The Characterization of Chalcone Synthase from the Moss, *Physcomitrella patens*, and the Investigation of the Cyclization Mechanism of Stilbene Synthase

A Thesis

Submitted to the Faculty of Graduate Studies and Research

In Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Biochemistry

University of Regina

by

Chenguang Jiang

Regina, Saskatchewan

August, 2006

Your file Votre référence
ISBN: 978-0-494-20216-6
Our file Notre référence
ISBN: 978-0-494-20216-6

UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

SUPERVISORY AND EXAMINING COMMITTEE

Chenguang Jiang, candidate for the degree of Master of Science, has presented a thesis titled, **The Characterization of Chalcone Synthase from the Moss, *Physcomitrella patens*, and the Investigation of the Cyclization Mechanism of Stilbene Synthase,** in an oral examination held on August 28, 2006. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner:	Dr. Harold Weger, Department of Biology
Supervisor:	Dr. Dae-Yeon Suh, Department of Chemistry and Biochemistry
Committee Member:	Dr. Tanya Dahms, Department of Chemistry and Biochemistry
Committee Member:	*Dr. R. Scott Murphy, Department of Chemistry and Biochemistry
Chair of Defense:	Dr. Randal Lewis, Department of Physics

*Unable to attend defense

Abstract

Polyketide synthases (PKS) catalyze the biosynthesis of various natural products in plants, fungi, and bacteria with different biological activities and pharmacological properties. The members of the chalcone synthase (CHS) superfamily, also known as the type III PKSs, function as homodimeric iterative PKSs with two independent multifunctional active sites. CHS is involved in the first committed step in the biosynthesis of flavonoids that serve as the primary UV protective pigments in plants. Flavonoids are found in liverworts and mosses (bryophyte) that are thought to be similar to the earliest extant land plants. Although higher plant CHSs have been extensively studied, little information is available on the enzymes from bryophytes. Using a reverse transcription polymerase chain reaction strategy, a CHS from the moss, *Physcomitrella patens* (*P. patens*) was cloned and characterized. *P. patens* CHS exhibited similar kinetic properties and substrate preference profiles as those of higher plants CHSs. *p*-Coumaroyl-CoA was the most preferred starter unit, suggesting that *P. patens* CHS is a naringenin producing CHS. Consistent with the evolutionary position of the moss, a phylogenetic analysis placed *P. patens* CHS at the base of the plant CHS clade, next to the microorganism CHS-like gene products. Therefore, *P. patens* CHS likely represents a modern day version of one of the oldest CHSs that appeared on earth. The evolutionary relationship of the CHS superfamily with other condensing enzymes was also investigated. Stilbene synthase (STS), another plant type III PKS, catalyzes condensation reactions of a phenylpropanoid-CoA with three molecules of malonyl-CoA to form the tetraketide intermediate and then cyclizes it to produce stilbene. The cyclization process of STS involves hydrolysis, decarboxylation, ring formation, and aromatization;

however, the sequence of these chemical events remained unknown. A putative intermediate, *p*-coumaroyltriacetic acid was prepared and converted to a stilbene, resveratrol, by STS. This indicates that the first step of the STS cyclization reaction is the hydrolysis of the thioester bond of the tetraketide intermediate; and provides a foundation to further investigate the sequence of the remaining cyclization steps in the STS cyclization reaction.

Acknowledgement

I am grateful to Dr. Dae-Yeon Suh (University of Regina, Regina, Canada) for his great help and encouragement of my research work, and critical reading and suggestions of the thesis. I thank Dr. N. Ashton (University of Regina, Regina, Canada) for providing gametophyte tissues of the moss *Physcomitrella patens*. I am grateful to Dr. Sun Young Kim for assistance and helpful suggestions, to my colleagues Graeme Gordon and Clark K. Schommer for material preparation and useful discussion. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Department of Chemistry & Biochemistry, and the Faculty of Graduate Studies and Research of the University of Regina.

This thesis is divided into two independent parts: The cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*; and the investigation of the cyclization mechanism of stilbene synthase.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3-D	three-dimensional
6-MSA	6-methylsalicylic acid
ACP	acyl carrier protein
ACS	Acridone synthase
BAS	benzalacetone synthase
BNY	bisnoryangonin
BPS	benzophenone synthase
CHS	chalcone synthase
CoA	coenzyme A
CTAA	*p*-coumaroyltriacetic acid
CTAL	*p*-coumaroyltriacetic acid lactone
CTAS	*p*-Coumaroyltriacetic acid synthase
DTT	dithiothreitol
EST	expressed sequence tag
FAS	fatty acid synthases
h	hours
IPTG	isopropyl-*β*-D-thiogalactopyranoside
KAS	β-ketoacyl synthase
KPi	potassium phosphate
MAAs	mycosporine like amino acids
min	minutes
MP	maximum parsimony

NJ	neighbour-joining
ORF	Open reading frame
PAGE	polyacylamide gel electrophoresis
PCR	Polymerase chain reaction
PIBPS	phloroisobutyrophenone synthases
PIVPS	phloroisovalerophenone synthases
PKS	polyketide synthases
RACE	Rapid Amplification of cDNA Ends
RP-HPLC	reversed phase-high performance liquid chromatography
RP-TLC	reversed phase-thin layer chromatography
s	seconds
SDS	sodium dedocyl sulfate
SPS	styrylpyrone synthase
STS	stilbene synthase
TFA	trifluoroacetic acid
UTR	Untranslated regions
UV	Ultraviolet
V_b	bed volume

Part I: Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*

1. Introduction

1. 1. Polyketide synthases

Natural products, mostly built up from an array of simple building blocks derived from single or multiple primary metabolism pathways, contain a diverse collection of small compounds providing various biological advantages to their hosts. Polyketides are a large family of structurally diverse natural products found in bacteria, fungi and plants, possessing broad ranges of biomedical and pharmacological properties.[1] They are biosynthesized from acyl coenzyme A (CoA) precursors by polyketide synthases (PKS), which catalyze the sequential Claisen condensations of two-carbon acetate units derived from malonyl-CoA with an activated starter unit into a growing polyketide chain, in a parallel manner that mirrors the fatty acid synthases (FAS) of primary metabolism. All PKSs possess the key β-ketoacyl synthase (KAS) domain involved in carbon-carbon bond formation.[2] Although PKSs share a common chemical strategy for the assembly of polyketide scaffolds, they can utilize various starter groups and extender units to control the size of the final polyketides, thus generating a broad range of molecular scaffolds.

On the basis of molecular genetics and structural and mechanistic characteristics, PKSs are classified into three groups. The type I PKS consists of one or more large multifunctional polypeptides that are organized into modules. Each module possesses distinct, non-iterative activities responsible for the catalysis of one cycle of polyketide chain elongation. The type I PKS, resembling the yeast and animal FASs, can be processive as in FAS and macrolide synthases. One example is 6-deoxyerythronolide synthase that contains six sets of most or all of the core FAS domains (Fig. 1).[3]

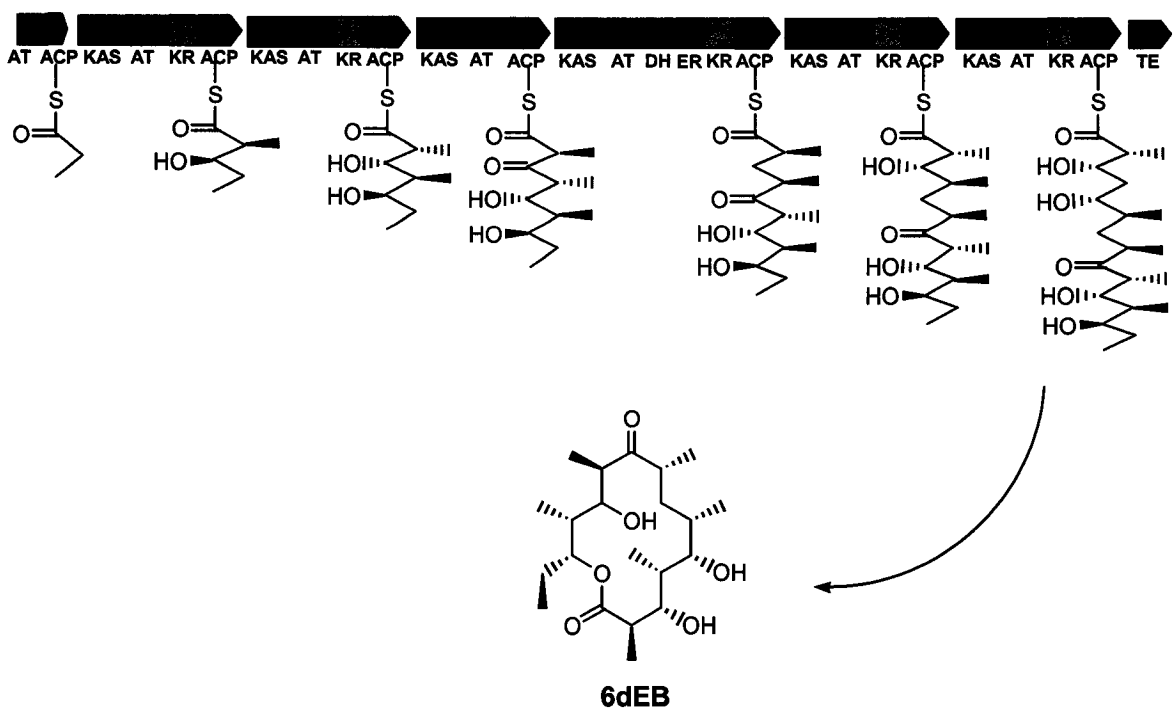
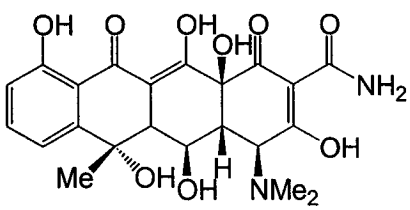


Figure 1. Genetic architecture of DEBS (deoxyerythronolide synthase), a processive modular type I PKS. AT: acetyl-CoA-ACP transacetylase; ACP: acyl carrier protein; KAS: β-ketoacyl-ACP synthase; KR: β-ketoacyl-ACP reductase; DH: β-Hydroxyacyl-ACP dehydratase; ER: Enoyl-ACP reductase; TE: thioesterase; 6dEB: 6-deoxyerythronolide B.

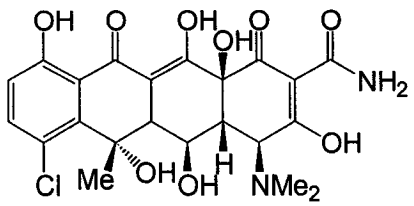
Alternatively, the type I PKS can be iterative as in the biosynthesis of 6-methylsalicylic acid (6-MSA) in the fungus *Penicillium patulum*. 6-MSA is produced by an homotetrameric type I PKS, using all FAS core domains to link four acetate units.[4]

Similar to the type II FAS systems found in plants and bacteria, the type II PKSs are iterative monofunctional protein complexes that generate poly-β-keto intermediates and catalyze cyclodehydrations to produce multicycle, aromatic polyketides. Actinorhodin[5] and tetracenomycin[6] are produced by such type II PKSs (Fig. 2). Each type II PKS contains a minimal set of three subunits, two β-ketoacyl synthase subunits and an acyl carrier protein (ACP) subunit to which the growing chain is attached. Additional subunits such as cyclase and aromatase are responsible for modification of the nascent chain to form the final cyclized structures.[7, 8]

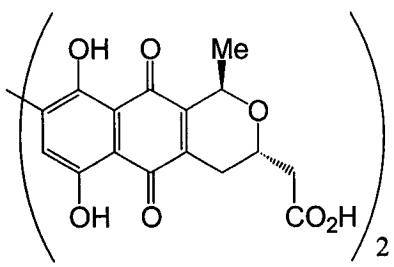
In type I and II PKSs, the acyl CoA substrates are transferred to the acyl carrier protein and the growing polyketide chain is covalently linked to the phosphopantetheine prosthetic group on ACP. The type I PKSs may retain the β-ketone group or modify it to a hydroxyl, methyne, or a methylene group, as dictated by the different domain activities as shown in Fig. 1. The linear sequential module corresponds to the sequence of extender units in the polyketide product, where the individual module has specific enzyme activities for each extension cycle and different modification. However, in type II PKSs, the β-ketone groups are predominantly left intact and the reactive polyketide backbone undergoes enzyme-catalyzed intramolecular cyclization reactions generating a range of aromatic structures. Although there have been NMR and X-ray crystallographic studies on the ACP and thioesterase domains from bacteria[9, 10] and fungi[11], there remains a lack of detailed information on the structure-based mechanisms for



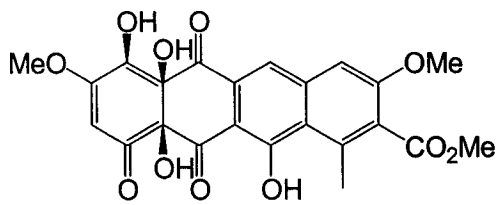
Oxytetracycline



Chlortetracycline



Actinorhodin



Tetracenomycin C

Figure 2. Aromatic polyketides produced by the type II polypeptide synthase.

the type I and II PKSs, which limits a complete understanding of the substrate selectivity, polyketide chain elongation, and cyclization specificity in these multi-functional enzymes.

The type III PKSs are structurally and mechanistically distinct from the type I and II PKSs. These enzymes, also known as the chalcone synthase (CHS) superfamily enzymes, are homodimers and iteratively act as condensing enzymes. The type III PKSs use CoA esters of simple carboxylic acids without the involvement of ACP for substrate delivery to the active site. These enzymes synthesize the small aromatic metabolites, as typified by CHS that catalyzes the biosynthesis of an aromatic polyketide, chalcone.[12] The type I and type II PKSs use multiple active sites or accessory enzymes for additional chemical modifications of the polyketide skeleton. On the other hand, each monomer of the type III PKSs contains an active site that loads the starter unit onto a catalytic cysteine, and carries out multiple decarboxylative condensation reactions and subsequent cyclization of the polyketide products.

Since the first reports of the different types of PKSs from diverse species, the current PKS paradigms have served to provide the molecular basis to classify and explain the vast structural and functional diversity of the polyketide natural products. However, some examples of recently discovered PKSs such as non-iterative type II PKSs and type I PKSs associated with discrete acyltransferase enzymes are leading to the argument that PKSs have much greater diversity in both structure and mechanism than previously realized.[13] This serves as an inspiration to search for novel PKSs with unprecedented biosynthetic machinery, which may provide further insights into polyketide biosynthesis.

1. 2. The chalcone synthase superfamily

The type III PKSs are structurally simple homodimers that utilize the conserved Cys-His-Asn catalytic triad as the key amino acid residues in a single active site to catalyze the iterative condensation of acetyl units to a starter CoA.[14] Following this carbon chain elongation, the linear polyketide intermediate is cyclized in the same active site. The simple architecture of type III PKS makes them much more amenable to *in vitro* manipulation and detailed structural and functional analysis than the type I and type II PKSs. Recent structural evidence suggests that the type III PKSs have acquired their function from the structurally similar homodimeric β-ketoacyl carrier protein synthase III (KAS III)[15], which initiates the fatty acid biosynthesis by linking two acetate units to produce a C_4 compound, acetoacetyl-CoA as illustrated in Fig. 3. However, type III PKSs have gained multiple activities that can handle a diverse set of much larger starter molecules. These mechanistically complex type III PKSs have gained ability to utilize four distinct polyketide cyclization mechanisms (This will be discussed in detail in Part II.), all the while maintaining a simple homodimeric organization of KAS III.

