Published as: *Curr Opin Plant Biol.* 2005 June ; 8(3): 249–253.

Structure–function relationships in plant phenylpropanoid biosynthesis

Joseph P Noel^{1,2}, Michael B Austin^{1,2}, and Erin K Bomati^{1,2}

¹Jack Skirball Chemical Biology and Proteomics Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

²Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92037, USA

Abstract

Plants, as sessile organisms, evolve and exploit metabolic systems to create a rich repertoire of complex natural products that hold adaptive significance for their survival in challenging ecological niches on earth. As an experimental tool set, structural biology provides a high-resolution means to uncover detailed information about the structure–function relationships of metabolic enzymes at the atomic level. Together with genomic and biochemical approaches and an appreciation of molecular evolution, structural enzymology holds great promise for addressing a number of questions relating to secondary or, more appropriately, specialized metabolism. Why is secondary metabolism so adaptable? How are reactivity, regio-chemistry and stereo-chemistry steered during the multi-step conversion of substrates into products? What are the vestigial structural and mechanistic traits that remain in biosynthetic enzymes during the diversification of substrate and product selectivity? What does the catalytic landscape look like as an enzyme family traverses all possible lineages *en route* to the acquisition of new substrate and/or product specificities? And how can one rationally engineer biosynthesis using the unique perspectives of evolution and structural biology to create novel chemicals for human use?

Introduction

Sessile organisms such as plants have evolved specialized biosynthetic pathways, the output of which are regio- and stereo-chemically complex small-molecule natural products. These chemicals of secondary metabolism often impart a species-specific chemical ‘signature’ upon an organism and confer a multitude of evolutionary advantages. Moreover, biologically active natural products have historically been exploited during the search for new pharmaceutical agents.

Diverse molecular changes that are associated with specialized metabolism are often preserved genetically, functionally and structurally as a result of the increased adaptability of their biosynthetic hosts in diverse and challenging ecosystems. Therefore, specialized metabolic pathways and their chemical products present us with a rich evolutionary record of where metabolic pathways, natural products and biosynthetic enzymes have been, what adaptive significance these complex enzymatic systems hold at present, and ultimately where these pathways might be heading in the future.

The rich chemical diversity of plants is the result of ongoing evolutionary processes. Recent advances in the molecular biology of plants, particularly in the area of large-scale genomics [1••], are revealing how enzymes of natural product biosynthesis arise through mutation and gene duplication, leading to the continued elaboration of new chemical structures that will be selected for if they impart an adaptive advantage on the plant [2]. Many of these compounds act as chemical cues for plants during their on-going interactions with physical and biotic factors in their environments. In addition, plant natural products have positive and negative impacts on human and animal health and nutrition. These diverse roles in plant and animal physiology provide scope for molecular approaches to crop improvement that are based upon the manipulation of natural product profiles.

Structural biology provides an important tool set for the detailed structure–function characterization of proteins at the atomic level [3,4••]. The level of functional understanding derived from experimentally determined structures or from realistic models that are constructed from homologous protein folds [5•] can lead to a more complete appreciation of complex biosynthetic pathways. Such information can elucidate the mechanisms of individual biosynthetic reactions [6], can afford new views at atomic resolution of the temporal and spatial architecture of multi-protein complexes that are vital to metabolic flux and channeling [7], and can provide a practical and rational basis for engineering useful metabolic traits [8,9] into medicinally and agriculturally important plants. This review focuses on the phenylpropanoid pathway of plants to provide a partial survey of the ways in which structural biology can impact our fundamental understanding of the evolution and biochemistry of secondary metabolism. We also discuss how atomic-resolution information can be used in a practical way to facilitate the rational engineering of enzymes.

Phenylpropanoid biosynthetic pathway

Plant phenylpropanoids encompass a group of phenylalanine-derived chemicals that comprise a structurally diverse group of secondary metabolites. They play vital roles in the interaction of plants with their surrounding environment. The structural diversity of phenylpropanoids is due to the action of enzymes and enzyme complexes that bring about regio-specific condensation, cyclization, aromatization, hydroxylation, glycosylation, acylation, prenylation, sulfation, and methylation reactions.

