Boodhoo A, Sariaslani S, Vannelli T. Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis. Biochemistry 2004;43:11403–11416. [PubMed: 15350127]
11. D'Auria JC. Acyltransferases in plants: A good time to be BAHD. Curr Opin Plant Biol 2006;9(3): 331–340. [PubMed: 16616872]
12. Davies C, Heath RJ, White SW, Rock CO. The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. Structure 2000;8:185–195. [PubMed: 10673437]
13. Dinkova-Kostova AT, Gang DR, Davin LB, Bedgar DL, Chu A, Lewis NG. (+)-Pinoresinol/(+)-lariciresinol reductase from *Forsythia intermedia*. Protein purification, cDNA cloning, heterologous expression and comparison to isoflavone reductase. J Biol Chem 1996;271(46): 29473–29482. [PubMed: 8910615]
14. Dixon RA, Lamb CJ, Masoud S, Sewalt VJH, Paiva NL. Metabolic engineering: Prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses—a review. Gene 1996;179(1):61–71. [PubMed: 8955630]
15. Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nature 1999;6:775–784.
16. Ferrer JL, Zubieta C, Dixon RA, Noel JP. Crystal structures of alfalfa caffeoyl coenzyme a 3-O-methyltransferase. Plant Physiol 2005;137(3):1009–1017. [PubMed: 15734921]
17. Fujita M, Gang DR, Davin LB, Lewis NG. Recombinant pinoresinol-lariciresinol reductases from western red cedar (*Thuja plicata*) catalyze opposite enantiospecific conversions. J Biol Chem 1999;274(2):618–627. [PubMed: 9872995]
18. Gang DR, Costa MA, Fujita M, Dinkova-Kostova AT, Wang HB, Burlat V, Martin W, Sarkanen S, Davin LB, Lewis NG. Regiochem-ical control of monolignol radical coupling: A new paradigm for lignin and lignan biosynthesis. Chem Biol 1999;6:143–151. [PubMed: 10074466]
19. Gang DR, Kasahara H, Xia ZQ, Vander Mijnsbrugge K, Bauw G, Boerjan W, Van Montagu M, Davin LB, Lewis NG. Evolution of plant defense mechanisms. Relationships of phenylcoumaran benzylic ether reductases to pinoresinol-lariciresinol and isoflavone reductases. J Biol Chem 1999;274(11):7516–7527. [PubMed: 10066819]
20. Gensheimer M, Mushegian A. Chalcone isomerase family and fold: No longer unique to plants. Protein Sci 2004;13(2):540–544. [PubMed: 14718655]
21. Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC. The 1.75 Å crystal structure of acetyl-coA synthetase bound to adenosine-5′-propylphosphate and coenzyme A. Biochemistry 2003;42(10):2866–2873. [PubMed: 12627952]
22. Guo D, Chen F, Inoue K, Blount JW, Dixon RA. Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell 2001;13(1):73–88. [PubMed: 11158530]
23. Hatfield R, Vermerris W. Lignin formation in plants. The dilemma of linkage specificity. Plant Physiol 2001;126(4):1351–1357. [PubMed: 11500535]
24. Hisanaga Y, Ago H, Nakagawa N, Hamada K, Ida K, Yamamoto M, Hori T, Arii Y, Sugahara M, Kuramitsu S, Yokoyama S, Miyano M. Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-coA synthetase dimer. J Biol Chem 2004;279(30):31717–31726. [PubMed: 15145952]
25. Hwang EI, Kaneko M, Ohnishi Y, Horinouchi S. Production of plant-specific flavanones by *Escherichia coli* containing an artificial gene cluster. Appl Environ Microbiol 2003;69(5):2699–2706. [PubMed: 12732539]
26. Inoue K, Sewalt VJH, Murray Ballance G, Ni W, Sturzer C, Dixon RA. Developmental expression and substrate specificities of alfalfa caffeic acid 3-o-methyltransferase and caffeoyl coenzyme a 3-o-methyltransferase in relation to lignification. Plant Physiol 1998;117(3):761–770. [PubMed: 9662519]