CHS catalyzes the formation of chalcone from a phenylpropanoid-CoA and three molecules of malonyl-CoA, the first committed step of flavonoid biosynthesis. It is not surprising that CHS was the first type III PKS that was discovered since it is ubiquitous and important in plant secondary metabolism. Duplication and functional divergence of the CHS gene in plants has given rise to an expanding superfamily of CHS-like enzymes, many of which have only recently been cloned and functionally characterized. This important class of divergent CHS-like enzymes are referred to as the CHS superfamily. The members of this superfamily can differ from CHS in their choice of CoA-tethered

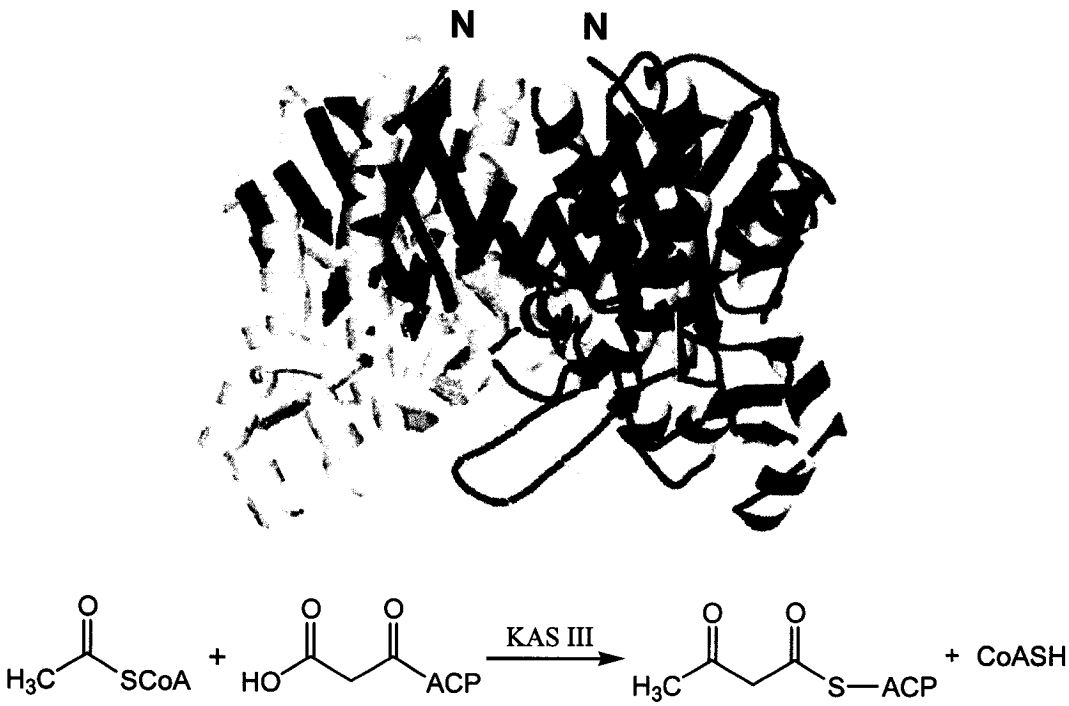


Figure 3. Ribbon diagram from the crystal structure of KAS III and its reaction scheme of KAS III. (One monomer is shown in yellow and green; the other is shown in cyan and magenta. ACP: acyl carrier protein. Helices and random coil are rendered in yellow and cyan, while beta sheets are rendered in green and magenta.)

starter molecule, in the number of polyketide elongation steps, and in their mechanism of intramolecular cyclization of the linear polyketide intermediate.

One group of the CHS superfamily catalyzes three-condensation reactions with starter units other than phenylpropanoid esters, while another group catalyzes only one or two condensation reactions. Acridone synthase (ACS), benzophenone synthase (BPS), phloroisovalerophenone and phloroisobutyrophenone synthases (PIVPS and PIBPS) belong to the first group while benzalacetone synthase (BAS) and styrylpyrone synthase (SPS) fall into the second group.[16, 17]

ACS, mostly found in the Rutaceae family (including some *Citrus* species), shares over 65% amino acid identity with most CHSs[18, 19], and catalyzes the pivotal reactions in the formation of acridone alkaloids (Fig. 4). Acridone alkaloids are synthesized from *N*-methylanthraniloyl-CoA and malonyl-CoA, and the new aromatic ring system is formed by a Claisen type cyclization. The product, 1,3-dihydroxy-*N*-methylacridone, is the result of a second ring closure and it is unknown whether this is an intrinsic activity of the enzyme or a non-enzymatic process.

Xanthones are a group of tricycle natural products with useful pharmaceutical properties and are found predominately in plants of the Gentianaceae and Hypericaceae families. Plant xanthones possess antifungal properties, although their physiological roles have not been established. Xanthone is synthesized by a region-specific ring closure of benzophenone, which is produced by BPS.[20] BPS catalyzes three acetyl additions to a benzoyl-CoA starter unit, followed by a CHS-like intramolecular Claisen cyclization and aromatization of the resultant tetraketide intermediate to give 1,3-dihydroxy-

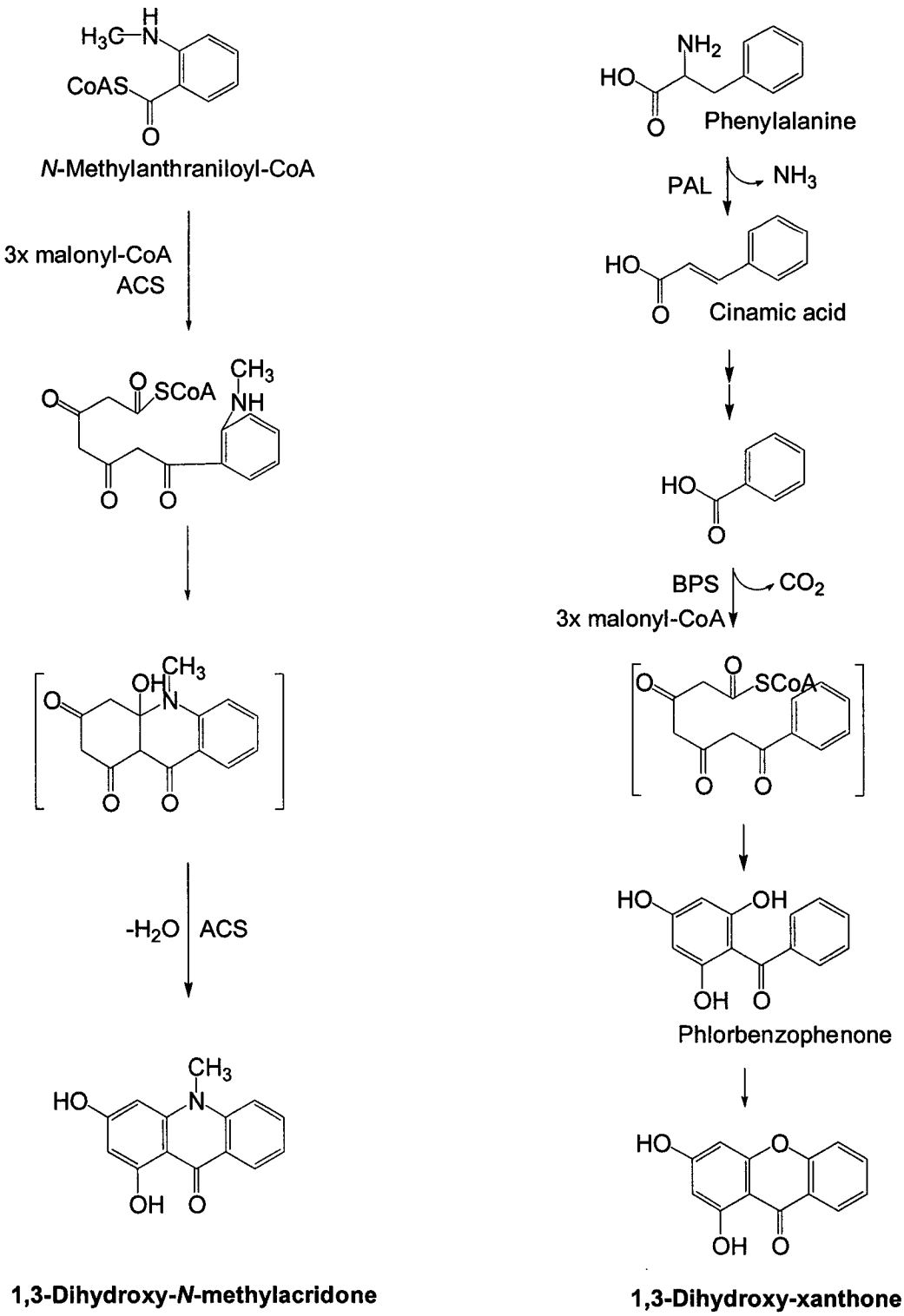


Figure 4. Reactions catalyzed by acridone synthase (ACS) and benzophenone synthase (BPS).

xanthone as the final product (Fig. 4).

The ripe cones of hops (*Humulus lupulus*, Cannabaceae) contain up to 20% of the polyol compounds including humulone and cohumulone. Based on the unambiguous detection of phloroisovalerophenone (PIVP) and phloroisobutyrophenone (PIBP) as intermediates, it was suggested that the formation of the aromatic ring precedes all of the prenylation steps.[21] PIVP and PIBP are synthesized by phloroisovalerophenone synthase and phloroisobutyrophenone synthase, respectively, *via* the CHS-like Claisen type cyclization, using isovaleryl-CoA and isobutyryl-CoA as starter units and three molecules of malonyl-CoA as extender units in condensation reactions (Fig. 5).

The characteristic compounds of the raspberry, *p*-hydroxyphenylbutan-2-one (*p*HPB, also termed raspberry ketone) or its glycosides have been found in *Pinus contorta*, *Vaccinium oxycoccus* (European cranberries), *Hippophae rhamnoides* (sea buckthorn) and *Scutellaria rivularis*. The two step biosynthesis of *p*HPB is initiated by a type III PKS called benzalacetone synthase (BAS). Isolated from raspberries and rhubarb, BAS is a plant-specific PKS that catalyzes a single decarboxylative condensation of malonyl-CoA to the *p*-coumaroyl-CoA followed by decarboxylation to form *p*-hydroxyphenylbut-3-ene-2-one (Fig. 6).[22]

Gametophytes and rhizomes of *Equisetum arvense* (horsetail) accumulate styrylpyrone as the major phenolic constituents. Styrylpyrones are common constituents in fungi, but also found in pteridophytes and angiosperm families.[23, 24] Styrylpyrones are presumably synthesized by styrylpyrone synthase (SPS) from the same starter and extender units used by CHS, but with only two condensation reactions followed by lactonization of the triketide intermediates to form the styrylpyrones as shown in Fig. 6.

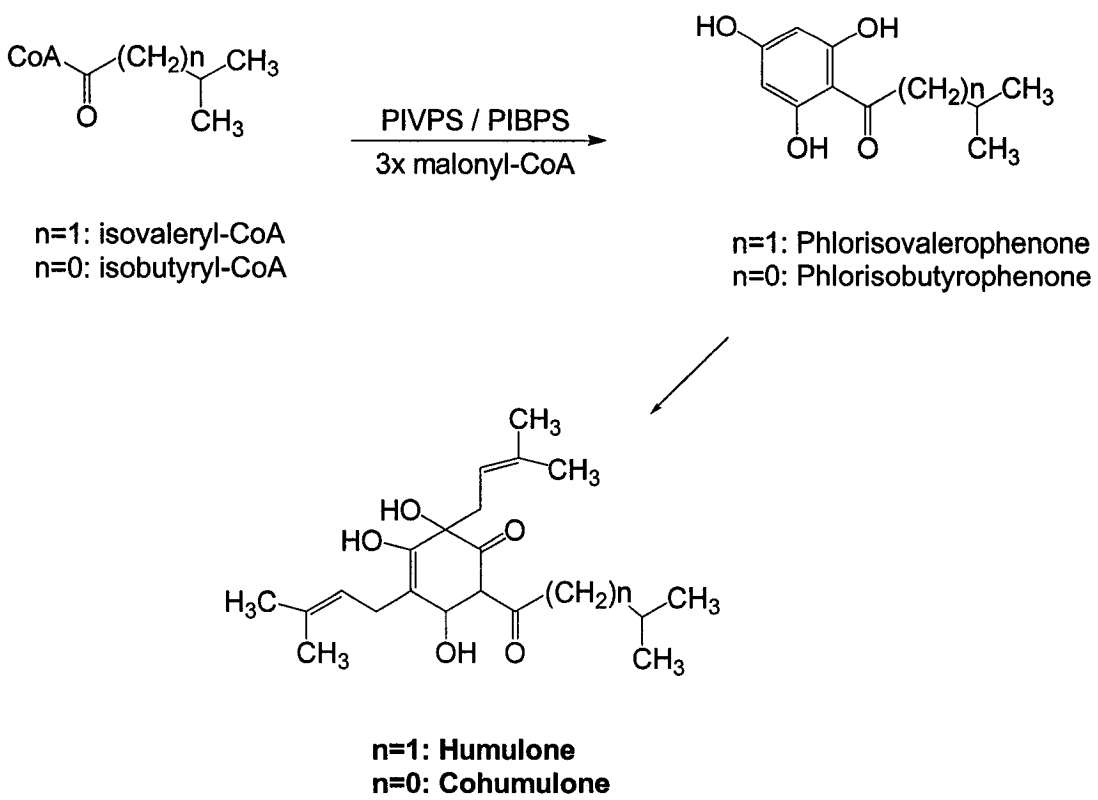


Figure 5. Reaction schemes for the phloroisovalerophenone and phloroisobutyrophenone synthases.

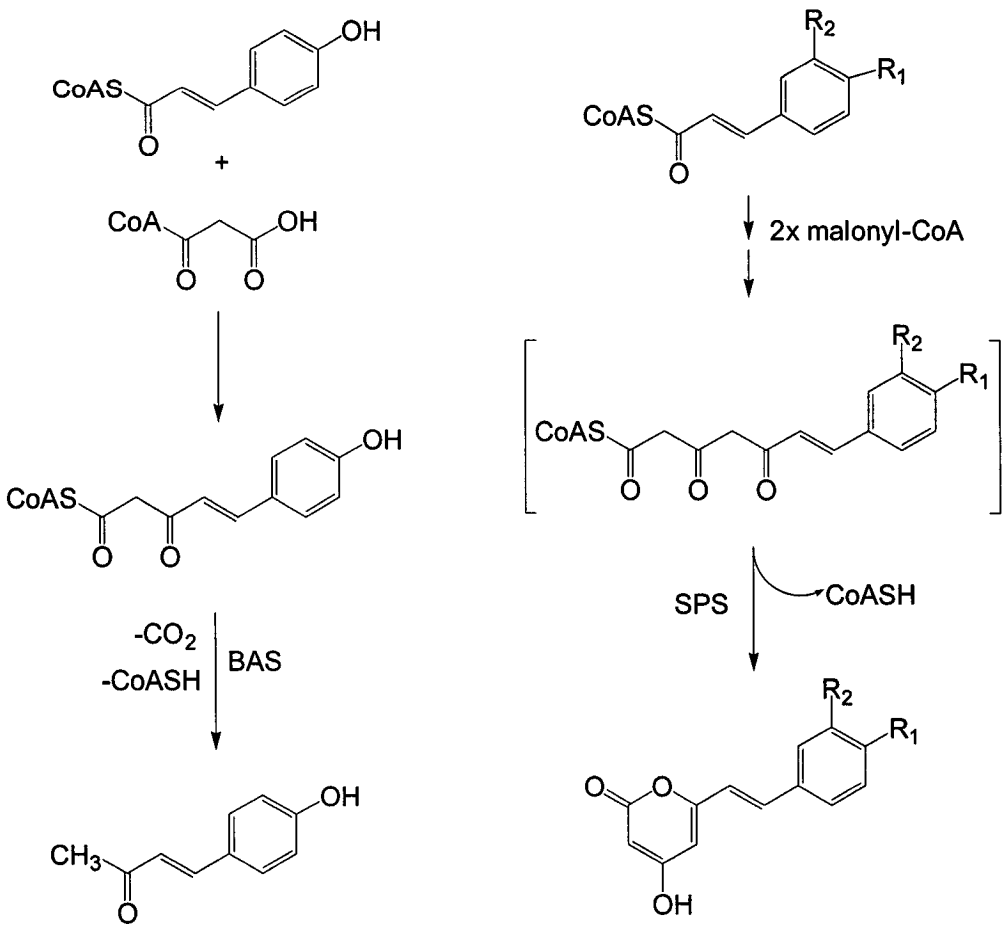


Figure 6. One and two condensation reaction schemes catalyzed by benzalacetone synthase (BAS) and styrylpyrone synthase (SPS).

Several CHS-like enzymes have also been found in bacteria and fungi. Based on their scarcity, lower sequence similarities to plant CHSs, and the uncertain evolutionary history of the type III PKSs, it had been suggested that a few bacteria had obtained CHS-like genes only recently *via* horizontal gene transfer from plants. However, the recent rapid expansion of available genome sequences has since revealed many more bacterial CHS-like sequences, most of which share over 25% amino acid sequence identity with plant CHSs and with each other. Further, functional similarities to KAS III domains indicate that bacterial CHS-like enzymes may have evolved from their KAS III ancestors similar to the evolutionary manner of the plant CHS superfamily. Bacterial CHS-like enzymes are much more structurally and functionally divergent than their plant counterparts, and participate in a variety of secondary metabolic pathways.[25]

Investigation of CHS superfamily enzymes from plants, bacteria and fungi has allowed us to integrate the structural basis of functional diversity within this superfamily with phytochemical and biological information. Mechanistic knowledge gained from the crystal structures of CHS and related condensing enzymes, combined with information gained from studies of several functionally divergent CHS superfamily enzymes offer a good example of the evolutionary process of metabolic divergence, the factor playing a significant role in the remarkable diversity of life on earth.