Many enzymes in these multiple-branched biosynthetic grids belong to an even larger superfamily of biosynthetic enzymes that utilize a core set of chemical transformations but that extend their synthetic capabilities beyond the general phenylpropanoid skeleton (C6–C3). In most species that maintain the phenylpropanoid pathway, three enzymes are required to transform phenylalanine into the Coenzyme A (CoA)-activated hydroxycinnamoyl (phenylpropanoid) esters that are capable of entering various downstream pathways. The deamination of phenylalanine by phenylalanine ammonia lyase (PAL) first produces cinnamic acid, which serves as the precursor for all of the phenylpropanoids of secondary metabolism. A cinnamic acid 4-hydroxylase (C4H) catalyzes the introduction of a hydroxyl group at the para position of the phenyl ring of cinnamic acid, producing coumaric acid. The carboxyl group of coumaric acid is then activated by the formation of a thioester bond with CoA, a process catalyzed by hydroxycinnamate CoA ligase (4CL). Notably, grasses and some fungi possess a dual-specificity ammonia lyase (PAL/tyrosine ammonia lyase [TAL]) that uses tyrosine as a substrate, reducing the number of enzymes that are essential for the production of p-coumaroyl-CoA from three in the general phenylpropanoid pathway to two [10].

Structural enzymology of PAL

PAL catalyzes the non-oxidative elimination of ammonia from L-Phe, yielding *trans*-cinnamate. Although PAL lacks a cofactor, the lyase reaction requires an electrophilic moiety, which is formed auto-catalytically from the cyclization and dehydration of an Ala-Ser-Gly segment at the active site. This autocatalytic process mirrors the reaction in the related histidine ammonia lyase (HAL) enzyme, which results in the covalent attachment of a 4-methylidene-imidazole-5-one (MIO) group to this enzyme [11]. Two recent crystal structures of PAL, one from the yeast *Rhodospiridium toruloides* [12••] and one from parsley [13•], provide new mechanistic insights into the catalytic mechanism and phylogenetic relationships of PALs and HALs. PAL is found in plants, and to a limited extent in fungi and bacteria, whereas HAL is Widespread.

Both bacterial HAL and fungal and plant PAL are tetramers that are largely made up of α helices. Unlike HAL, PAL possesses a mobile amino-terminal region that encompasses 24 residues and two helices inserted into the MIO-containing domain, together, these features form a 'fan' [12••] or 'shielding domain' [13•]. Ritter and Schulz [13•] propose that this shielding domain fortifies the interface with the core domain, which contains the active site that is shared with HAL, thus restricting substrate access.

Calabrese *et al.* [12••] found that the *R. toruloides* PAL has the same overall structure as that of parsley, except that the polypeptide loops residing over the active site are disordered. These authors point out that six of the seven helices of the *Rhodospiridium* PAL are aligned so that their respective helical dipoles create an electropositive platform for the MIO moiety, thus strengthening the MIO electrophilic 'co-factor'. This three-dimensional arrangement of secondary structural elements is used to argue that the generally accepted reaction scheme for the non-oxidative elimination of ammonia, through a carbocation species centered on the phenyl ring of a covalently bound phenylalanine substrate, is not feasible given the presence of six positive helical dipoles that point into the active site. Instead, Calabrese *et al.* [12••] propose an alternative scheme in which MIO forms a transient bond with the amino group of phenylalanine before ammonia is eliminated. If true, this later mechanism could open up new avenues for the engineering of alternative amino-acid specificities into PAL/HAL, and thus could facilitate the generation of wholly novel substrates for downstream enzymes in plant biosynthetic pathways.

Plant polyketide biosynthesis

A major class of molecules in the phenylpropanoid pathway is derived from a backbone wherein the phenylpropanoid unit that is derived from phenylalanine and activated through conjugation to CoA is extended by three acetyl moieties. These moieties condense with one and other and subsequently undergo cyclization and aromatization reactions to form a second polyhydroxylated ring. The group of natural products that are formed in this way includes a large number of flavonoids, isoflavonoids and stilbenes [14].

In all plants characterized to date, this branch of the phenylpropanoid pathway is initiated by the elongation of the p-coumaroyl-CoA unit to a C15 skeleton through the iterative action of an ubiquitous plant enzyme known as chalcone synthase (CHS). A large variety of stress-induced compounds and pigments are formed from this C15 skeleton after biosynthetic elaboration by a wide array of downstream enzymes. 4,2',4',6'-tetrahydroxychalcone (chalcone), the typical cyclized C15 molecule, is synthesized by CHS via the sequential condensation of one p-coumaroyl-CoA and three malonyl-CoA molecules. Stilbenes, such as resveratrol, are produced in a similar way from an identical set of starting materials and by the activity of an enzyme family that is homologous to CHS, known as stilbene synthase (STS). The biosynthesis of stilbenes, however, involves an additional decarboxylation step,

which occurs after the condensation mechanism that leads to a C14 backbone, in which the two aromatic rings are separated by two carbons instead of three. These enzymatic transformations are analogous to those performed by a large number of polyketide synthases (PKSs) found in Nature [15•].