27. Jaeck E, Dumas B, Geoffroy P, Favet N, Inzé D, Montagu MV, Fritig B, Legrand M. Regulation of enzymes involved in lignin biosynthesis: Induction of *O*-methyltransferase mRNAs during the hypersensitive reaction of tobacco to tobacco mosaic virus. Mol Plant Microbe Interact 1992;5:294–300. [PubMed: 1515665]
28. Jez JM, Bennet MJ, Schlegel BP, Lewis M, Penning TM. Comparative anatomy of the aldo-keto reductase superfamily. Biochem J 1997;326:625–636. [PubMed: 9307009]
29. Jez JM, Bowman ME, Dixon RA, Noel JP. Structure and mechanism of the evolutionary unique plant enzyme chalcone isomerase. Nature 2000;7:786–791.
30. Jez JM, Bowman ME, Noel JP. Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. Biochemistry 2002;41(16):5168–5176. [PubMed: 11955065]
31. Jez JM, Ferrer JL, Bowman ME, Austin MB, Schröder J, Dixon RA, Noel JP. Structure and mechanism of chalcone synthase-like polyketide synthases. J Ind Microbiol Biotechnol 2001;27(6):393–398. [PubMed: 11774005]
32. Jez JM, Noel JP. Mechanism of chalcone synthase. Pka of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. J Biol Chem 2000;275(50):39640–39646. [PubMed: 11006298]
33. Jez JM, Noel JP. Reaction mechanism of chalcone isomerase. PH dependence, diffusion control, and product binding differences. J Biol Chem 2002;277(2):1361–1369. [PubMed: 11698411]
34. Joernvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D. Short-chain dehydrogenases/reductases (SDR). Biochemistry 1995;34(18):6003–6013. [PubMed: 7742302]
35. Jouanin L, Goujon T, de Nadai V, Martin MT, Mila I, Vallet C, Pollet B, Yoshinaga A, Chabbert B, Petit-Conil M, Lapierre C. Lignification in transgenic poplars with extremely reduced caffeic acid *O*-methyltransferase activity. Plant Physiol 2000;123(4):1363–1374. [PubMed: 10938354]
36. Koeduka T, Fridman E, Gang DR, Vassao DG, Jackson BL, Kish CM, Orlova I, Spassova SM, Lewis NG, Noel JP, Baiga TJ, Dudareva N, Pichersky E. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. Proc Natl Acad Sci USA 2006;103(26):10128–10133. [PubMed: 16782809]
37. Koutaniemi S, Toikka MM, Karkonen A, Mustonen M, Lundell T, Simola LK, Kilpelainen IA, Teeri TH. Characterization of basic *p*-cou-maryl and coniferyl alcohol oxidizing peroxidases from a lignin-forming *Picea abies* suspension culture. Plant Mol Biol 2005;58(2):141–157. [PubMed: 16027971]
38. Lapierre C, Pollet B, Petit-Conil M, Toval G, Romero J, Pilate G, Leple JC, Boerjan W, Ferret V, De Nadai V, Jouanin L. Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid *O*-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. Plant Physiol 1999;119(1):153–164. [PubMed: 9880356]
39. Lawton MA, Lamb CJ. Transcriptional activation of plant defense genes by fungal elicitor, wounding, and infection. Mol Cell Biol 1987;7(1):335–341. [PubMed: 3561393]
40. Lewis, NG.; Davin, LB. Lignans: biosynthesis and function. In: Barton, DHR., editor. Comprehensive Natural Products Chemistry, Polyketides and Other Secondary Metabolites Including Fatty Acids and their Derivatives. Vol. 1. Elsevier; 1999. p. 639-712.
41. Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X. Crystal structure of *Medicago truncatula* UGT85H2-insight into structural basis of a multifunctional (iso)flavonoid glycosyltransfer-ase. J Mol Biol 2007;370:951–963. [PubMed: 17553523]
42. Li L, Popko JL, Umezawa T, Chiang VL. 5-hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms. J Biol Chem 2000;275(9):6537–6545. [PubMed: 10692459]
43. Liu CJ, Deavours BE, Richard SB, Ferrer JL, Blount JW, Huhman D, Dixon RA, Noel JP. Structural basis for dual functionality of isoflavonoid *O*-methyltransferases in the evolution of plant defense responses. Plant Cell 2006;18:3656–3669. [PubMed: 17172354]
44. Louie GV, Baiga TJ, Bowman ME, Koeduka T, Taylor JH, Spassova SM, Pichersky E, Noel JP. Structure and reaction mechanism of basil eugenol synthase. PLoS ONE 2007;2(10):e993. [PubMed: 17912370]