1. 3. Chalcone synthase

CHS, the best studied type III PKS, catalyzes the condensation of one *p*-coumaroyl-CoA and three malonyl-CoA molecules into chalcone as shown in Fig. 7.[26]

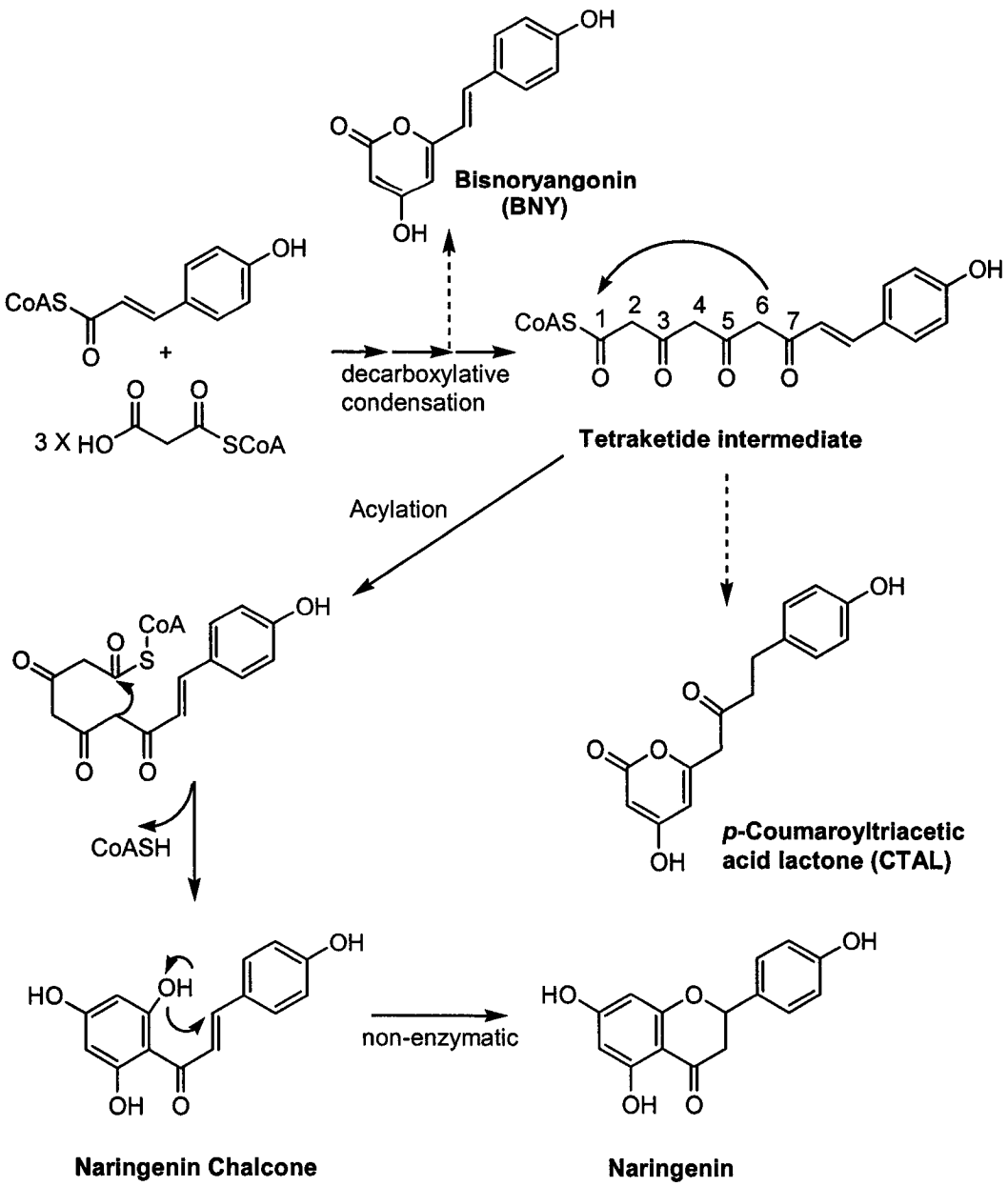


Figure 7. The reaction scheme for the chalcone synthase-catalyzed reaction and proposed mechanism.

First, CHS loads the *p*-coumaroyl starter moiety from the CoA-linked starter molecule onto the catalytic cysteine at the active site to initiate the reaction. Then the sequential condensation of three acetate units, derived from three malonyl-CoA molecules, with the enzyme-bound coumaroyl moiety forms the tetraketide intermediate. The final step is the cyclization of the linear tetraketide intermediate *via* an intramolecular Claisen condensation followed by aromatization to produce the final product, chalcone. It was proposed that the ring is closed through an internal proton transfer from the nucleophilic carbon to the carbonyl oxygen.[27] It would follow that the naringenin chalcone is synthesized by the subsequent aromatization of the trione ring through a second series of internal proton transfers.[28] Under certain *in vitro* reaction conditions, bisnoryangonin (BNY, the derailed lactone after two condensations) and *p*-coumaroyltriacetic acid lactone (CTAL, the derailed lactone after three condensations) are also produced as byproducts during the CHS reaction, due to nonenzymatic hydrolysis of the polyketide intermediates (Fig. 7).[29]

Chalcones are crucial secondary metabolites in the biosynthesis of anthocyanin pigments, antimicrobial phytoalexins, and flavonoid inducers of *Rhizobium* nodulation genes. In the 1950s, the wide distribution of anthocyanin flower pigments and other flavonoid products in higher plants[30] attracted the attention of phytochemists, who then predicted the incorporation of acetate units into the flavonoid skeleton. In 1972, Kreuzaler and Hahlbrock demonstrated the biosynthesis of naringenin chalcone from malonyl-CoA and the phenylalanine-derived *p*-coumaroyl-CoA by isotopic labelling studies.[31] It was subsequently discovered in 1983 that, *in vivo*, the released chalcone product of CHS undergoes an additional ring closure rapidly and stereospecifically

through the catalysis of chalcone isomerase.[32] This Michael-type ring closure can also be spontaneous but nonstereospecific in aqueous solution following the release of chalcone from the CHS active site; thereby naringenin appears to be the main product in the *in vitro* enzyme assay as depicted in Fig. 7.[33]

Due to their biochemical diversity, multiple *chs* genes are present in most plants. The expression patterns of these multiple genes vary across different plant species, within different tissues, and at different developmental stages during the life cycle of an individual plant. Some of these genes are constitutively expressed while others are transcriptionally induced by environmental factors, such as UV light and pathogen infestation, reflecting the diverse biological roles of the various flavonoid metabolism products in plants.[16]

The crystal structure of CHS[15], as shown in Fig. 8, revealed that the enzyme contains the characteristics conserved in all condensing enzymes including the location of the active site, the five-layer αβαβα structure, and the dimerization interface. The CHS dimerization interface contains both hydrophilic and hydrophobic residues, and is a fairly flat surface with the exception of a pair of entwined N-terminal helices believed to serve as the intracellular localization signal that does not affect the CHS reaction.[34] Each CHS monomer consists of two structural (upper and lower) domains and the active site of CHS is buried within an interior cavity located at the cleft between these domains of each monomer. The upper domain, exhibiting a similar motif to KAS III in fatty acid synthesis, maintains the three-dimensional position of the catalytic residues like Cys, His, and Asp. Cys164 serves as a nucleophile in the condensation reaction and shuttle reaction intermediates *via* CoA thioester-linked molecules. In the lower

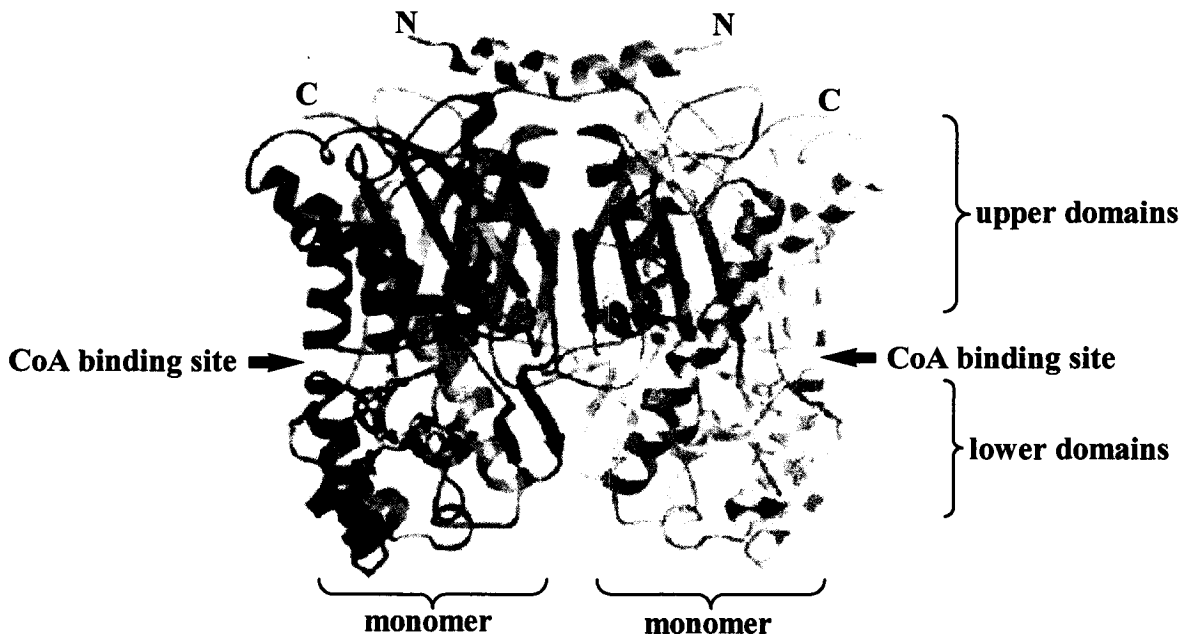


Figure 8. Ribbon diagram of the chalcone synthase crystal structure. (One monomer is shown in gold; the other is shown in blue. Enzyme-bound substrates (red and whilte) are shown as stick models.)

domain of CHS, there is a larger active site cavity compared to that of KAS III, providing space for the tetraketide intermediate.

There are surprisingly few chemically reactive amino acids in the active site. Four active site residues including Cys164, Phe215[15], His303, and Asn336[27, 35] and the G^{372}FGPG loop[29] largely define the catalytic machinery of CHS (amino acid numbering according to alfalfa CHS). Unlike the corresponding flexible loop of SFGFG in KAS II, the G^{372}FGPG loop of the CHS superfamily enzymes contains a strictly conserved proline residue, which is believed to provide some rigidity and unique configuration in this loop (Fig. 9).[29]

When this Pro375 was mutated to Gly, CTAL was produced at the expense of chalcone, indicating that the integrity of the loop is essential for the proper folding for cyclization of the linear intermediate.[29] Although much is known about the CHS-catalyzed reactions, a few important mechanistic questions of the CHS reaction remain to be addressed, including the kinetic equivalence of the malonyl-CoA decarboxylation and condensation steps and the movement of intermediate CoA-thioesters into and out of the active site.

1. 4. Evolution of the chalcone synthase superfamily in plants

Life started very early on this planet; however, most of the life on this planet made its history entirely under water, partly since it took several billions of years for life to conquer the hostile radiation climate on land. Ultraviolet (UV) radiation can damage various cell components, including membrane lipids, functional proteins and DNA. The proposition that UV radiation was the major threat for early terrestrial life is supported

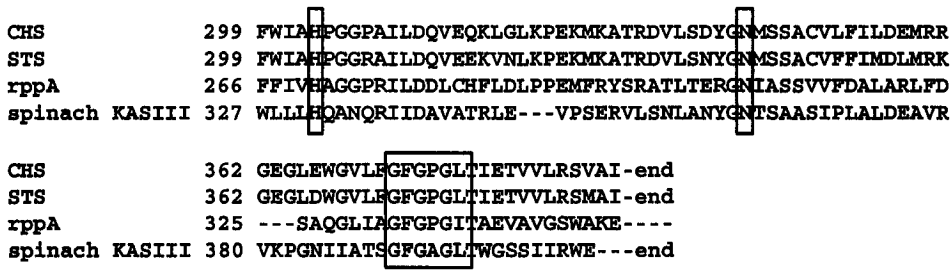


Figure 9. Alignment of the partial amino acid sequences of type III ketoacyl-acyl carrier protein synthase (KAS III), chalcone synthase, and stilbene synthase. RppA is a bacterial CHS-like enzyme.[25]

by the discovery that some of the repair systems for UV-damaged DNA are evolutionarily ancient.[36, 37] It is believed that aquatic green algae (charophytes) are the ancestors of plants[38], which first appeared on land more than 450 million years ago, and that the earliest land plants resembled modern bryophytes including hornworts, liverworts and mosses.[39] In green algae, the most important group of UV absorbing pigments are the mycosporine-like amino acids (MAAs), inefficient UV-protective compounds that are required in plentiful amounts.[40] MAAs also play a significant role in the UV defense system of cyanobacteria[41] and some aquatic animals like sea cucumbers and sea urchins.[42] Before the appearance of the first terrestrial plant, charophyte green algae already possessed the shikimate pathway or a part of the phenylpropanoid pathway, a crucial preparation for various aspects of terrestrial life.[39] These pathways are thought to have led to the biosynthesis of more efficient UV protective compounds, the most important of which are the flavonoids. Flavonoids have not been discovered in the green algae and hornworts, which are thought to be the earliest species among the bryophytes. On the other hand, about 40% of liverworts and over 50% of mosses contain flavonoids, which are ubiquitous in the posterities of bryophytes including the ferns and higher plants.[43] Thus, the chemical and functional development of UV absorbing compounds and the corresponding phenolic polymers such as flavonoids and lignins has likely played a major role in the evolution of land plants.

Enzymes of secondary metabolism have undoubtedly been derived from pre-existing enzymes, ultimately from those in primary metabolism *via* gene duplication and random mutations. For example, synthases, hydroxylases, and reductases of the flavonoid pathways may have evolved from the primary metabolism.[44, 45] Flavonoid evolution was

preceded by the evolution of the phenylpropanoid and malonyl-CoA pathways, as the initial step of flavonoid biosynthesis requires substrates from both the phenylpropanoid and malonyl-CoA pathways.[46] Solar UV-B radiation likely stimulated the development of the branch-point enzymes of the phenylpropanoid pathway such as CHS and phenylalanine ammonium lyase, the latter of which catalyses the transformation of phenylalanine to *trans*-cinnamic acid leading to the formation of complex phenolic compounds including flavonoids.[47] It is commonly believed that the development of CHS has paralleled the evolution of flavonoids. Thus, CHS and flavonoids are useful in understanding the evolution of plants.

As representative members of the CHS superfamily, CHS and stilbene synthase (STS) share common features in structure and function including their high amino acid sequence similarity and the same catalytic mechanism (decarboxylative Claisen condensation), leading to a common tetraketide intermediate. However, the ensuing cyclization reactions in CHS and STS are different, giving rise to chalcone and stilbene, respectively as shown in Fig. 7 and 10.[48] As the homodimeric proteins, CHS and STS share highly conserved amino acid residues in their active sites such as Cys, His, Asn and the GFGPG loop in the cyclization pocket.[35] Their sequences share over 60% identical positions with small differences only revealed by detailed inspection. These differences are distributed throughout the sequences instead of being confined to certain regions of the proteins.[49]

While CHS is ubiquitous in plants, STS is only found in a limited number of unrelated plants. According to Schroder's phylogenetic analysis using 34 CHS and 4 STS amino acid sequences, STSs are always grouped with CHSs from the same or related

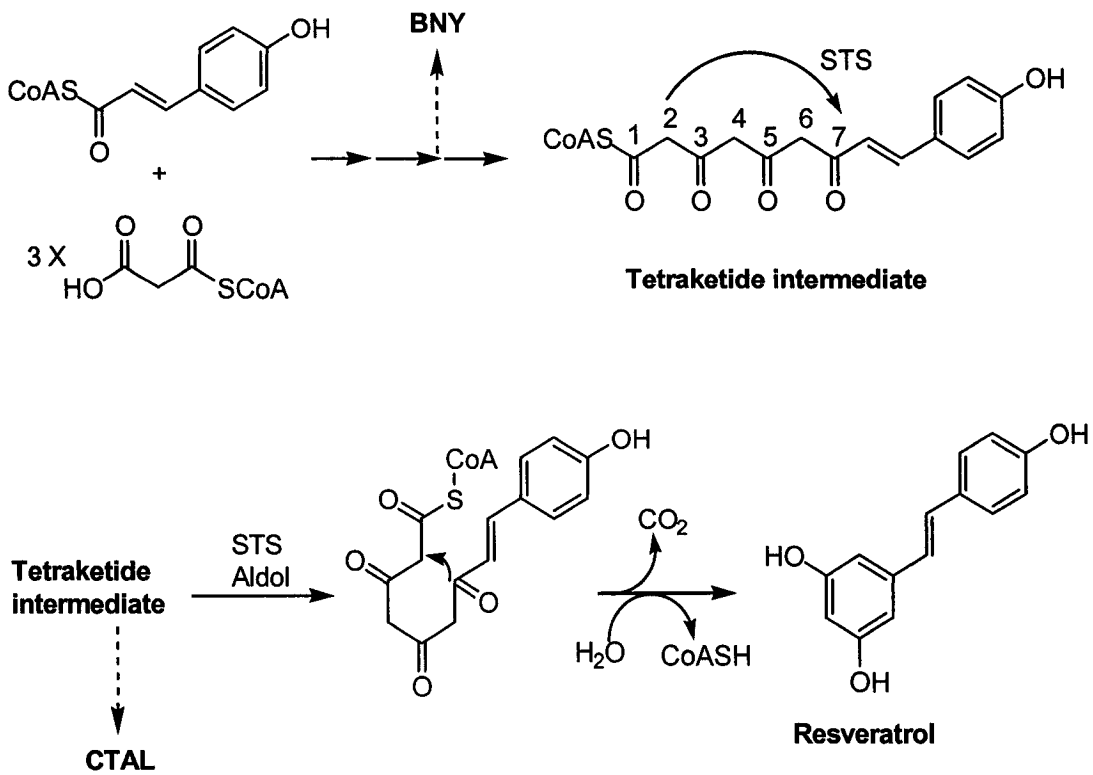


Figure 10. The reaction scheme of the stilbene synthase-catalyzed reaction.

plants, without forming a separate cluster. These results suggested that STSs have evolved from CHSs several times independently and by a limited number of amino acid replacements.[12, 49] An active interfamily CHS/STS hybrid generated by site-directed mutagenesis[49] favours Schroder's hypothesis that a relatively small number of amino acid exchanges are sufficient to alter CHS sequences to a STS function. It has been inferred from Schroder's study that other members of the CHS superfamily followed a similar evolutionary process as STS. However, due to the limited number of CHS and other CHS-like superfamily enzymes, it has not been possible to draw conclusions on which enzyme might have arisen first or whether there were ancestor STS genes. In this study, an extensive phylogenetic analysis was carried out with all the enzymes in the database that belong to the CHS superfamily, to investigate the evolutionary relationship among CHS and other members of the CHS superfamily in more detail.