Remarkably, until recently, the polyketide origins of chalcone and stilbene biosynthesis in plants were largely ignored by the larger community of natural product chemists and biochemists. This lack of attention was partly due to the early belief that polyketide biosynthesis in plants was restricted to the formation of chalcones and, in limited plant taxonomic groups, stilbenes. However, recent discoveries of plant-like PKSs in bacteria [16], their structural conservation with plant CHSs [17•,18•], and the growing number of biosynthetically unique plant PKSs (see below) garnered widespread appreciation of plant PKSs as important targets for mechanistic and structural analysis. This resulted in the generally accepted inclusion of CHS-like PKSs in the larger family of biosynthetically and genetically diverse PKSs and their designation as type III PKSs [15•].

CHS exists as a homodimer that contains two distinct bilobed active-site cavities, which are situated at the bottom edge of each monomer's conserved $\alpha\beta\alpha\beta\alpha$ core. Identical six-residue loops from each monomer, which meet at the dimer interface, separate the active sites of the two monomers from each other. A series of larger loops surround the bottom half of the active site, forming an additional domain. The active-site cavity is buried except for a 16 Å CoA-binding tunnel through which CoA-linked substrates and intermediates are delivered to the catalytic machinery [14].

STS has evolved in a limited number of phylogenetically distinct plants by gene duplication and subsequent mechanistic divergence from CHS [19]. Although it employs the same substrates as CHS, STS catalyzes a carbon 2 to carbon 7 aldol condensation that forms the stilbene backbone of resveratrol and related antifungal phytoalexins. The first STS crystal structure confirmed the overall structural homology of CHS and STS but revealed the previously obscure structural basis for the evolution of STSs from their CHS ancestors [20••]. Unexpectedly, the mechanism of STS functional divergence arises from the upregulation of a cryptic thioesterase activity in the active site, which is explained by an alternative hydrogen-bonding network termed the 'aldol switch'. This mechanism is distinct from those predicted by previous models of the functional divergence of type III PKSs in plants and bacteria, which relied on the modulation of the shape of the active-site cavity to direct substrate specificity and the cyclization fate of elongated intermediates [14]. The subsequent structure-guided mutagenic conversion of alfalfa CHS into a functional stilbene synthase confirmed the architectural underpinnings of the 'aldol switch' and the structural changes responsible for aldol-cyclization specificity in STSs [20••].

Indeed, CHS, STS, and CHS-like enzymes constitute an expanding family of plant and microbial PKSs, which give rise to chemical diversity and ultimately physiological and ecological diversity in their host organisms. This large family of enzymes serves as the central catalysts in the biosynthesis of an emerging class of synthetically related plant compounds that are not derived from phenylalanine. Several plant PKSs that are related to CHS and STS by a significant level of primary sequence homology, including bibenzyl synthase (BBS), acridone synthase (ACS) and pyrone synthase (PS), share a common chemical mechanism with CHS and STS but differ from CHS in their substrate specificity and/or in the regio-chemistry of the polyketide cyclization reaction [14].

Recent members of the type III PKS family from plants

The St. John's wort family of plants (*Hypericum*) produce bioactive benzophenone derivatives. A novel member of the CHS-like family, known as benzophenone synthase

(BPS), has been cloned from cell cultures of *Hypericum androsaemum* [21]. BPS shares considerable sequence identity with authentic CHS cloned from the same cell culture (60%), but it displays a distinct specificity for a benzoyl-CoA starter that is elongated by the stepwise condensation of three malonyl-CoAs to produce 2,4,6-trihydroxybenzophenone. When three residues that line the active-site cavity of CHS were replaced by the corresponding residues from BPS, the result was the preferential loading of a benzoyl moiety onto the mutant CHS [21].

Until recently, one mechanistic feature of the type III PKSs, which have a simpler architecture than their type I and type II cousins, has been their restriction to the formation of smaller polyketides of no more than five ketide units. However, Abe *et al.* [22••] obtained a cDNA clone from rhubarb (*Rheum palmatum*) that is capable of forming an aromatic heptaketide, known as aloesone, from an acetyl-CoA starter and six successive condensation reactions with malonyl-CoA before a terminating cyclization reaction. Probable specificity determinants for the unprecedented production of a heptaketide in a type III biosynthetic scaffold have been elucidated on the basis of the 60% sequence identity between aloesone synthase (ALS) and rhubarb CHSs, and the resultant three-dimensional homology models [22••].