45. Louie GV, Bowman ME, Moffitt MC, Baiga TJ, Moore BS, Noel JP. Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases. Chem Biol 2006;13:1327–1338. [PubMed: 17185228]
46. Min T, Kasahara H, Bedgard DL, Youn B, Lawrence PK, Gang DR, Halls SC, Park H, Hilsenbeck JL, Davin LB, Lewis NG, Kang C. Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. J Biol Chem 2003;278:50714–50723. [PubMed: 13129921]
47. Moffitt MC, Louie GV, Bowman ME, Pence J, Noel JP, Moore BS. Discovery of two cyanobacterial phenylalanine ammonia lyases: Kinetic and structural characterization. Biochemistry 2007;46(4):1004–1012. [PubMed: 17240984]
48. Munro AW, Girvan HM, McLean KJ. Variations on a (t)heme—novel mechanisms, redox partners and catalytic functions in the cytochrome p450 superfamily. Nat Prod Rep 2007;24:585–609. [PubMed: 17534532]
49. Ni W, Fahrendorf T, Ballance GM, Lamb CJ, Dixon RA. Stress responses in alfalfa (*Medicago sativa* L.). XX. Transcriptional activation of phenylpropanoid pathway genes in elicitor-induced cell suspension cultures. Plant Mol Biol 1996;30(3):427–438. [PubMed: 8605296]
50. Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J 2006;25(6):1396–1405. [PubMed: 16482224]
51. Onwueme K, Ferreras J, Buglino J, Lima CD, Quadri LEN. Mycobacterial polyketide-associated proteins are acyltransferases: proof of principle with *Mycobacterium tuberculosis* PapA5. Proc Natl Acad Sci USA 2004;101:4608–4613. [PubMed: 15070765]
52. Petit P, Granier T, Langlois d'Estaintot B, Manigand C, Bathany K, Schmitter J-M, Lauvergeat V, Hamdi S, Gallois B. Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. J Mol Biol 2007;368:1345–1357. [PubMed: 17395203]
53. Pincon G, Maury S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M. Repression of *O*-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth. Phytochemistry 2001;57(7):1167–1176. [PubMed: 11430989]
54. Pond KR, Ellis WC, Lascano CE, Akin DE. Fragmentation and flow of grazed coastal bermudagrass through the digestive tract of cattle. J Anim Sci 1987;65(2):609–618. [PubMed: 3624104]
55. Ritter H, Schulz GE. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. Plant Cell 2004;16(12):3426–3436. [PubMed: 15548745]
56. Ros Barcelo A. Lignification in plant cell walls. Int Rev Cytol 1997;176:87–132. [PubMed: 9394918]
57. Schwede TF, Retey J, Schulz GE. Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. Biochemistry 1999;38(17):5355–5361. [PubMed: 10220322]
58. Shao H, Dixon RA, Wang X. Crystal structure of vestitone reductase from alfalfa (*Medicago sativa* L.). J Mol Biol 2007;369:265–276. [PubMed: 17433362]
59. Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. Plant Cell 2005;17(11):3141–3154. [PubMed: 16214900]
60. Shockey JM, Fulda MS, Browse J. *Arabidopsis* contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme a synthetases. Plant Physiol 2003;132(2):1065–1076. [PubMed: 12805634]
61. Shomura Y, Torayama I, Suh DY, Xiang T, Kita A, Sankawa U, Miki K. Crystal structure of stilbene synthase from *Arachis hypogaea*. Protein Struct Funct Bioinformatics 2005;60:803–806.
62. St Pierre B, De Luca V. Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. Recent Adv Phytochemistry 2000;34:285–315.
63. Stafford H. Flavonoid evolution: An enzymic approach. Plant Physiol 1991;96:680–685. [PubMed: 16668242]

64. Unno H, Ichimaida F, Suzuki H, Takahashi S, Tanaka Y, Saito A, Nishino T, Kusunoki M, Nakayama T. Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. J Biol Chem 2007;282(21):15812–15822. [PubMed: 17383962]

65. Wang X, He XZ, Lin J, Shao H, Chang Z, Dixon RA. Crystal structure of isoflavone reductase from alfalfa (*Medicago sativa* l.). J Mol Biol 2006;358:1341–1352. [PubMed: 16600295]

66. Waters E. Molecular adaptation and the origin of land plants. Mol Phylogenet Evol 2003;29:456–463. [PubMed: 14615186]

67. Watts KT, Mijts BN, Lee PC, Manning AJ, Schmidt-Dannert C. Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. Chem Biol 2006;13:1317–1326. [PubMed: 17185227]