1. 5. Evolutionary relationship of the chalcone synthase superfamily and other condensing enzymes

The mechanism of carbon-carbon bond formation in the biosynthesis of fatty acids and polyketides is either decarboxylative or non-decarboxylative Claisen condensation.[50] The enzymes that catalyze such reactions are collectively referred to as condensing enzymes, and share overall similarity in protein structure and function.[50] Comparing the CHS superfamily with other condensing enzymes should lead to a more comprehensive understanding of the evolutionary relationship among the architecturally similar domains of these related condensing enzyme families.

β-Ketoacyl-acyl carrier protein synthase (KAS) domain of FAS catalyzes the addition of a two-carbon acetate unit to the enzyme bound thioester end of the fatty acid and has been hypothesized to be the ancestor enzyme of the chalcone synthase superfamily based on their similar mechanism of repetitive C-2 units condensation. There are three different types of homodimeric KAS enzymes, varying in specificity for both substrate and product length.[51] KAS I (protein name: FabB) catalyzes the elongation reaction of C_4 to C_{16} fatty acids, whereas KAS II (FabF) further extends the chain length from C_{16} to C_{18} by adding a single acetate unit as illustrated in Fig. 11. Meanwhile, KAS III (FabH) serves as the initiating enzyme to start the biosynthetic process of fatty acid production by catalyzing the first elongation step from C_2 to C_4 as shown in Fig. 3.

Malonyl-ACP is the extender unit for all three KAS enzymes. Only KAS III accepts acetyl-CoA as the starter substrate and releases its product as an ACP thioester, while the other two can only accept acyl-ACP as the start unit. KAS III is generally believed to have arisen first and acted as the ancestor for other KAS enzymes based on its role as the initiator of fatty acid synthesis.[52, 53] Similar to the KAS enzymes in primary metabolism, the β-ketosynthase (KS) domain of type I and II PKS is responsible for the condensation of an extender unit onto the growing polyketide chain. The KS domain shares more overall similarities with the type I and II KAS instead of KAS III, including its key amino acid residues and reaction mechanism.[2]

Thiolases are classified into two functional types: 3-ketoacyl-CoA thiolase (thiolase I), and acetoacetyl-CoA synthase (biosynthetic thiolase or thiolase II).[54] Thioalse I, the degradative thiolase involved in fatty acid β-oxidation, catalyzes the removal of an acetyl group from an acyl-CoA[55], while thiolase II catalyzes the

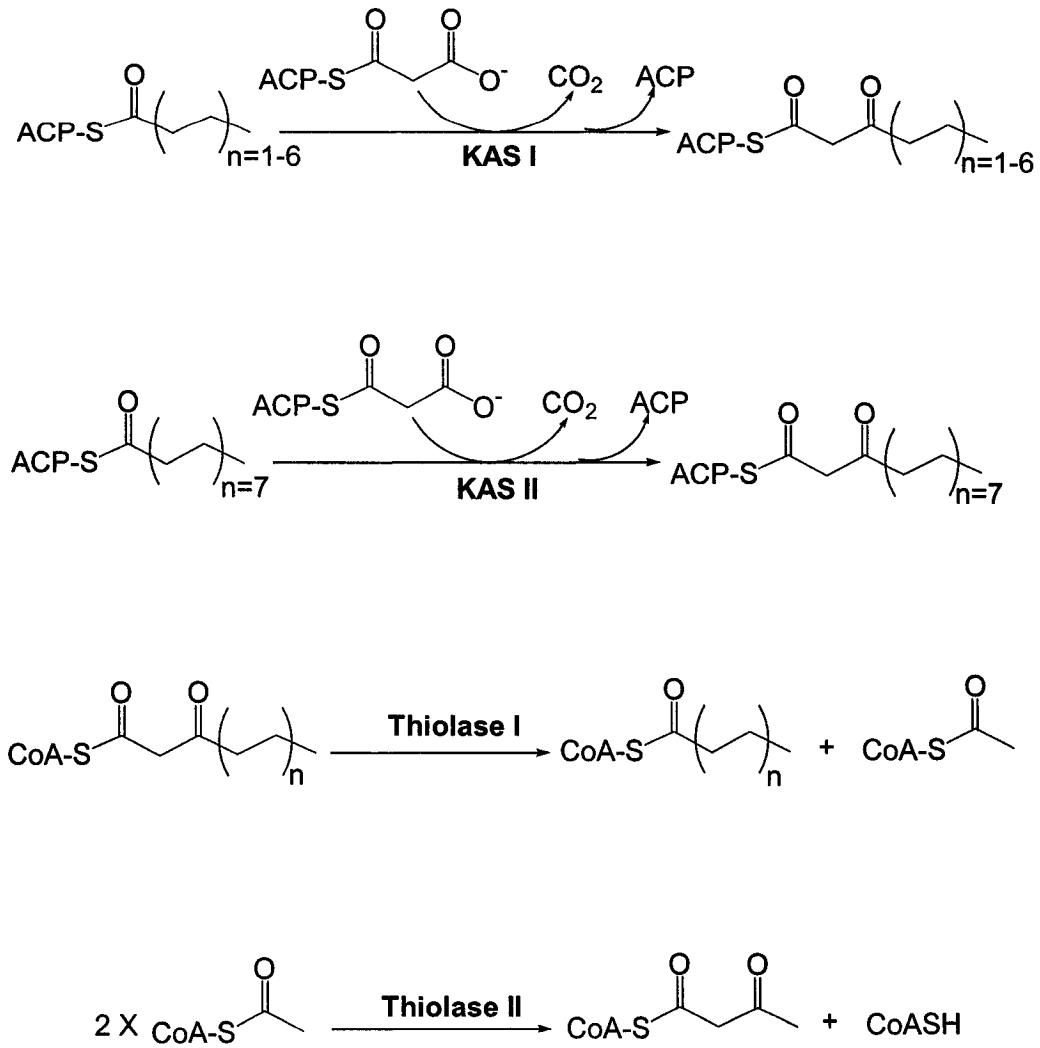


Figure 11. The reaction schemes for KAS I and II, and thiolase I and II.

non-decarboxylative Claisen condensation of two acetyl-CoA to form acetoacetyl-CoA as shown in Fig. 11.[56] Thiolase II differs from KAS enzymes, and KS domains in FAS and PKS most remarkably in their way of generating the acetyl carbanion moiety. Condensing enzymes in FAS and PKS generate the acetyl carbanion by decarboxylation of malonyl-ACP or malonyl-CoA, whereas thiolase II generates the acetyl carbanion *via* an activated cysteine, in an anionic thiolate form, to remove a proton directly from an acetyl-CoA.

All condensing enzymes catalyzing either decarboxylative or non-decarboxylative condensations are believed to have a common evolutionary root based on their conserved structural features.[16] Phylogenetic analysis of these condensing enzymes may reveal the evolutionary history and relationship among all condensing enzymes and provide further insights into the early evolution of metabolism.

The non-decarboxylative condensing enzymes (thiolase II) possess a conserved Cys-His-Cys (C-H-C) architecture.[57] The thiol of the first Cys is responsible for the substrate loading process to form the enzyme-acetyl-CoA complex.[58, 59] The His residue is involved in the activation of the first Cys as well as forming a hydrogen bond with the thioester carbonyl to stabilize the tetrahedral transition intermediate during the condensation reaction.[57] The second Cys serves as a general acid in the transfer reaction of the first acetate group from CoA to the first Cys, and functions as a base in the deprotonation of the second acetyl unit to form the carbanion. On the other hand, there are two variants within the decarboxylative condensing enzyme families: the Cys-His-Asn (C-H-N) catalytic triad found in KAS III[14] and the CHS superfamily[15], or the Cys-His-His (C-H-H) triad found in KAS I and II[60-62], and the KS domains in type I & II

PKSs as depicted in Fig. 12.[63] The decarboxylative condensing enzymes with either C-H-N or C-H-H at the active site are likely to share similar reaction mechanisms. In both cases, the highly conserved active site Cys is activated and functions in the same way as that in thiolase. In the CHS superfamily, the second His also plays a supportive role during the loading reaction of the starter unit by forming an thiolate-immidazolium ion pair.[27, 35] In this study, a phylogenetic analysis is based on the amino acid sequences of all types of condensing enzymes from bacteria, fungi and plants to explore the evolutionary relationship of these condensing enzymes, and to demonstrate the different selections of the active site architectures (Cys-His-Cys, Cys-His-Asn or Cys-His-His) by different types of condensing enzymes during the evolutionary process.

1. 6. Cloning of chalcone synthase from the moss, *Physcomitrella patens*

The earliest land plants, bryophytes, are believed to have appeared more than 450 million years ago.[64] Bryophytes are categorized into three classes: mosses (Bryophta or Musci), liverworts (Hepatophyta), and hornworts (Anthoceraphyta).[65] Mosses and their allies are simple green land plants with leaves, a stem without roots and protonemata. Mosses reproduce themselves through alternating generations. The first generation, the *gametophyte*, forms the green leafy structure. The sperm and gamete produced by the *gametophyte* will grow into the next generation, the *sporophyte* or spore-bearing structure. The capsule released from spores will eventually grow into a new generation of gametophytes.[66, 67] Evolutionary studies support the monophyletic origin of land plants, that is, all the extant plant species belong to a common evolutionary lineage.[68,69]

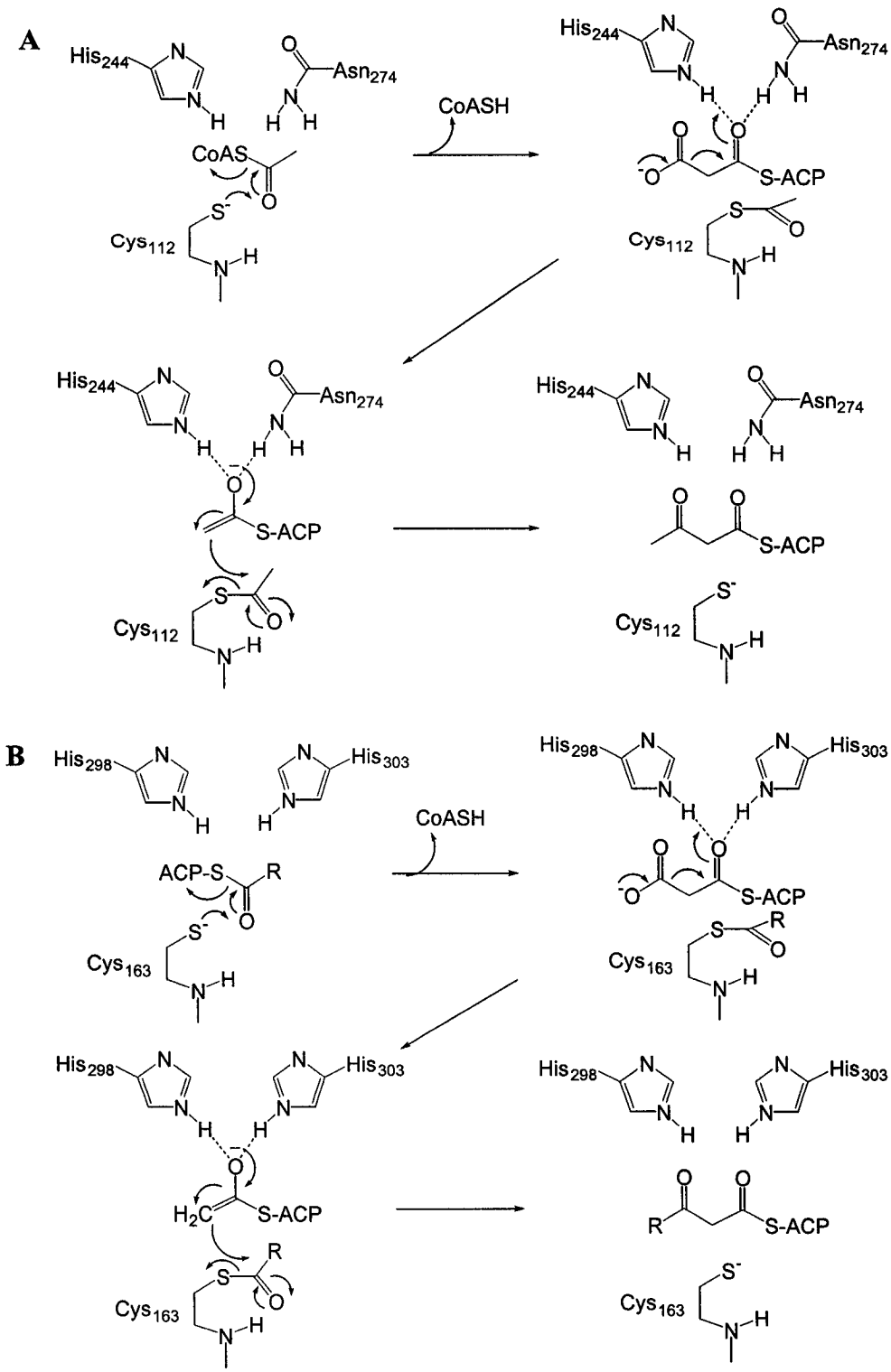


Figure 12. The schematic presentation of the proposed mechanisms of decarboxylative condensing enzymes (**A**: KAS III; **B**: KAS I and II; **C**: thiolase II) having different key active site architectures.