Evolution of phenylpropanoid biosynthesis

Many derived classes of phenylpropanoids and biosynthetically related compounds are present in advanced plant groups, and certain biosynthetic enzymes are similar to those of primary metabolism. These discoveries have given rise to considerable speculation that phenylpropanoid biosynthetic systems in plants have been built up in a stepwise fashion by successive recruitment of enzymes from primary metabolism and/or as the product of an endocellular cooperation of a green alga and a fungus-like organism [23]. Although these pathways have been extensively characterized in several higher plants, there is little information about the pathway during its first 400–700 million years of existence. The information that we do have about the early evolution of phenylpropanoid biosynthesis has been inferred from limited examination of basal members of the terrestrial plant family [24], from the origins of the pathway in the green algae during the Late Precambrian Era until the emergence of the core conserved group of genes that are found in nearly all advanced angiosperms in the Early Cretaceous Period. A recent report describing the fossilized remains of plants that contain spores dated to 474 million years ago, together with ultrastructural analysis of the spore walls, supports a close relationship of these fossilized plants with modern-day liverworts [25•]. The discovery of novel flavones and flavanones in the liverwort *Tylimanthus renifolius* [26], the recent isolation of a CHS-like gene from the liverwort *Marchantia paleacea* [27•], and gene discovery efforts in ‘living fossils’ of ancient seed plants such as Cycads [28] promise to reveal the early roots of phenylpropanoid biosynthesis. When combined with comparative structural studies across kingdoms of a wide diversity of phenylpropanoid-metabolizing enzymes and structurally related enzymes of primary and secondary metabolism, a logical model for the emergence and continued functional divergence of polyketide-associated secondary metabolism in plants and other organisms might be within reach.

Conclusions

Biodiversity exists on multiple levels from the ecological to the atomic. At the atomic level, the metabolic and chemical diversity found in Nature constitutes a rich record, reporting on the evolution of enzymes and metabolic pathways. Nowhere is this record more expansive than in metabolic pathways that are associated with so-called secondary metabolism in microbes and plants. The tools of structural biology hold great promise for expanding our

understanding of the mechanistic enzymology that governs the formation of the chemicals of secondary metabolism: chemicals that confer a multitude of evolutionary advantages on the host, particularly with regard to host defense, increased fitness and the establishment of mutually beneficial inter-species relationships.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. MCB-0236027 and MCB-0312466 to J.P.N.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1. Borevitz JO, Ecker JR. Plant genomics: the third wave. *Annu Rev Genomics Hum Genet.* 2004; 5:443–477. [PubMed: 15485356] This review constitutes a comprehensive analysis of the methods and approaches employed in modern-day genomic analyses in plants.
2. Pichersky E, Gang DR. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 2000; 5:439–445. [PubMed: 11044721]
3. Kim SH, Shin DH, Choi IG, Schulze-Gahmen U, Chen S, Kim R. Structure-based functional inference in structural genomics. *J Struct Funct Genomics.* 2003; 4:129–135. [PubMed: 14649297]
4. Pal D, Eisenberg D. Inference of protein function from protein structure. *Structure.* 2005; 13:121–130. [PubMed: 15642267] This paper details how one can extract important functional inferences from the structures of proteins.
5. Goldsmith-Fischman S, Honig B. Structural genomics: computational methods for structure analysis. *Protein Sci.* 2003; 12:1813–1821. [PubMed: 12930981] This overview of structural genomics provides a flow chart for computational approaches to analyze and extract information from protein structures.
6. Naismith JH. Chemical insights from structural studies of enzymes. *Biochem Soc Trans.* 2004; 32:647–654. [PubMed: 15493979]
7. Winkel BS. Metabolic channeling in plants. *Annu Rev Plant Biol.* 2004; 55:85–107. [PubMed: 15725058]
8. Whittle E, Shanklin J. Engineering delta 9–16:0-acyl carrier protein (ACP) desaturase specificity based on combinatorial saturation mutagenesis and logical redesign of the castor delta 9–18:0-ACP desaturase. *J Biol Chem.* 2001; 276:21500–21505. [PubMed: 11294879]
9. Singh SP, Zhou XR, Liu Q, Szymne S, Green AG. Metabolic engineering of new fatty acids in plants. *Curr Opin Plant Biol.* 2005; 8:197–203. [PubMed: 15753001]
10. Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. *Plant Cell.* 1995; 7:1085–1097. [PubMed: 12242399]
11. Baedeker M, Schulz GE. Autocatalytic peptide cyclization during chain folding of histidine ammonia-lyase. *Structure.* 2002; 10:61–67. [PubMed: 11796111]
12. 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] The authors describe the crystal structure of a fungal PAL that bears a striking structural homology with plant PALs. They provide a provocative new mechanistic hypothesis to account for the PAL-mediated conversion of phenylalanine to *trans*-cinnamic acid, and thus set the stage for possible rational routes to the production of novel precursors for plant and bacterial polyketide synthases.
13. Ritter H, Schulz GE. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. *Plant Cell.* 2004; 16:3426–3436. [PubMed: 15548745]

The structural studies outlined in this paper provide the first structure of a plant PAL and describe the architectural features that distinguish it from previously reported structures of bacterial HALs.