68. Welford RWD, Clifton IJ, Turnbull JJ, Wilson SC, Schofield CJ. Structural and mechanistic studies on anthocyanidin synthase catalysed oxidation of flavanone substrates: The effect of c-2 stereochemistry on product selectivity and mechanism. Org Biomol Chem 2005;3:3117–3126. [PubMed: 16106293]

69. Welle R, Schröder G, Schiltz E, Grisebach H, Schröder J. Induced plant responses to pathogen attack. Analysis and heterologous expression of the key enzyme in the biosynthesis of phytoalexins in soybean (*Glycine max* l. Merr. Cv. Harosoy 63). Eur J Biochem 1991;196(2):423–430. [PubMed: 1840523]

70. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE. Mammalian microsomal cytochrome p450 monooxygenase: Structural adaptations for membrane binding and functional diversity. Mol Cell 2000;5(1):121–131. [PubMed: 10678174]

71. Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, Schofield CJ. Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. Structure 2002;10:93–103. [PubMed: 11796114]

72. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 2001;126:485–493. [PubMed: 11402179]

73. Winkel-Shirley B. It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. Plant Physiol 2001;127:1399–1404. [PubMed: 11743081]

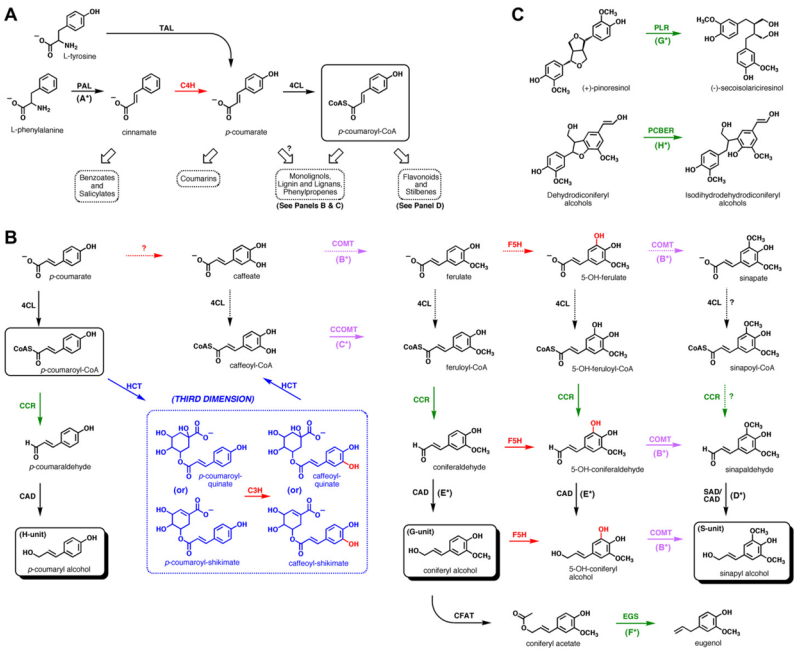
74. Xia ZQ, Costa MA, Pelissier HC, Davin LB, Lewis NG. Secoisolariciresinol dehydrogenase purification, cloning, and functional expression. Implications for human health protection. J Biol Chem 2001;276(16):12614–12623. [PubMed: 11278426]

75. Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. Role of anthocyanidin reductase, encoded by banyuls in plant flavonoid biosynthesis. Science 2003;299(5605):396–399. [PubMed: 12532018]

76. Youn B, Camacho R, Moinuddin SGA, Lee C, Davin LB, Lewis NG, Kang C. Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. Org Biomol Chem 2006;4:1687–1697. [PubMed: 16633561]

77. Zubieta C, He XZ, Dixon RA, Noel JP. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant *O*-methyltransferases. Nat Struct Biol 2001;8(3):271–279. [PubMed: 11224575]

78. Zubieta C, Kota P, Ferrer JL, Dixon RA, Noel JP. Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-*O*-methyltransferase. Plant Cell 2002;14(6):1265–1277. [PubMed: 12084826]