C

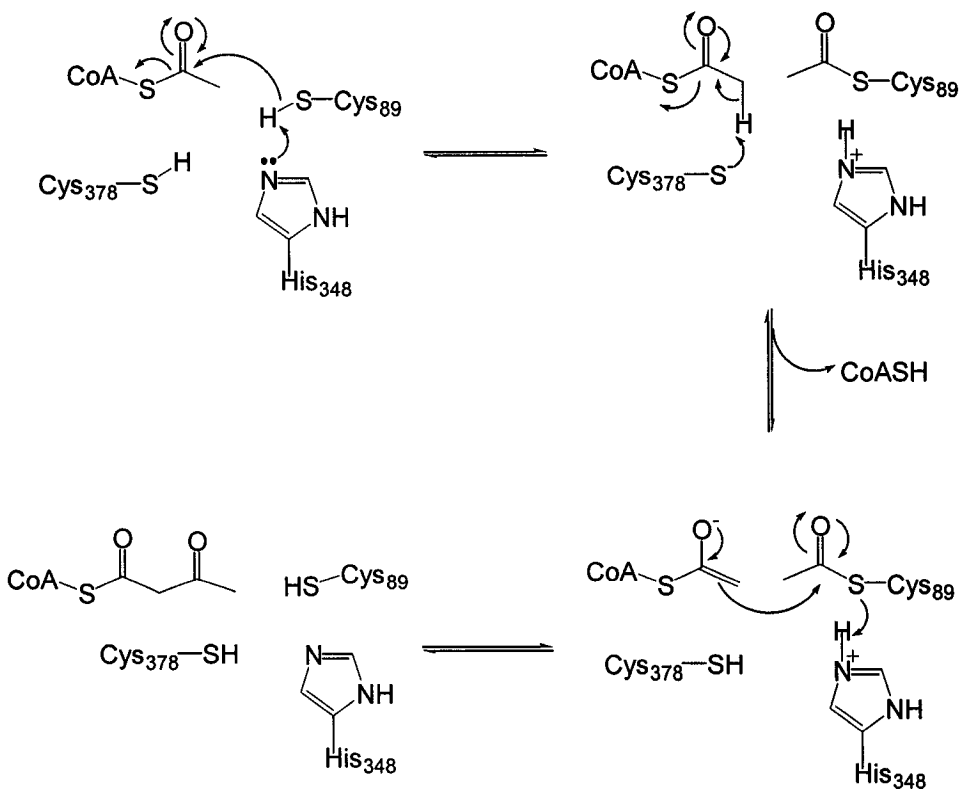


Figure 12. (Continued)

Mosses are widely used as a model molecular system for land plants due to their relatively simple developmental pattern, suitability for cell lineage analysis, common physiological responses to plant growth factors and environmental stimulations, and the facile genetic approaches resulting from the dominance of the gametophyte in their life cycle. Among various mosses, *Physcomitrella patens* (*P. patens*) is frequently selected by plant physiologists as a model system.[70] *P. patens* requires relatively simple growth conditions for the completion of its short life cycle (2 to 3 months) and possesses a high regeneration capacity.[71, 72] As the first successfully transformed moss, *P. patens* is the first multi-cellular eukaryote in which there is gene targeting at an exceptionally high efficiency, similar to that in the yeast. Furthermore, the known set of 25,000 *P. patens* genes covers 50% of the gene content of higher plants and there are a large number of EST sequences currently available.[73, 74]

Although flavonoids are common in higher plants and the primitive vascular plants, they are found in a limited proportion in the bryophytes (50% of mosses, 40% of liverworts and no hornworts). As discussed above, flavonoids provide these plants with protection from UV radiation and are believed to have played a significant role in the migration of early plants from the sea to the land during evolution.[75] Earlier phytochemical studies have found various flavones (apigenin, saponarine and vitexin), biflavones (dicranolomin and bartramiaflavone) and triflavones (triluteolins) in different moss species.[76] A few hundreds of CHS superfamily enzymes have been cloned from various plants, but only a few from the lower primitive plants. It is important to study the CHS superfamily enzymes from primitive plants to provide further information on the evolutionary process of CHS and other CHS superfamily enzymes. Currently, five

diverse CHS superfamily proteins from *Psilotum nudum* (whisk fern), a CHS from the green sprouts of *Equisetum arvense*[77], two CHS-like genes from the liverwort, *Marchantia paleacea var. diptera* and *Lunularia cruciata*[78, 79], and a putative stilbenecarboxylate synthase from *Marchantia polymorpha* (unpublished, Schroder et al.) have been cloned, but no CHS superfamily enzymes had been studied from moss. In this study, a CHS was, for the first time, cloned and characterized from the moss, *P. patens*. In Dr. Suh's laboratory, the sequence of the coding region and 5′- and 3′-untranslated regions (UTR) of a CHS from *P. patens* had previously been obtained from the three-week old gametophore tissues.

2. Materials and Experimental Procedures

2. 1. Plant material

Gametophore tissues of *Physcomitrella patens* (Hedw.) pabB4 strain were kindly provided by Dr. N. Ashton (University of Regina). Three-week-old tissues, grown on agar plates under continuous light at 16 °C as described previously[80], were collected free of agar, chilled in liquid nitrogen, and stored at –80 °C.

2. 2. Cloning of the *P. patens* CHS gene

The open reading frame (ORF) of the *P. patens* CHS was PCR (polymerase chain reaction)-amplified with the 5′-RACE (Rapid Amplification of cDNA Ends) cDNA mixture as a template. The 5′-flanking primer was 5′-TTGTCCCATGGCTTCTGCT-GGGGATG (*Nco* I site underlined), and the 3′-flanking primer was 5′-TGGAATTCTAAGCGGAGTTGGGAGCGGCGC (*EcoR* I site underlined). The PCR program was: 3 min of denaturation at 94 °C, 35 cycles of 30 s at 94 °C, 45 s at 64 °C, 45 s at 72 °C, and 7 min of final extension at 72 °C. Thus amplified PCR product was gel-purified, digested with restriction enzymes, and subcloned into the *Nco* I/*EcoR* I-treated pET-32a(+) (Novagen). The resulting plasmid was transformed into the *E. coli* AD494(DE3)pLysS cells (Novagen) using the heat-shock method.

The nucleotide and amino acid sequences of *P. patens* CHS obtained in this study have been deposited in the EMBL/GeneBank/DDBJ databank (Accession no. DQ255627).

2. 3. Heterologous expression and purification

The cloned *P. patens* CHS, as well as *Pueraria lobata* CHS[81], were expressed as a thioredoxin-$(His)_6$-fusion protein for improved solubility and easy purification. *E. coli* AD494 cells, transformed with the expression plasmids, were grown at 37 °C in Luria-Bertani (LB) medium containing chloramphenicol (34 μg/mL) and ampicillin (100 μg/mL). The overexpression was induced with 1 mM isopropyl-*β*-D-thiogalacto-pyranoside (IPTG). Following an induction period (20 h, 20 °C), the expressed cells were harvested after centrifugation at 5,000 *g* for 15 min. The cells were washed by 1/10 culture volume of 50 mM Tris-HCl buffer (pH 8.0) and suspended in 1/10 culture volume of the starter buffer (20 mM potassium phosphate (KPi), pH 7.2, 300 mM NaCl, 0.1% Triton X-100, 1 mM dithiothreitol (DTT), 5 mM imidazole, 5 mM *β*-mercaptoethanol) at room temperature. Taking advantage of the $(His)_6$-tag, the fusion proteins were highly purified using Ni^{2+}-affinity chromatography.[82] After brief sonication (10 s × 5) on ice and centrifugation (15,000 *g*, 4 °C, 15 min), the soluble fraction was recovered and applied to a Ni^{2+}-affinity chromatography column (His-Bind Resin, Novagen), pre-equilibrated with the starter buffer. Following sequential wash steps with starter buffer (5×bed volume (V_b)) and 80 mM imidazole in the starter buffer ($6\times V_b$), the target protein was eluted with 300 mM imidazole in the starter buffer ($3\times V_b$). The column effluent was then buffer-exchanged to the enzyme storage buffer (100 mM KPi, pH 7.2, 0.1% Triton X-100, 1 mM DTT) using ultrafiltration on a PD-10 column (Pharmacia), for subsequent functional assay. Protein concentration was determined by the Bradford assay (Bio-Rad) [83] with BSA as standard, and the purity was evaluated by sodium dedocyl sulfate (SDS) polyacylamide gel electrophoresis (PAGE).[84]

2. 4. *In vitro* functional assay and steady-state kinetic analysis

To analyze the *in vitro* activity of the *P. patens* CHS, the enzyme assay was carried out with appropriate amounts of protein (2-5 μg of purified enzyme), 0.1 mM *p*-coumaroyl-CoA, and 17 μM [2-^{14}C] malonyl-CoA (1.74 GBq/mmol, NEN) in 0.1 M KPi buffer (pH 7.2) containing 0.1 % Triton X-100 (enzyme assay buffer). Following incubation at 37 °C for 60 min, the reaction was terminated with 7.5 μL of 1 N HCl and the reaction products were extracted with 200 μL ethyl acetate.[82] After a brief centrifugation, a portion (50 μL) of the extract was analyzed by reversed phase thin layer chromatography (RP-TLC) on an RP18 TLC plate (Merck 1.15389) with a solvent system of methanol:H_2O:acetic acid (60:40:1, volume ratio). The radioactive products were quantified and analyzed using an imaging plate analyzer (Molecular Dynamics Storm 860, Amersham Pharmacia) with a compound having a known specific activity as a standard. Authentic naringenin (R_f = 0.3) was used as internal standard to identify the enzyme reaction products. Derailment products, BNY and CTAL were identified by their published R_f values (BNY: 0.4 and CTAL: 0.6).[82] The specific enzyme activity was expressed in pmol of the product produced $s^{-1}mg^{-1}$ (pkat/mg).

All kinetic experiments were performed in the enzyme assay buffer at pH 7.2. The $K_{m(app)}$ values for *p*-coumaroyl-CoA were determined at the malonyl-CoA concentration of 17 μM; and those for malonyl-CoA in the condensing reaction were determined at 150 μM *p*-coumaroyl-CoA. The reactions were carried out with purified enzyme (4 μg) at 37 °C for 30 min, and terminated by 0.1 N HCl (7.5 μL). The reaction products were extracted with 200 μL ethyl acetate, and then analyzed by RP-TLC. Five different

substrate concentrations were used that cover the range of 0.2 ~ 4 $K_{m(app)}$. Data were plotted as v_0 versus [S] and fitted to the Michaelis-Menten equation, the rate equation for a one-substrate enzyme-catalyzed reaction ($v_0=V_{max}[S] / (K_m+[S])$). The V_{max} and $K_{m(app)}$ values were calculated using a non-linear regression program (Enzyme Kinetics Pro, ChemSW).

2. 5. Preparation of starter CoA esters

p-Coumaroyl-CoA and other starter CoA esters were synthesized from their respective *N*-hydroxysuccinimide esters according to the active ester exchange method established by Stöckigt and Zenk.[85] Briefly, *p*-coumaric acid or other carboxylic acids (2.5 g) and *N*-hydroxy succinimide (1.8 g) were dissolved in tetrahydrofuran (20 mL). *N,N*-Dicyclohexyl carbodiimide (3.5 g) was added to the reaction mixture, which was stirred overnight in the dark. The reaction was analyzed by silica TLC with dichloromethane:methanol (50:1) as the solvent system. The reaction mixture was filtered and evaporated to remove the by-product, dicyclohexylurea and remaining solvents. The resulting active esters were dissolved in ethyl acetate (20 mL), washed with 1 M $NaHCO_3$ solution, and further purified by silica gel chromatography with dichloromethane as solvent.

To exchange the ester, the corresponding active ester (65 mg) and $NaHCO_3$ (45 mg) were dissolved in acetone (4 mL), and the solution was dropped into a degased H_2O solution (10 mL) containing coenzyme A (40 mg). The resultant solution was stirred and incubated overnight at 4 ℃, and then desalted with a Dowex 50W-X8 resin column. The CoA ester was then eluted with H_2O. After extraction with ethyl acetate to remove the

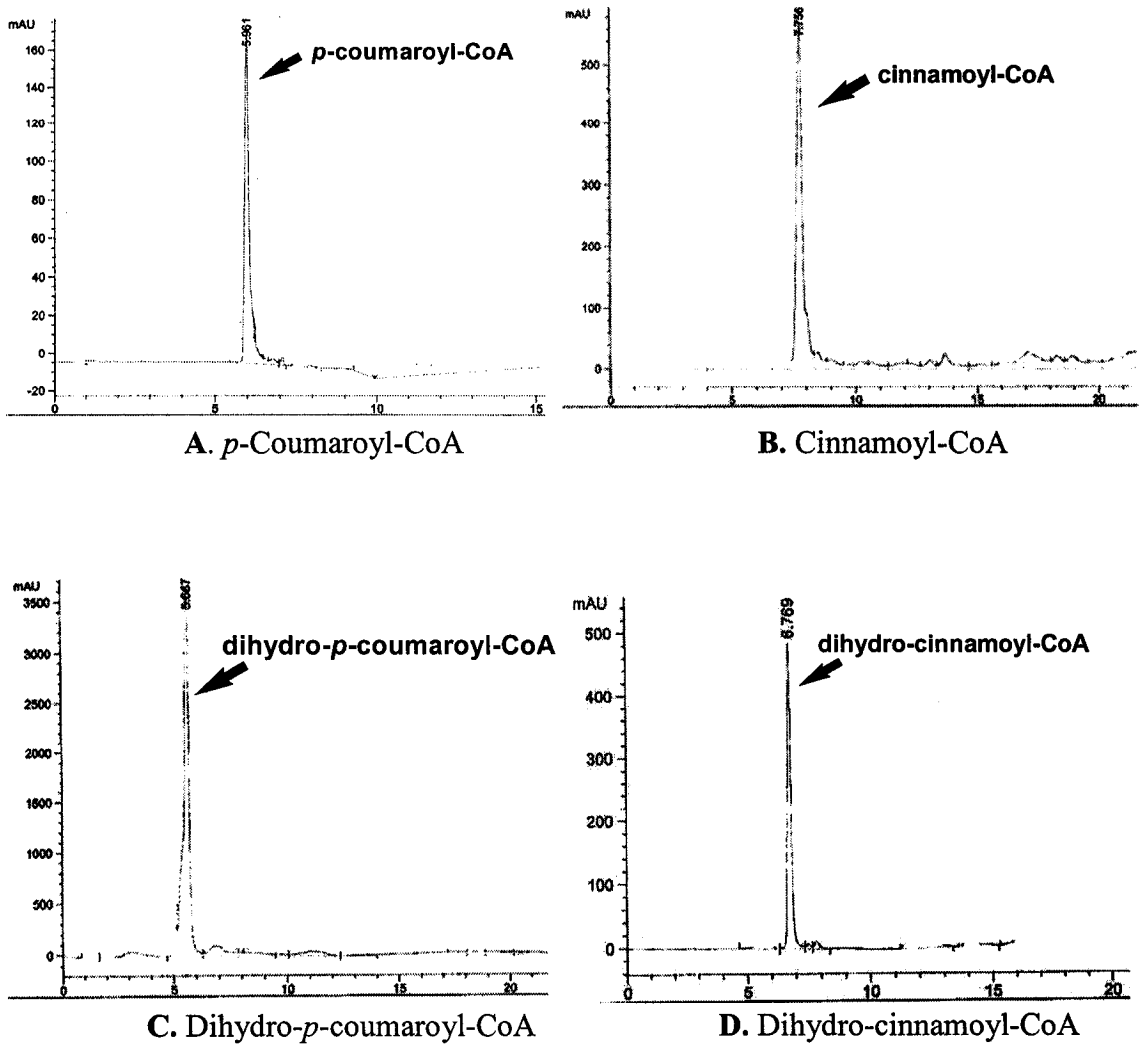
excess active ester, crude CoA ester was isolated by evaporating all the remaining solvent.

Starter CoA esters were further purified using reverse phase-high pressure liquid chromatography (RP-HPLC) (Apollo™ 5C 18-MS; 4.6 × 250 mm; 0.6 mL/min; UV 254 nm) as showed in Fig. 13. A gradient solvent system, consisting 0.1% trifluoroacetic acid (TFA) in H_2O (solvent A) and acetonitrile (solvent B), was used for the separation: 0 ~ 1 min, 20% B; 1 ~ 31 min, 20 ~ 40 % B. Under these conditions, the retention times of *p*-coumaroyl-CoA, dihydro-*p*-coumaroyl-CoA, cinnamoyl-CoA, and dihydro-cinnamoyl-CoA were 6.0 5.7, 7.8, and 6.8 min, respectively, as shown in Fig. 13.

2. 6. Phylogenetic analysis

The amino acid sequences for a total of 13 KAS III, 19 KAS I and II and KS domains of PKS, 45 representative CHS superfamily enzymes in plants, and 6 CHS-like enzymes in microorganisms were subjected to a phylogenetic analysis, in which the type III KAS (FabH protein) from several archaea served as outgroups. Amino acid sequence alignments were created using the CLUSTALW[86] (European Bioinformatics Institute, http://www.ebi.ac.uk) and CLUSTALX programs.[87]

Phylogenetic relationships for the amino acid alignments were reconstructed and evaluated with the Bayesian inference method using the MrBayes program (v. 3.1.2).[88] The MrBayes program incorporates various evolutionary models to analyze the amino acid data, including the Poisson and Blosum62 model, where both the stationary state frequencies and the substitution rates are fixed. In this study the "mixed" amino acid analysis model was used, which allows model jumping between various fixed-rate amino



A. *p*-Coumaroyl-CoA **B.** Cinnamoyl-CoA

C. Dihydro-*p*-coumaroyl-CoA **D.** Dihydro-cinnamoyl-CoA

Figure 13. HPLC chromatograms of purified starter CoA esters. (0 ~ 1 min, 20% acetonitrile in H_2O (0.1 % TFA), 1 ~ 31 min, 20 ~ 40 % acetonitrile in H_2O (0.1 % TFA); flow rate: 0.6 ml/min)

acid models during the analysis process. The Markov chain Monte Carlo analysis (MCMC) was performed for 1 million generations with four independent chains, and the Markov chain was sampled every 100 generations. The state of convergence was evaluated and identified by two diagnostic parameters output by the MrBayes. The standard deviation of split frequency was close to 0.006, and the potential scale reduction factors approached 1.00 for all the run parameters. All trees that were sampled before reaching the convergence state (350,000 generations) were discarded (burn-in = 3,500), while the remaining trees were accumulated to construct a consensus tree (conexus type data) and to calculate the posterior clade probabilities and bootstrap values. The TreeView program[89] was used to translate the consensus tree data and construct the phylogenetic tree.