14. Austin MB, Noel JP. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep.* 2003; 20:79–110. [PubMed: 12636085]
15. Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol.* 2003; 7:285–295. [PubMed: 12714063] In this review, the author describes various strategies used by Nature to elaborate complex polyketide natural products.
16. Saxena P, Yadav G, Mohanty D, Gokhale RS. A new family of type III polyketide synthases in *Mycobacterium tuberculosis*. *J Biol Chem.* 2003; 278:44780–44790. [PubMed: 12941968]
17. Austin MB, Izumikawa M, Bowman ME, Udway DW, Ferrer JL, Moore BS, Noel JP. Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. *J Biol Chem.* 2004; 279:45162–45174. [PubMed: 15265863] This study, along with the next [18•], demonstrates the close architectural correspondence of plant and bacterial type III PKSs.
18. Sankaranarayanan R, Saxena P, Marathe UB, Gokhale RS, Shanmugam VM, Rukmini R. A novel tunnel in mycobacterial type III polyketide synthase reveals the structural basis for generating diverse metabolites. *Nat Struct Mol Biol.* 2004; 11:894–900. [PubMed: 15286723] This study, along with the previous [17•], demonstrates the close architectural correspondence of plant and bacterial type III PKSs.
19. Tropf S, Lanz T, Rensing SA, Schroder J, Schroder G. Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. *J Mol Evol.* 1994; 38:610–618. [PubMed: 8083886]
20. Austin MB, Bowman ME, Ferrer JL, Schroder J, Noel JP. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. *Chem Biol.* 2004; 11:1179–1194. [PubMed: 15380179] The structure–function relationship of stilbene synthases are elucidated, providing an unexpected explanation for the divergent cyclization behavior of stilbene synthases and a proof-of-principle method for modulating the cyclization behavior of plant type III PKSs.
21. Liu B, Falkenstein-Paul H, Schmidt W, Beerhues L. Benzophenone synthase and chalcone synthase from *Hypericum androsaemum* cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases. *Plant J.* 2003; 34:847–855. [PubMed: 12795704]
22. Abe I, Utsumi Y, Oguro S, Noguchi H. The first plant type III polyketide synthase that catalyzes formation of aromatic heptaketide. *FEBS Lett.* 2004; 562:171–176. [PubMed: 15044020] The landmark discovery described in the title of this paper demonstrates the amazing elongating potential of an architecturally simple type III PKS from plants.
23. Jorgensen R. The origin of land plants: a union of alga and fungus advanced by flavonoids? *Biosystems.* 1993; 31:193–207. [PubMed: 8155852]
24. Yamazaki Y, Suh DY, Sitthithaworn W, Ishiguro K, Kobayashi Y, Shibuya M, Ebizuka Y, Sankawa U. Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. *Planta.* 2001; 214:75–84. [PubMed: 11762173]
25. Wellman CH, Osterloff PL, Mohiuddin U. Fragments of the earliest land plants. *Nature.* 2003; 425:282–285. [PubMed: 13679913] The authors of this paper examined of the fossil record of early land plants and established a relationship between the ancestors of modern terrestrial plants and primitive liverworts that are still in existence.
26. Feld H, Zapp J, Becker H. Secondary metabolites from the liverwort *Tylimanthus renifolius*. *Phytochemistry.* 2003; 64:1335–1340. [PubMed: 14629996]
27. Harashima S, Takano H, Ono K, Takio S. Chalcone synthase-like gene in the liverwort, *Marchantia paleacea* var. *diptera*. *Plant Cell Rep.* 2004; 23:167–173. [PubMed: 15168073] The work described in this report uncovered the first CHS-like gene in a primitive terrestrial plant that might represent a modern-day relative of the earliest terrestrial plants.
28. Brenner ED, Stevenson DW, McCombie RW, Katari MS, Rudd SA, Mayer KF, Palenchar PM, Runko SJ, Twigg RW, Dai G, et al. Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant. *Genome Biol.* 2003; 4:R78. [PubMed: 14659015]