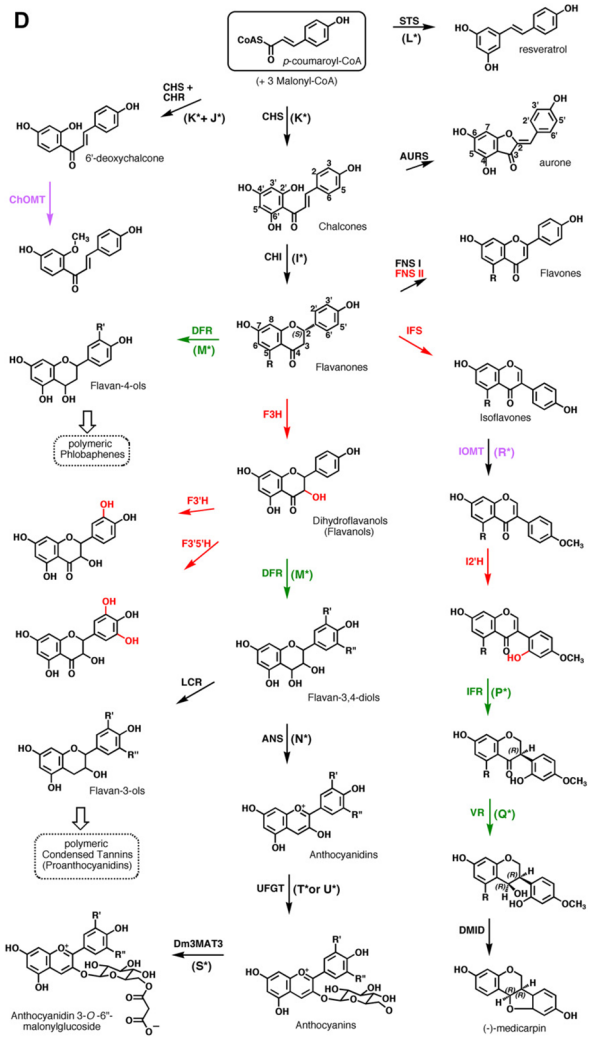


Fig. 1.

Phenylpropanoid biosynthetic pathways. Enzymes and representative reactions of each major pathway branch are shown. In all panels, capital letters marked with asterisks designate corresponding panels of Fig. 2 for structurally characterized pathway enzymes, and color indicates highly represented enzyme classes: cytochrome P450-catalyzed reactions are highlighted in red, short-chain dehydrogenase/reductase (SDR) reactions are green, and *O*-methyltransferase (OMT) reactions are lavender. (A) The general phenylpropanoid pathway leading from phenylalanine to *p*-coumaroyl-CoA, the entry point to each major downstream pathways. Additional offshoot product classes are indicated. (B) Monolignol biosynthetic grid leading to polymeric lignin, oligomeric lignans, and monomeric phenylpropenes. Newly discovered ‘third dimension’ shown in blue. Monolignol end products are in shadowed boxes. (C) Structurally characterized lignan reductase reactions. (D) Flavonoid pathway leading to anthocyanin pigments, isoflavonoid ‘phytoestrogens’, polymeric phlobaphenes and condensed tannins, and various antimicrobial phytoalexins. Related biosynthesis of resveratrol is also shown.