In addition to the Bayesian estimation method, the data was analyzed with neighbour-joining (NJ) and maximum parsimony (MP) methods using the CLUSTALW and MEGA (v. 3.1) software packages.[87, 90] For the NJ analysis, the models were Blosum62 and Pam, and the bootstrap analysis was conducted with 111 seeds and 1000 pseudo-replicate sequences. The MP analysis was performed using the step-wise calculated heuristic search option, with gaps treated as missing data, and branch swapping by tree-bisection-reconnection (TBR) to exclude the constant and uninformative characters from the analysis. The consensus trees produced by the NJ and MP methods were highly similar to the one constructed by Bayesian estimation, leading to the same phylogenetic relationships.

3. Results and Discussion

3. 1. cDNA isolation and sequence analysis

The *P. patens* CHS cDNA contained an ORF of 1194 bp encoding an approximate 42.6 kDa protein of 397 amino acids. The nucleotide sequence surrounding the designated start codon (TCCAATGGC) is similar to the initiation consensus sequence of AACAATGGC in plants.[91] The presence of an in-frame stop codon (TAG) 30 bp upstream of the start codon and the absence of any other start codon in the 5′-untranslated region (278 bp) lend further support to the identity of the designated ATG as the start codon. The ATG start codon of *P. patens* CHS is followed by GCTTCT, thus making the second and third residues Ala and Ser, which are the predominant residues at corresponding positions in highly expressed plant genes.[92]

Currently there are a total of 40 putative CHS expressed sequence tag (EST) sequences with significant clarity and length in *P. patens* EST databases; 8 in GenBank, 16 in the Physcomitrella EST Programme (www.cosmoss.org)[74], and another 16 in the PHYSCO base (moss.nibb.ac.jp).[73] Multiple EST sequences were apparently generated from an identical transcript since there are only 11 separate ESTs or contigs as shown in Table 1 and Fig. 14. In these EST sequences, there are seven different coding sequences (A~G) and five different 5′-UTR sequences (5a~5e) as depicted in Fig. 15 and Fig. 16, respectively, each of which is associated with a different coding sequence. The type A coding sequence was most frequently encountered in the databases. It was identified in the *P. patens* CHS sequence obtained in this study as well as in 38% (15/40) of all the ESTs.

Table 1. Analysis of putative chalcone synthase EST sequences from *P. patens*

Accession no.	Length (nt)	Sequence identity to PpCHS (%)[a]		Sequence type			Comments
		Nucleotide	Amino acid	5′-UTR	cds	3′-UTR	
DQ275627 **(This study)**	1629	(100)	(100)	5a	A	3a	Contains Contig4041, Contig4179, Contig7312 PPP_7231_C2, BJ177589, BJ583475.
Contig9794	2276	98	100	5a	A	3b	Contains PPP_182_C2. B is similar to A and contains 30 nt insertions.
Contig7184	1762	88	94	5b	B	3c	Contains PPP_7231_C1, PPP_1175_C1, PPP_7571_C1.
Contig4180	1519	98	100	-	A	3d	Overlaps with PPP_182_C3. D is 1363 nt long.
Contig4042	1182	98	100	-	A	3e	E contains parts of B and D.
Contig7183	1060	85	95	-	B	3f	Contains PPP_2112_C1.
Contig466	856	80	89	-	C	3g	Contains PPP_142_C1.
PPP_7327_C1	836	81	86	5c	D	-	Contains Contig467. Extra 9 aa at the N-end.
BJ610253	776	87	82	5d	E	-	Identical to rpphn50i18.
Contig10984	749	84	91	-	F	3h	Identical to Contig11524. Contains PPP_2843_C1.
Contig 11523	718	91	90	5e	G	-	Contains BJ157102, Contig 10983.
BJ589895	700	95	-	-	-	3i	Identical to Contig14228. Contains BJ604022. I is similar to D with insertion/deletion.

[a] Sequence identity was calculated with the matching sequences only.

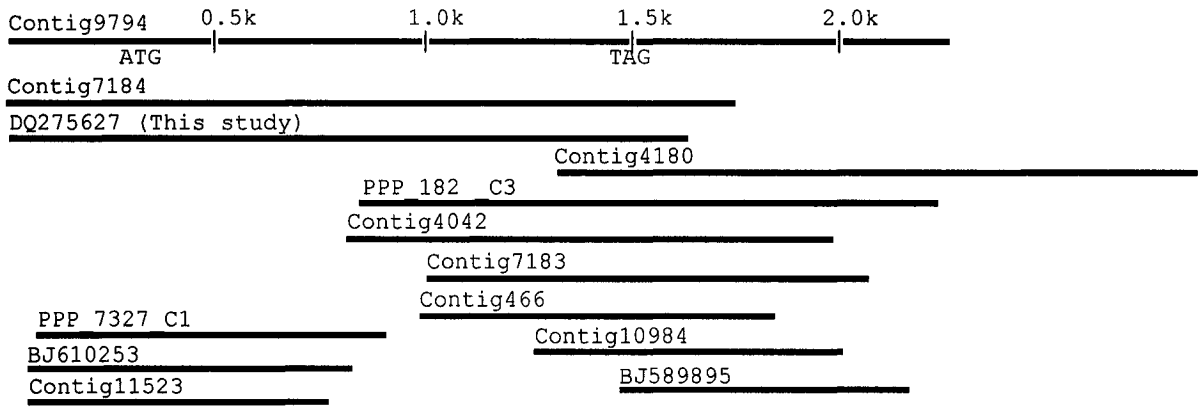


Figure 14. Alignments of the *P. patens* chalcone synthase EST sequences.

```
type A          ----------------MASAGDVTRAALPRAQPRAEGPACVLGIGTAVPPAEFLQSEYP
type B          -----------MAPSGEVDVQGAATRSALPRGQPRAEGPACVLGVGTAVPPAEFLQSEYP
type D          -------MAPPSGESISASAEEPIALSVLPRGQPRAEGPASVLGIGTAVPPTEFLQSEYP
type E          -----------MAPAGEVEAEVRARRAVLPKGQPRAEGPACVSRNGTAVPSTEFLKSDYP
type G          MAPRAGELDIAASDEQVAAAPLVRMHAPIPRGQPRAEGPACVLGIGTAVPPTEFLQSEYP

type A          DFFFNITNCGEKEALKAKFKRICDKSGIRKRHMFLTEEVLKANPGICTYMEPSLNVRHDI
type B          DFFFNITNCGEKDALKAKFKRICDKSGIRKRHMFLTEEVLKANPGICTYMEPSLNVRHDI
type D          DFFFEVTKCSEKEALKAKFKRICDKSGIRKRYLFLTKEVLEANPGIATYMEPSLNVRHDI
type E          DFFFNITNTCDKEALKVKFKRICDKSCIRKRHMFLTEKVLKANPGICTYMQPSLNVRHDI
type G          DFFFNITNTSEKEALKAKFKRICDKSGIRKRHMFLTEEVLKANPGICTYMEPSLNVRHDI

type A          VVVQVPKLAAEAAQKAIKEWGGRKSDITHIVFATTSGVNMPGADHALAKLLGLKPTVKRV
type B          VVVQVPKLAAEAAQRAIKEWGGRKSDITHIVFATTSGVNMPGADHALAKLLGLKPTVKRV
type D          VVVQVPKLAAEAAVKAIKEWGGRKSEITHIVFATTSGVNMPGADHAMAKLLGLKPTVKRV
type E          VVVQVPKLAAEAAQKAIKEWGGRKSNINHIVFATTSGVNML-ADHALVKLLGLKPTVKRV
type G          VVVQVPKLAAEAAQKAIKEWGGRKSDITHIVFATTSGVNMPGA----------------

type A          MMYQTGCFGGASVLRVAKDLAENNKGARVLAVASEVTAVTYRAPSENHLDGLVGSALFGD
type B          MMYQTGCFGGASVLRVAKDLAENNKGARVLAVASEVTAVTYRAPSENHLDGLVGSALFGD
type D          MLYQTGCFGGATVLRVAKDLAENNKNARVLAVCSEV----------------------
type E          MMYQILEL-----IKYQK------------------------------------------

type A          GAGVYVVGSDPKPEVEKPLFEVHWAGETILPESDGAIDGHLTEAGLIFHLMKDVPGLISK
type B          GAGVYVVGSDPKPEVEKALFEVHWAGETILPESDGAIDGHLTEAGLIFHLMKDVPGLISK
type C          --------------EKPLFEVHWAGETILPESDGAIDGHLTEAGLIFHLMKDVPGLISK

type A          NIEKFLNEARKPVGSPAWNEMFWAVHPGGPAILDQVEAKLKLTKDKMQGSRDILSEFGNM
type B          NIEKFLNEAR-CVGSPDWNEMFWAVHPGGPAILDQVEAKLKLTKDKMQGSRDILSEYGNM
type C          NIEKFLSEARKCVGSPDWNDMFWAVHPGGPAILDQVEAKLKLSKDKMQGSRDVLSEFGNM
type F          ------------------------------------AKLKLTKDKMQGSRDILSEYGNM

type A          SSASVLFVLDQIRHRSVKMGASTLGEGSEFGFFIGFGPGLTLEVLVLRAAPNSA
type B          SSASVLFVLDQIRQRSVKMGASTLGEGSEFGFFIGFGPGLTLEVLVLRAAATV-
type C          SSSSVLFVLDQIRQRSMKMGASTTGEGNDFGFFIGFGPGLTLEVLVLRSMPIV-
type F          SSSSVLFVLDQIRQRSVKMGASTLGEGSDFGFFIGFGPGLTLEVLVLRAAANV-
```

Figure 15. Amino acid sequence alignment of the *P. patens* ESTs annotated as putative CHS genes.

A. 5′-UTR sequences

Type 5a

GGGTCGATACGAGCTTGAAAACGGAAGTCGCCACCCAACTACACCGTCTATCATTTCTCTGGCAATCTCTCC
GATCGTTGTACTCAAGTTACTTACGAAAATAAACTAGCTCGTGGCGAACTAAGAGAGTGGGGGTGTCTCCGC
AGATCCCGACTGGTAGCATAACTGAATCGGTGCGTAGAACAACCTCTCTTGTTGCAGGATTAGCAGACAGAC
AAAGTCGGAATACGTTTGGCAGCACAGATTGAGAACTAGAGTGAGGAGATCGACCGGGGTTGTCCAATG

Type 5b

GGAGAGAACTTCTCTGTGCACTCAACGCATTCAGGAAGTGTAGCAGATAGAGCAACAAACTTGTGAGCTAAC
AGTTCAGAGCGAGACGTCTGCTCAACTCGGGCCACCTTCTCGCCGGCACATTGTGCGTTGCATATAAGCCCC
CTGATTTTGGCAGGTTAAGAGATCCTAGGCAGTGGCTGCAGGGCCAGTTCACCGGAACAAATTGAACGCTCA
TTGTCCGGCCGCTTCGAGACAACCCCGAGTGAGCAGCTCGGACTCAACATCCATTCCGAACCCAAACGTGCC
CATG

Type 5c

AGTGCAACTTATATTGCCGCCCCGCAAGCTCACAGTGGAGTATCAGCGCTGGACCTGAAGGTGTCACCATCT
AGAGATCTGTCGAAGCATTTTTATTTCGAGTGTTATTTCCTCAGGGTCTGCACAGGTCTTACTAGCAGCATC
TCTGCACGCCTCGAGAGGTGTGTGTGTGTTTTTCTTTGCTGAGAGGTTTAGTGCCAGCGAGCAATG

Type 5d

TCCGATCGCTGTGTTGAATTCACTTACGAATGTAGAGTTGCTCTTGTTGCAAACAAAGAGAGTCGGGTGCGT
GTGCAGATTCCAACTGGTAACATTTACAGAGTCGGCACGTAAGAGGCATAAACCGGTGTGTTGTCGCAGAAT
TAGGAAGGAGAGAAAGCGGGAGTACAGGTTGCGCTACACAGGCTGAGAACAGTTGGGTGAAGTAATAGAGCA
GGGTTGTGGAATG

Type 5e

ACAGGCAATCCATCCACTCAAGGTTCGCACATCGCAGCGTCGCTGCCGATTCCGACCCCTTTTCCCGATTC
AGCGAGCGTTAGAGTTTTGGTTGTAGTTCTCGACCCGAAAACGTTTGCAAAGCGTGTCGCTTGTTGCTCGC
TCGGGTGCCTATCCATTCCGAAGTTTCTGTGCAGACGACCTAGTTTGGGATTCACGTCGCAGGCTCTAGCA
CATCCGTTGGCAGCAATG

B. 3′-UTR sequences

```
type_3b         ---------------------------------TAGATTCCACCCCACCTCTCCTTTTTAC
type_3e         ---------------------------------TAGATTCCACCCCACCTCTCCTTTTTAC
type_3a         ---------------------------------TAGATTCCACCCCACCTCTCCTTTTTAC
type_3d         ---------------------------------TAGATTCCACCCCACCTCTCCTTTTTAC
type_3i         ---------------------------------TAGATTCCACCCCACNTCTCCTTTTTAC
type_3g         --------------------------------TGAAGTCTCTGTCGATAGATGATTATCAG
type_3c         --------------------------------TGAGCGGCGTCCTAATGGCGTTGTTCGA
type_3h         TAAATGGGGGAAGCAGGGCGCGAGAAACGCATCGGATGAGGGACAGGGTGTGTGCAGGGG
type_3f         -----TGAGTTGTAGCAACGGGCACGGGCAAATCAATATCGTCTCCCTGGTTTTCCTTTT

type_3b         TCCCCATTCTCACCAATTTGTATCATAT--CATATT-ATTGCATAGCATTGCATAGCATA
type_3e         TCCCCATTCTCACCAATTTGTATCATAT--CAAATT-ATTGCATAGCAT-----------
type_3a         TCCCCATTCTCACCAATTTGTATCATAT--CAAATT-ATTGCATATCAT-----------
type_3d         TCCCCATTCTCACCAATTTGTATCATAT--CAAATT-ATTGCATATCAT-----------
type_3i         TCCCCATTCTCACCAATTTGTATCATAT--CAAATT-ATTGCATAGCAT-----------
type_3g         TCATCATTC-----AGTCTCGATCAACC--CAAACTCACCACATTTTTC-----------
type_3c         CCGTGGTGA----CTATTTGGAAGGTCA--TTTGTC-ATTGTACCCTCC-----------
type_3h         CGGTGGCCGGGTGCGGGTGTGGACCACG--CTCCTTCGCCGTGCATGATTCAGATCAGCA
type_3f         GCGCGCTGGAGTGGAGATCGTATGAGTGAGCGAATT-ATGATGATTCAT-------GGGT
```

Figure 16. The 5′-UTR sequences of the *P. patens* ESTs and sequence alignment of the 3′-UTR sequences annotated as putative CHS genes. The start codons and stop codons are underlined.

```
type_3b         GCATAGCATAGCATAGCATAGAAGTGTGCAATTCAATTT-CAATGGTATTGCTTCCTGA-
type_3e         --------------AGCATAGAAGTGTGCAATTCAATTT-CAATAGTAT----TCCTGA-
type_3a         --------------AGCATAGATGTGTCCAATTCAATTT-CAAT-GTAT----TCCTGA-
type_3d         --------------AGCATAGATGTGTCCAATTCAATTT-CAAT-GTAT----TCCTGA-
type_3i         --------------AGCATAGAAGTGTGCAATTCAATTT-CAAT-GTAT----TCCTGA-
type_3g         -----------CTATCCACCGGAATGTACTATAGCGCTC-AAAACGCACATCGCTCAGAT
type_3c         ---------------GAATCGAGGTCTAGAGAGCGCCCC-TTGGTACAT-----------
type_3h         GCGAGCCGTG-CGACCAGCCGCTGGACATGGAGCCGCTCGTCGGTGAGCAGATGCGGAGG
type_3f         CAGAGAGCATCCAGGTTGTGGCACCAGAGGGTTTTACATTCAGGTCTGCAGCAAATGGAC

type_3b         -GCT-AGAGTTCGCGCTTAATTTTAAAAGA--GAAATCAGGT---TATCGAATCAACCTA
type_3e         -ACT-AGAGTTCGCGCTTAATTTTAAAAGC--GAAATCAGGT---TATCGAATCAACCTA
type_3a         -ACT-TGAGTTCGCGCTTAATTTTAAAAGA--GAAATCAGGT---TATCGAATCAAAAAA
type_3d         -ACT-TGAGTTCGCGCTTAATTTTAAAAGA--GAAATCAGGT---TATCGAATCAACCTT
type_3i         -ACT-TGAGGTTGCGCTTAATTTCAAAAGA--GAAATCAGGT---TATCGAATCAACCTA
type_3g         GGAT-CGGAGATGCACTGTCTTCGAAAATCTCAAGATTAGAC---TAATACTGTGAGCCG
type_3c         TATT-CTGATTAGAAGCTACCTTGACCGTTTCGTAGTTGATT---TTTTGCATCAGGGCT
type_3h         CGCTACGGACACGGACGTCCCCCGGGTAGCATCCAAGCGGATGGATGTGAAGCCGAGGCC
type_3f         TTTCGATGAAATTTGGTCAGTTGGATCAGCGCGCTACCGCCTCCGTCCACCAGGGGTTCT

type_3b         -CGTTCTT--GCTTCAGCGAGCTC-TTTACAC---ACTCTGAAATCTAGGGTTCCACTGT
type_3e         ACGTTCTT--GCTTCAGCGAGCTC-TTTACAC---ACGCTAAAATCTAGGGTTCCACTGT
type_3a         AAAA-------------------------------------------------------
type_3d         -CGTTCTT--ACTTCAGCGAGCTCATTTACAC---AGTCTAAAATCTAGGGTTCCATTGT
type_3i         -TGTTCTT--TCTTCAGCAAGCTC-TTTACAC---ACTCTAAAATNTAGGGTTCCATTGA
type_3g         AAAGGTAT--TTAGAAGCAGAGAAAAGT-CAC---CGGGTAAAGTTTCAGA---CAGACT
type_3c         ----TTCT--CTGTCATTTCAGTGAAGTTAAG---CCGCAGATATCC---ATTCCATCCA
type_3h         CAAAGTGTAACATGCAACACCGCAAAGGAAAG---AGTTTGACAGATATTGTTTCGGCGA
type_3f         GCAGCGGCG-ACGGCGACAGAGGAGGTCGCATGGACAGTTGCAAGTTGCTGAGACGTTGG

type_3b         TTCGTATCCGGAACCTGAACGATCTGGCGCCTGTGGGTATATTTCTCTCGCTGTGCTAAA
type_3e         TTCGTATCCAGAACCTGAACGAGCTGGCGCCTGTGGGTATATTTCTCTCGCTGTGCTAAA
type_3d         TTCGTATCCAGAACCTGAACGATCTGGCGCCTGTGGGTATATTTCTCTCGCTGTGCTAAA
type_3i         TTCGTATCCAGAACCTGAACGATCTGGCGTCTGTGGATATATTTCTCTTGCTGTGCTAAA
type_3g         CAAAGATATGAAGACTATACAGGAGAGTCATTTTTCGTATGAACTAGCCATCGAGCATGT
type_3c         TTCAAATCACGCTACTC--CGAAGAAATTATAGATAGTGTAAACTGTATCATAATTCAAT
type_3h         TATGGA-CTAGCTCCTGAAAGCTCTGGTTTTCAAGCTTTTGACGGTTGTTGTTGTCATGT
type_3f         AGCGGGCACAAATGTCCAAGGCAGGGATCATGTTGGAGGGGTGAAAGGACTTGAAGTGTT

type_3b         TTGG------CAGTGCATCTACTCTAGCGTTCTTATTGTTCACTCAATTTATTTCTGGAA
type_3e         TTGG------CAGTGCATCTACCCTAGCGTTCTTATTGTTCACTCAATTTATTTCTGGAA
type_3d         TTAG------CAGTGCATCTAATCTAGCGTTCTTATTGTTCACTCAGTTTATTTCTGGAA
type_3i         TTAG------CAGTGCATCTACTCTAGCATTGTTGTTGTTGACTCAATTCATTTCTGGAA
type_3g         TT--------CAGTGCAAACATTATGGAATGGTCTATGCTGACTTA----CTTTCTAATA
type_3c         A----------------------------------------------------------
type_3h         TCGAAGT---CGTTGCAGTGGCCGTGATTTTGCTGGAGCAGGAGCAAAATTTCGTTGCAC
type_3f         CTGAGTTGCCTTTCGAAGGAGGGCAGGAATTATTTACGTAATCTGCAGCCATGCTTAGTC

type_3b         GACAGGAGTTTGGCAATGTATAGTCACAATGCAATCAGAGGTGCGAAGTACAGTCTTATT
type_3e         GACAGGAGTTTGGCAATGTATGGTCACAATGCAATCAGAGGTGCGAAGTACAGTCTTATT
type_3d         GACAGGAGTTTGGCAATGTATAGTCACAATGCAATCAGAGGTGCGAAGTACAGTTTTATT
type_3i         GACAAGAGTTTGGCAATGTACAGTCACAATGCAATCAGAGGTGTGAAGTACAGTTTTATT
type_3g         AAATATTTATTGCCATTTCGCAATGTGA--GAAATCAAAAGTTAAAAA------------
type_3h         AAAGGATTCGGTGTAGTGTATATTTAGGTGGAGCTTTAGATTAAACTGGGCACAGACAAG
type_3f         ACATGGTTTGAGAAATTGTAGGAACTGT--GCGATTCTTGTTGATATGAGGCACAATTTA
```

Figure 16. (Continued)

```
type_3b         TAGAAAATTAGGGCTGGAAGAAGATGAAGTGGGAGTGTGTAGTGCGCATTGTGCCGATGG
type_3e         TAGAAAATTTAGGTTGGAAGAAGATGAAGTGGGAGTGTCTAGTGCGCATTGTGCCGATGG
type_3d         TTGAAAATTTAGGTTGGAAGAAGGTGAAGTGGGAGTGTCTAGTGCGTACTGTGTGCTAAG
type_3i         TAGAAAATTTAGGTTGGAAGAAGATGAAGTGGGAGTGTCCAGTGCATACTGTGTANTAAG
type_3h         TGCTCTGCTGGCAAGGCTGCATGAGCATAACATATCTTGTGTCCAATGTATTATTCTGCG
type_3f         TGAGCATTTTGCAAGTTGAATGTGTGGCGACATTTTGCCCCACCCGAAGTGTGGTTTTAG

type_3b         ATCTGAAATTCATCAGGAATGTAATTCATCGGTGAATCTCCGGACCATGTGTTCGTTTGA
type_3e         ATCTGAAATTC---AGGAATGTAATTCATCGGTGAATCTCCGGACCATGTGTTCGTTTGA
type_3d         TTAGGTCTAGCCTTAAGGATTCAAAGTGTGGTATGGGTACAGATTTTGGCAAACATGGGA
type_3i         TTAGGTCTAGCCTTAAGGATTCAAAGTGTGGTATGGGTACAGATTTTGGCAAACATGAGA
type_3h         ATG-------------------------------------------------------
type_3f         TTAAAAAGCATTTGATGCAGAGATCACTTTATTTGGTTCCTCAAAGAACAAAAATAAACA

type_3b         GC-TTTTATAATTGCAGCCTCTAAGTCAGCTGCGATTGTTCGACACTAC--CCACTGTTT
type_3e         GC-TTT----------------------------------------------------
type_3d         GTGCTTCGTGA-----------------------------------------------
type_3i         GTGCTTCGTGACAAAGTGTGGTATGGGTACAGATTTTGGCAAACATGAGAGTGCTTCGTG
type_3f         AATTGGTAACTTG---------------------------------------------

type_3b         CTTTGAGCTGTTTCTCGTCACTTTTGAATCTGTTGCTGTAGCATTGGAAACTGAGAAGAG
type_3d         -TGTAGAATAAGTTTGTACAATAATACGTTTCCTGTCGCTCCAGAAGCCATCCCTGTTTT
type_3i         ATGTAGAATAAGTGTGTACAATAATACATGCTTCGTCGCTCCAGAAGCGATCCTTGTTTT

type_3b         AAGGGTCTGCAAGGAATTGGTTGGTTGCATTGAACTCATGGTAGCATGGCCCAGCCAACA
type_3d         CCCTCAGAGCGAGCAAATGTTGTTTTCTGTAGCACTCATAGGTGCGTCGCATCGCAAGGA
type_3i         CCCTTAGAGCCAGCAAATGTTGTTTTCTGTAGCACTCATAGGTGCGTAGCAT-GGAAGGA

type_3b         ATTAATGAAAACTGTGATTTGTTGTCTGTCGGTCATTGAGAAAGTTAATCCTCCACCAGT
type_3d         AATACAACCTTGAATTACATGGTTTATTTGAGAAGCAATCCTACTTGGAATTCGATTCTA
type_3i         AATACAA---------------------------------------------------

type_3b         TTCAGTGATATCTGGACACAAA------------------------------------
type_3d         CGAATTTGAAATATGTGTCTTGAGGTTAGAGTATCTGCGACTTCAACCATTTTGGAGTAA

type_3d         AATAAATTGTTTGATTTCTGATATTTTGAAATAGTGCTGTGTCGTAGTTTTGACAAATCA
type_3d         TCCATTGTCTAAGGTTTTCTCCACAACCGAGGCTGAGACCATTCATAGTGAAGAATTTGA
type_3d         AGGTGCAGGATTTGCGGGGCGACTGGTCTCAACCAATCGGGAGTGTTGTATTTAATTTTT
type_3d         TTAGAGAAACTGTTCATTACTTGGAAGCCAAGACACTTCTTCTCCCAACTGCGACGATAA
type_3d         TACACGTGACTTCATGGGTTCTTCTGGAGGATCGACAGAGCTGAAAAGCTCGTCAGCGTC
type_3d         ACATTCGAACTTCCTAGCGGAGTTGATGCCGAAAAAGGAGATTGGGGTCTCGCTGCAAAC
type_3d         ATCCGCAACTTGGTTCTGTTAAGTGGCAAACATCTGCAACTTGTGCTATAAGTAAAGTGA
type_3d         TCTTTTAATTGCTCTAGTTGACATTCCTAATTTAAAAACGAGGATTGTAGTTATTTGTAA
type_3d         TCAATCCCTCTCCCCTCCCCCCCTCAATGCTCGCATAACTAGAAGCTTTAATGACATTTC
type_3d         TTAGTACTATTTTCATAGAAGAGGTGGAAGAAATATCGAAGTAATTTGATCATCTTCAAATA
```

Figure 16. (Continued)

The amino acid sequence identities among these seven coding regions were 82~95%, while the nucleotide sequence identities were >80%. Amino acid sequence identities between CHS and other CHS superfamily enzymes in *Psilotum nudum* are lower than 60%, whereas the nucleotide sequence identity varies from 78 to 98% among the eight CHS genes in *Pisum sativum*.[77] The high degree of sequence identity found among the *P. patens* putative CHS ESTs strongly suggests that these ESTs most likely represent CHS genes and, thus, that there may be as many as seven CHS genes in *P. patens*. These genes are likely expressed at different developmental stages or under different growth conditions. Multiple copies of CHS genes in higher plants are well documented, and this is directly related to the functional diversity of flavonoids. Some of these genes are constitutively expressed, while others are transcriptionally induced by environmental factors such as UV light and pathogenic attack. Further studies may prove that similar functional divergence among multiple CHS genes operates in mosses as well. Since *P. patens* is applicable to genetic manipulations including gene targeting, the moss CHS genes may be found useful in studying, for example, light response gene regulation in primitive land plants.

A more complex pattern was observed in the 3′-UTR sequences of the moss CHS ESTs. In plants, the position of cleavage and polyadenylation of mRNA 3′-end can be heterogeneous within a single transcript.[93] The hexamer motif, AAUAAA, known to be required for the 3′-end processing in mammalian cells is less conserved or absent in plants, leading to the 3′-end heterogeneity.[94] In addition, there is often alternative splicing in non-coding regions of the plant mRNAs, thus generating diverse 3′-UTR sequences.[95] All these complex features were found in the moss CHS ESTs. A total of nine different

3′-UTR sequences were found (Fig. 16), where five sequences (3a, 3b, 3d, 3e, and 3i) were closely related, and the remaining four sequences (3c, 3f, 3g, and 3h) were distinct. The type A coding sequence was associated with four closely related 3′-UTR sequences, which are likely the products of alternative cleavage and/or splicing of a single transcript. Indeed, of the nine 3′-RACE products sequenced in Dr. Suh's laboratory, there were four 3′-UTR variants containing identical coding sequence. The 3′-termini of all clones analyzed were followed by poly(A) tails, suggesting that the observed multiple polyadenylation sites in the *P. patens* CHS gene are not cloning artifacts.

Multiple polyadenylation has been observed with higher plant genes such as chloroplast RNA-binding proteins in *Nicotiana plumbaginifolia* and a H^+-ATPase subunit A gene in *Arabidopsis thaliana*.[95] It has been suggested that variants in the non-coding region, resulting from alternative splicing are likely to be associated with different transcript stability, the biology of different cell types or, uncontrolled errors in the splicing process.[94, 95] The soon-to-be completed genomic structure of *P. patens* will help determine the legitimacy of the 3′-UTR heterogeneity in the moss CHS genes. Further investigation on the 3′-UTR sequences of the moss CHS genes may provide new insights into how the 3′-end formation is controlled in primitive plants.

The amino acid sequence identity of *P. patens* CHS was determined to be 62% as compared to other primitive plant CHSs, such as *Psilotum nudum* (BAA87922) and *Equisetum arvense* (BAA89501), and generally in the range of 60~65% with higher plant CHSs. The *P. patens* CHS sequence showed the highest amino acid sequence identity to a liverwort CHS-like gene product (BAD42329) (67% identity, 88% similarity), reflecting their close evolutionary relationship.

The strictly conserved CHS active-site residues, Cys170, His309, and Asn342 as well as the highly conserved CHS signature sequence, G378FGPG appear in the deduced amino acid sequence of *P. patens* CHS as shown in Fig. 17. The two Phe residues (Phe221 and Phe271), important in determining the substrate specificity of CHS[96], were also present.

3. 2. Functional expression of *P. patens* CHS in *Escherichia coli* and enzyme assay

To determine the enzyme activity of the cloned protein, the ORF of *P. patens* CHS was cloned into an *E. coli* expression vector pET-32a(+) vector, which introduces thioredoxin and a $(His)_6$-tag. Therefore the enzyme was heterologously overproduced as a thioredoxin-fusion protein (~62 kDa) and the $(His)_6$-tag enabled us to obtain the enzyme to a higher purity after a single purification step with Ni^{2+}-chelation chromatography.

As a representative SDS-PAGE result (Fig. 18) demonstrates, the expression levels and solubility of the recombinant thioredoxin fusion protein of *P. patens* CHS were comparable to those of *P. lobata* CHS.[82] Under the conditions employed, a good portion of the overproduced *P. patens* CHS in *E. coli* was recovered in soluble fractions with a yield of 4 mg/100 mL culture after purification. Earlier studies with CHSs from other plants showed that there is no functional difference between the thioredoxin fusion enzyme and the native enzyme recovered after removing the fusion parts.[33] Therefore, the thioredoxin fusion protein was used for further analysis in this study.

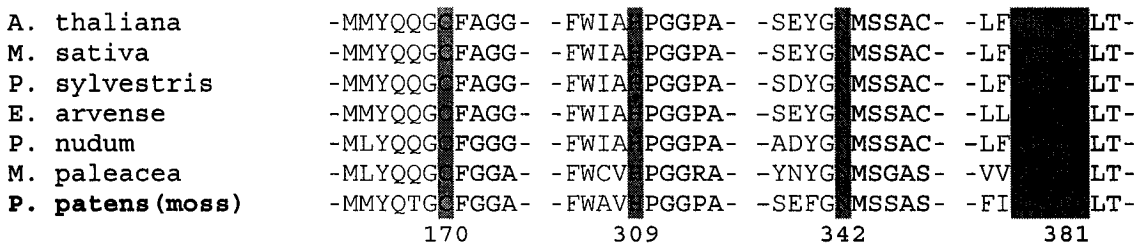


Figure 17. Alignment of the chalcone synthase sequence fragments from *Arabidopsis thaliana* (CAI30418), *Medicago sativa* (P30074), *Pinus sylvestris* (CAA43166), *Equisetum arvense* (Q9MBB1), *Psilotum nudum* (BAA87922), *Marchantia paleacea* (BAD42328), and *Physcomitrella patens* (ABB84527, this study). Sequences around the active site residues, Cys, His, and Asn, and the signature GFGPG loop are shown. The residue numbers correspond to those of *P. patens* chalcone synthase.

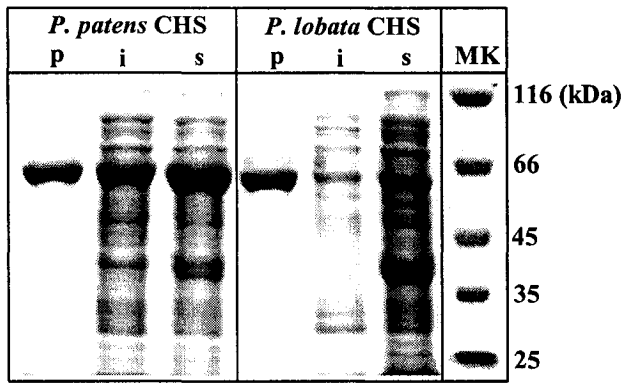


Figure 18. SDS-PAGE analysis for the overexpression and purification of *P. patens* CHS and *P. lobata* CHS. s: soluble fraction; i: insoluble fraction; p: purified protein; MK: protein marker.