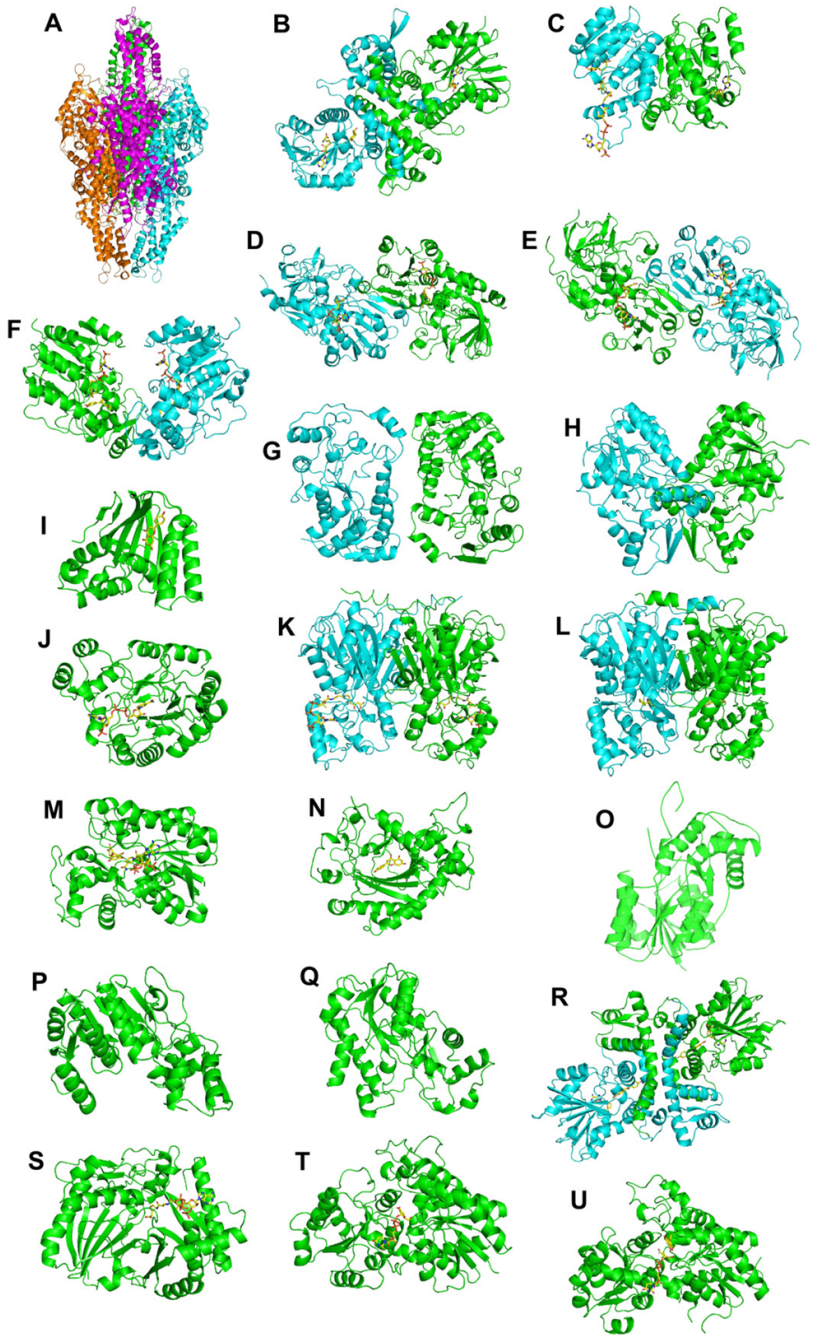


Fig. 2.

Elucidated structures of the phenylpropanoid pathways. Crystal structures of: (A) PAL. (B) COMT in complex with SAH and ferulic acid. (C) CCoAOMT in complex with sinapoyl-CoA. (D) SAD in complex with NADP+. (E) CAD in complex with NADP+. (F) EGS in complex with NADP+ and (7*S*,8*S*)-ethyl (7,8-methylene)-dihydroferulate. (G) PLR. (H) PCBER. (I) CHI in complex with naringenin. (J) CHR in complex with NADP+. (K) CHS in complex with malonyl-CoA. (L) STS in complex with resveratrol. (M) DFR in complex with NADP and dihydroquercetin. (N) ANS in complex with naringenin. (O) ANR. (P) IFR. (Q) VR. (R) IOMT in complex with SAH and 4′-hydroxy-7-methoxyisoflavone. (S) Dm3MAT3 in complex with malonyl-CoA. (T) UGT71G1 in complex with UDP-glucose.

(U) VvGT1 in complex with UDP-2FGlc and kaempferol. This figure was produced with PyMOL (http://www.pymol.org).

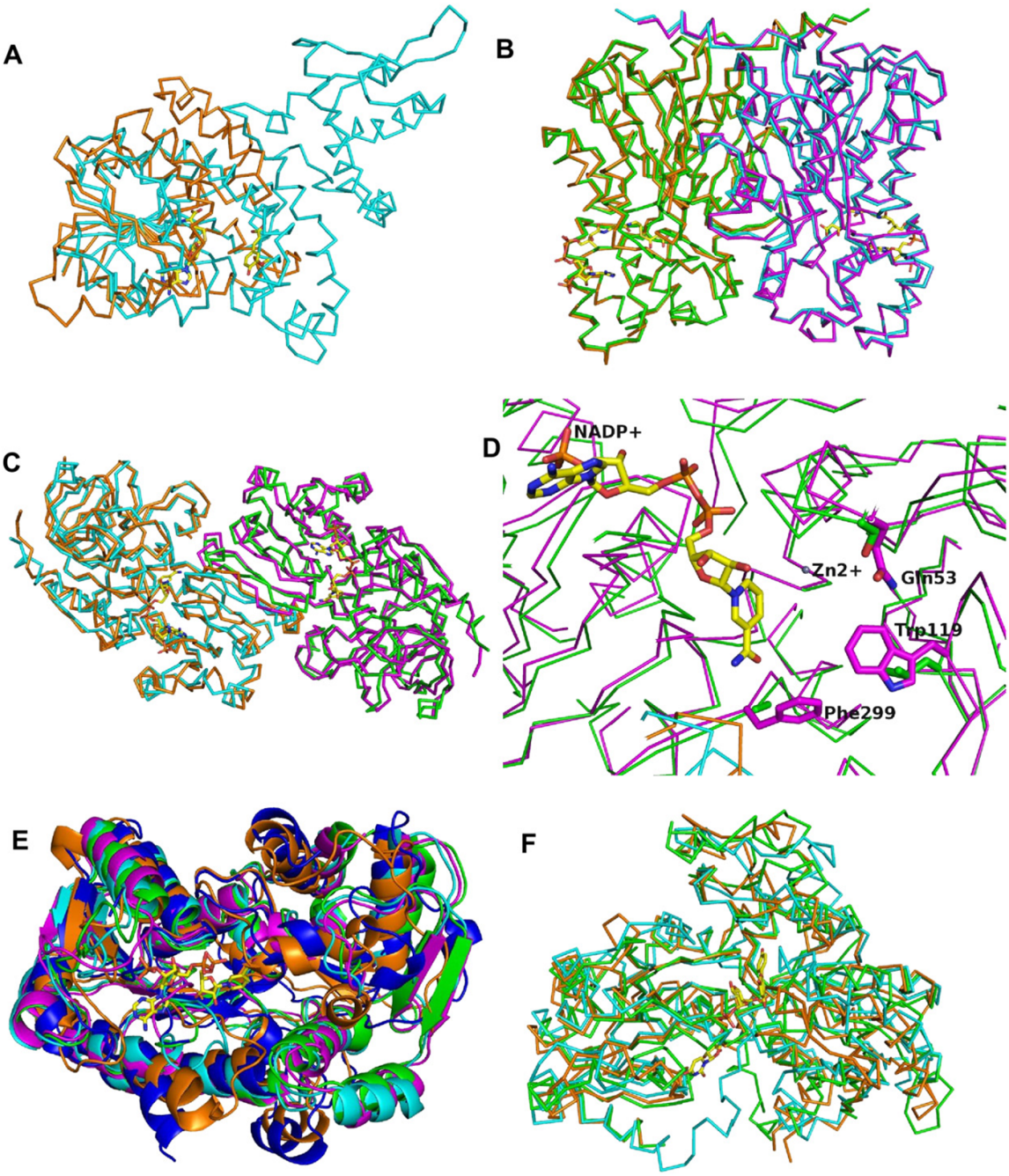


Fig. 3.

Structural comparisons within the same families. (A) Comparison of the overall structure of COMT and CCoAOMT monomers (colored in cyan and orange respectively). SAH and ferulic acid are represented as yellow sticks. (B) Comparison of the overall structure of stilbene synthase (STS: monomers in orange and magenta) and chalcone synthase in complex with malonyl CoA (CHS: monomers in cyan and green; malonyl-CoA as yellow sticks). (C) Comparison of the overall structure of cinnamyl/sinapyl alcohol dehydrogenase (CAD: monomers in orange and magenta) and sinapyl alcohol dehydrogenases (SAD: monomers in cyan and green). (D) Comparison of the active sites of CAD and SAD (same colors as panel C). Residues are labeled according to CAD sequence. NADP+ is represented

as yellow sticks. (E) Comparison of the different reductases of the phenylpropanoid pathway structurally characterized (PCBER is represented in green, PLR in cyan, DFR in orange, IFR in magenta and VR in blue). (F) Comparison of glycosyltransferases of the phenylpropanoid (UGT85H2 in green, UGT71G1 in cyan and VvGT1 in orange). UDP-2FGlc and kaempferol from VvGT1 structure are represented as yellow sticks. This figure was produced with PyMOL (http://www.pymol.org).

Table 1

Key structures of the plant phenylpropanoid pathway

Acronym	Name	Substrate	PDB code	Reference
PAL	Phenylalanine ammonia-lyase	Apoenzyme	**1W27**	[55]
COMT	Caffeic acid *O*-methyltransferase	5-Hydroxyconiferaldehyde complex	1KYW	[78]
		Ferulic acid complex	**1KYZ**	[78]
CCoAMT	Caffeoyl-CoA *O*-methyltransferase	Feruloyl-CoA complex	1SUI	[16]
		Sinapoyl-CoA complex	**1SUS**	[16]
SAD	Sinapyl alcohol dehydrogenase	Apoenzyme	1YQX	[7]
		NADP+ complex	**1YQD**	[7]
CAD	Cinnamyl alcohol dehydrogenase	Apoenzyme	2CF5	[76]
		NADP+ complex	**2CF6**	[76]
EGS	Eugenol synthase	Apoenzyme	2QYS	[44]
		NADP+ complex	2QW8	[44]
		NADPH complex	2R6J	[44]
		NADP+ and EMDF complex	**2QZZ**	[44]
		K132Q mutant, EMDF complex	2R2G	[44]
PLR	Pinoresinol-lariciresinol reductase	Apoenzyme	**1QYD**	[46]
PCBER	Phenylcoumaran benzylic ether reductase	Apoenzyme	**1QYC**	[46]
CHS	Chalcone synthase	Apoenzyme	1BI5	[15]
		CoA complex	1BQ6	[15]
		Naringenin complex	1CGK	[15]
		Resveratrol complex	1CGZ	[15]
		Hexanoyl-CoA complex	1CHW	[15]
		Malonyl-CoA complex	**1CML**	[15]
		C164A mutant	1D6F	[32]
		N336A mutant	1D6H	[32]
		H303Q mutant	1D6I	[32]
		‘18x’ mutant	1U0V	[2]
		‘18x’ mutant, resveratrol complex	1U0W	[2]
STS	Stilbene synthase	Apoenzyme	1U0U	[2]
		Apoenzyme	1XES	PDB entry only
		Methylmalonyl-CoA	1XET	PDB entry only
		Apoenzyme	1Z1E	[61]
		Resveratrol complex	**1Z1F**	[61]
CHI	Chalcone isomerase	Apoenzyme	1EYP	[31]
		Naringenin complex	**1EYQ**	[31]
		5-Deoxyflavanone complex	1FM7	[33]
		5,4′-Dideoxyflavanone complex	1FM8	[33]
		4′-Dydroxyflavanone complex	1JEP	[33]
		Y106F mutant	1JX0	[30]
		T48A mutant	1JX1	[30]

Acronym	Name	Substrate	PDB code	Reference
CHR	Chalcone reductase	NADP+ complex	**1ZGD**	[7]
DFR	Dihydroflavonol 4-reductase	NADP/dihydroquercetin complex	**2C29**	[52]
IFR	Isoflavone reductase	Apoenzyme	**2GAS**	[65]
VR	Vestitone reductase	Apoenzyme	**2P4H**	[58]
HI4′OMT	Isoflavonoid *O*-methyltransferase	SAH complex	1ZHF	[43]
		SAH and (2r,3r)-2,7-dihydroxi-3-(4-hydroxyphenyl)-2,3-dihydro-4h-chromen-4-one complex	1ZG3	[43]
		SAH and (6ar,12ar)-6h[1,3]dioxolo[5,6][1]benzofuro [3,2-C]chromen-3,6a(12ah)-diol complex	1ZGA	[43]
		SAH and (6ar,12ar)-3-(hydroxymethyl)-6h[1,3]dioxolo [5,6][1] benzofuro[3,2-C]chromen-6a(12ah)-ol complex	1ZGJ	[43]
ChOMT	Chalcone *O*-methyltransferase	SAM complex	1FPQ	[77]
		SAM and 2′,4,4′-trihydroxychalcone complex	1FP1	[77]
IOMT	Isoflavone *O*-methyltransferase	SAM complex	1FPX	[77]
		SAH and 4′-hydroxy-7-methoxyisoflavone complex	**1FP2**	[77]
UGT85H2	(Iso)flavonoid glycosyltransferase	Apoenzyme	2PQ6	[41]
UGT71G1	Triterpene/flavonoid glycosyltransferase	UDP complex	2ACV	[59]
		UDP-glucose complex	**2ACW**	[59]
VvGT1	Flavonoid 3-*O*-glycosyltransferase	UDP complex	2C1X	[50]
		UDP-2FGlc and kaempferol complex	**2C1Z**	[50]
		UDP and quercetin complex	2C9Z	[50]
ANS	Anthocyanidin synthase	Naringenin complex	**2BRT**	[68]
		Trans-dihydroquercetin complex	1GP5	[71]
		Trans-dihydroquercetin complex	1GP6	[71]
Dm3MAT3	Anthocyanin malonyltransferase	Apoenzyme	2E1U	[64]
		Malonyl-CoA complex	**2E1T**	[64]

List of relevant structures deposited in the Protein Data Bank. In bold: structures shown in Fig. 2.