In vitro enzyme assays using *p*-coumaroyl-CoA and [2-^{14}C] malonyl-CoA as substrates confirmed that this enzyme is indeed a CHS. The reaction products were identified by either using internal standards or by their known R_f values.[82] As shown in Fig. 19, naringenin was produced by *P. patens* CHS as the major product while CTAL (the derailment lactone after three condensations) and BNY (the derailment lactone after two condensations) were also produced as byproducts. This product profile was identical to that of *P. lobata* CHS, a representative higher plant CHS as illustrated in Table 2. The specific activity of *P. patens* CHS was determined to be 50 pkat/mg, comparable to that of *P. lobata* CHS (57 pkat/mg).[82]

3. 3. Characterization of *P. patens* chalcone synthase

P. patens CHS was further characterized for its kinetic properties. The K_m (app) values for *p*-coumaroyl-CoA and malonyl-CoA were 48.0 ± 5.2 μM (n=3, mean ± standard deviation) and 3.5 ± 0.3 μM, respectively, as shown in Fig. 20 and Table 3. These kinetic parameters were comparable to those of *P. lobata* CHS obtained under similar conditions.

The starter-CoA preference of *P. patens* CHS was studied using [2-^{14}C] malonyl-CoA and various CoA esters as the starter substrate and compared to those of *P. lobata* CHS as demonstrated in Fig. 21 and Fig. 22. The various cyclization products (marked by asterisks in Figs. 21 and 22) were identified based on their independence of reaction pH during extraction with ethyl acetate. The amounts of extracted cyclization products were essentially constant with or without acidification or alkalification. On the other hand, the amounts of extractable derailment lactone derivatives increased or decreased

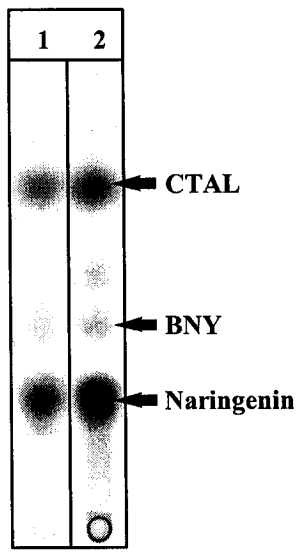
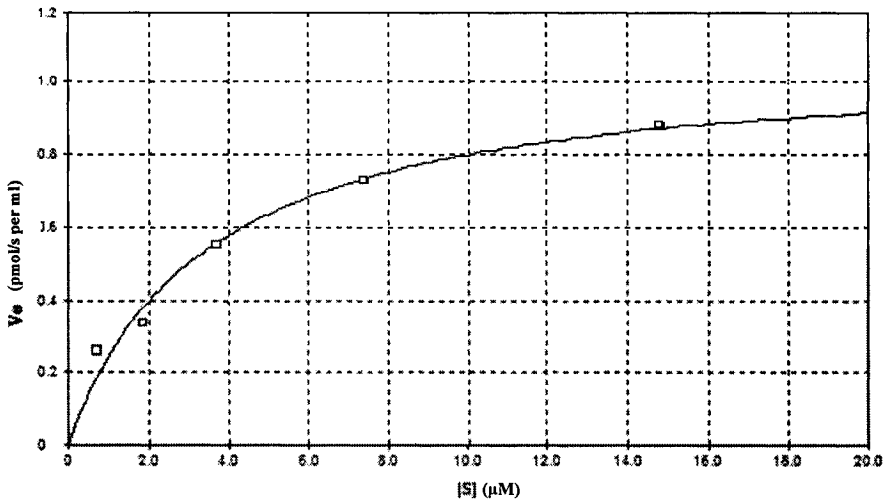


Figure 19. Radio-TLC of the CHS reactions. (1: *P. patens* CHS; 2: *P. lobata* CHS)

Table 2. Product profile of *P. patens* CHS and *P. lobata* CHS

	Product Profile (%)		
	Naringenin	CTAL	BNY
P. patens CHS	66	26	8
P. lobata CHS	70	23	7

A



B

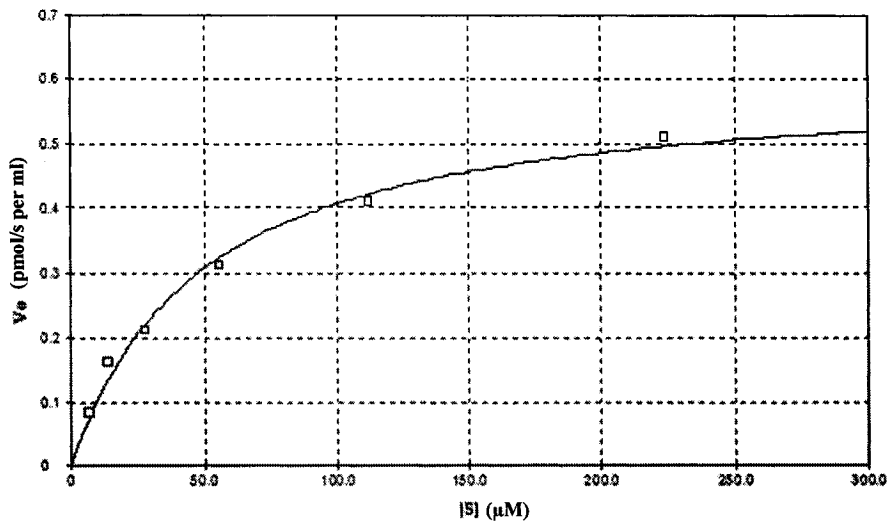


Figure 20. Michaelis-Menten plots of the *P. patens* CHS reaction. A: Michaelis-Menten plot for the determination of K_m(app) for malonyl-CoA; B: Michaelis-Menten plot for the determination of K_m(app) for *p*-coumaroyl-CoA.

Table 3. Kinetic parameters of *P. patens* CHS

Enzyme	K_m (app) (μM)		V_{max} (pmol/s per ml)
	p-Coumaroyl-CoA	Malonyl-CoA	
P. patens CHS	48.0 ± 5.2	3.5 ± 0.3	0.7 ± 0.1
P. lobata CHS	50.0 ± 7.1	3.0 ± 1.2	0.5 ± 0.3

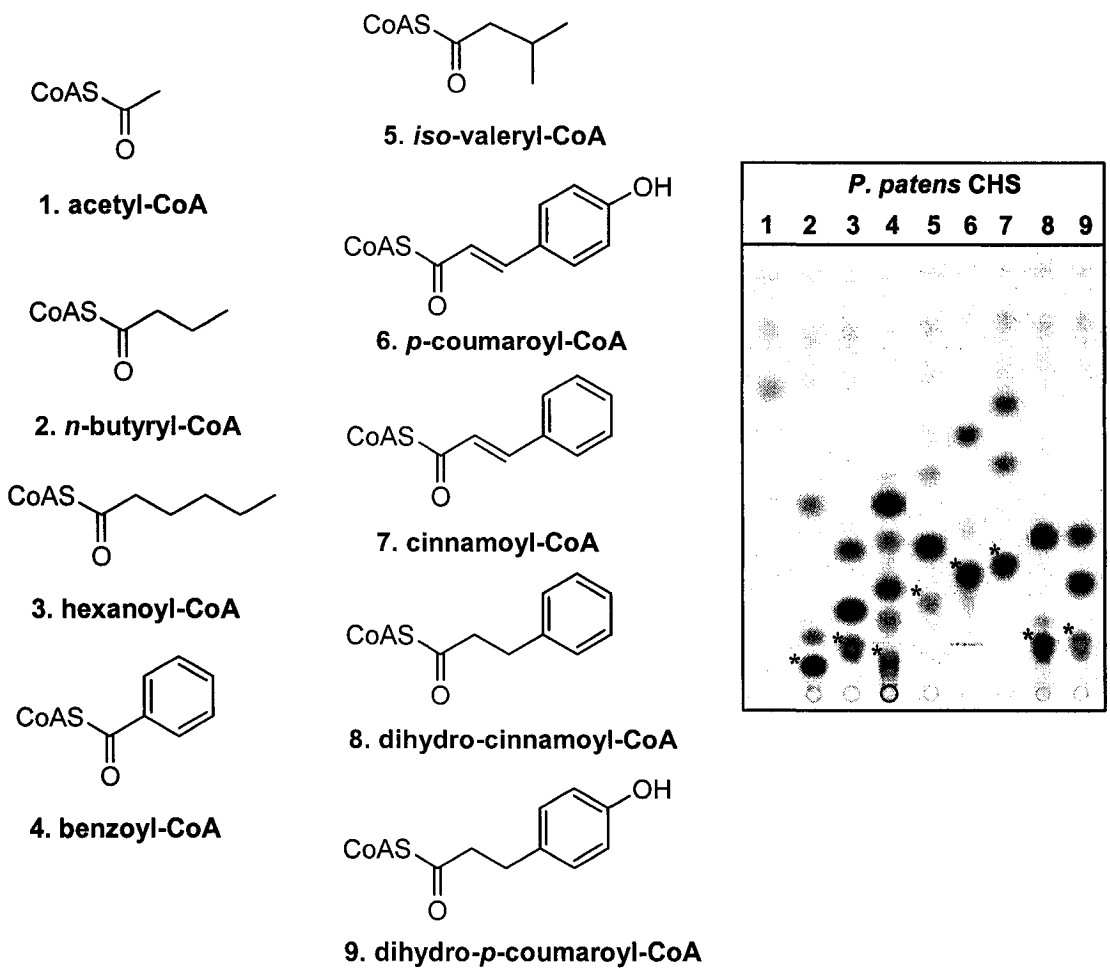


Figure 21. Radio-TLC of *P. patens* CHS reactions using various CoA esters as starter units. (1: acetyl-CoA; 2: *n*-butyryl-CoA; 3: hexanoyl-CoA; 4: benzoyl-CoA; 5: *iso*-valeryl-CoA; 6: *p*-coumaroyl-CoA; 7: dihydro-*p*-coumaroyl-CoA; 8: cinnamoyl-CoA; 9: dihydro-cinnamoyl-CoA). The structures of starter CoA esters are also shown.

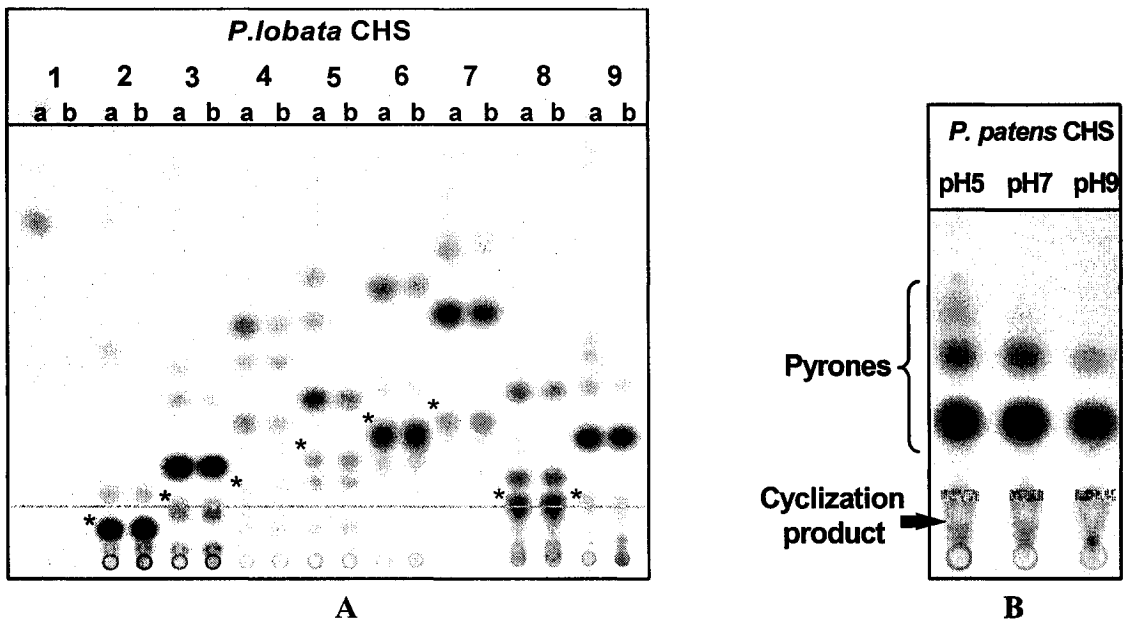


Figure 22. Radio-TLC of CHS reactions using various CoA esters as starter units. (**A**: *P. lobata* CHS; 1: acetyl-CoA; 2: *n*-butyryl-CoA; 3: hexanoyl-CoA; 4: benzoyl-CoA; 5: *iso*-valeryl-CoA; 6: *p*-coumaroyl-CoA; 7: dihydro-*p*-coumaroyl-CoA; 8: cinnamoyl-CoA; 9: dihydro-cinnamoyl-CoA; a: acidification before extraction; b: without acidification before extraction. **B**: Product extraction at different pH values. The *P. patens* CHS reaction with dihydro-cinnamoyl-CoA was conducted under standard conditions. The pH of the reaction mixture was then adjusted to different values with dilute HCl or NaOH before ethyl acetate extraction.)

substantially when the reaction mixtures were acidified or alkalified prior to the extraction as shown in Fig. 22-B.

P. patens CHS exhibited a preference profile towards CoA esters comparable to that of *P. lobata* CHS as depicted in Table 4. Among the CoA esters examined, *p*-coumaroyl-CoA was by far the most preferred substrate for the cyclization reaction of *P. patens* CHS. No study has been reported on the types of flavonoids produced by *P. patens*. However, apigenin and luteolin, both derived from *p*-coumaroyl-CoA, have been found in several moss species.[76] Cinnamoyl-CoA and dihydro-*p*-coumaroyl-CoA were converted to the corresponding chalcones by *P. patens* CHS at 50~60% efficiency as compared to *p*-coumaroyl-CoA, suggesting that *p*-coumaroyl-CoA is likely the *in vivo* substrate for *P. patens* CHS. *n*-Butyryl-CoA, hexanoyl-CoA, benzoyl-CoA were accepted with only 30-40% efficiency. It was also shown that *P. patens* CHS was less active with dihydro-*p*-coumaroyl-CoA than with *p*-coumaroyl-CoA, whereas *P. lobata* CHS displayed higher preference towards dihydro-*p*-coumaroyl-CoA. Both *P. patens* CHS and *P. lobata* CHS exhibited a general preference towards phenylpropanoid-CoA esters, a hallmark characteristic for plant CHSs.

Yamazaki et al.[82] reported that *P. nudum* CHS has no *in vitro* activity with benzoyl-CoA as substrate; however, a few CHSs from the higher plants such as *Hypericum androsaemum*[97], *Scutellaria baicalensisthe*[98], and *Cassia alata*[99] were reported to be able to accept benzoyl-Co as the starter substrate and convert it to the corresponding cyclization and lactone products. *P. patens* CHS seemed to produce a few cyclization and derailment lactone products with comparable efficiency to that of *P. lobata* CHS. In this regard, *P. patens* CHS is more similar to the higher plant CHSs in its

Table 4. Starter CoA preference of *P. patens* CHS and *P. lobata* CHS

Substrate	Relative activity (%)	
	P. patens CHS	*P. lobata* CHS
p-coumaroyl-CoA	(100)	(100)
acetyl-CoA	n.d.	n.d.
n-butyryl-CoA	43	39
hexanoyl-CoA	37	19
benzoyl-CoA	30	32
iso-valeryl-CoA	15	20
dihydro-*p*-coumaroyl-CoA	50	130
cinnamoyl-CoA	62	91
dihydro-cinnamoyl-CoA	23	25

(Relative activity was based on the cyclization products only. n.d.: not detected)

reactivity toward benzoyl-CoA, and differs from *P. nudum* CHS even though *P. nudum* CHS is more closely related to *P. patens* CHS. The identity of these benzoyl-CoA derived products and the mechanism of their production will be studied in the future.

3. 4. Phylogenetic analysis of CHS superfamily and other condensing enzymes

Earlier phylogenetic analyses carried out by Schroder et al. with a total of 34 CHS and 4 STS amino acid sequences showed that STSs always grouped with CHSs from the same or related plants in the phylogenetic tree. Based on these results, Schroder et al. proposed that there is no common ancestor STS gene and that STSs evolved from CHSs several times independently throughout evolution.[49] It has since been inferred that other members of the CHS superfamily evolved in the same manner as STS. Since Schroder's pioneering study in 1994, many new members of the superfamily have been discovered and a few phylogenetic studies have been reported. However, these studies were restricted to the evolution of the CHS superfamily in a single species (*Pisum sativum*[100], *Ipomea purpurea*[101], *Glycine max*[102]), in a single genus (*Ipomea*[103]), or in a single family (Asteraceae[104, 105], Poaceae[106]), and no study has included all members of the CHS superfamily in plants. In this study, Schroder's proposition was re-examined with all of known members of the CHS superfamily to date by using three different algorithms including the neighbour-joining (NJ) method, the maximum parsimony (MP) method, and the Bayesian estimation.

The results thus obtained showed that the majority of the CHS-like enzymes (Groups 1 to 6 in Fig. 23) indeed formed clusters with CHSs from the same or related species, in agreement with Schroder's original proposal.[49] However, a few CHS-like

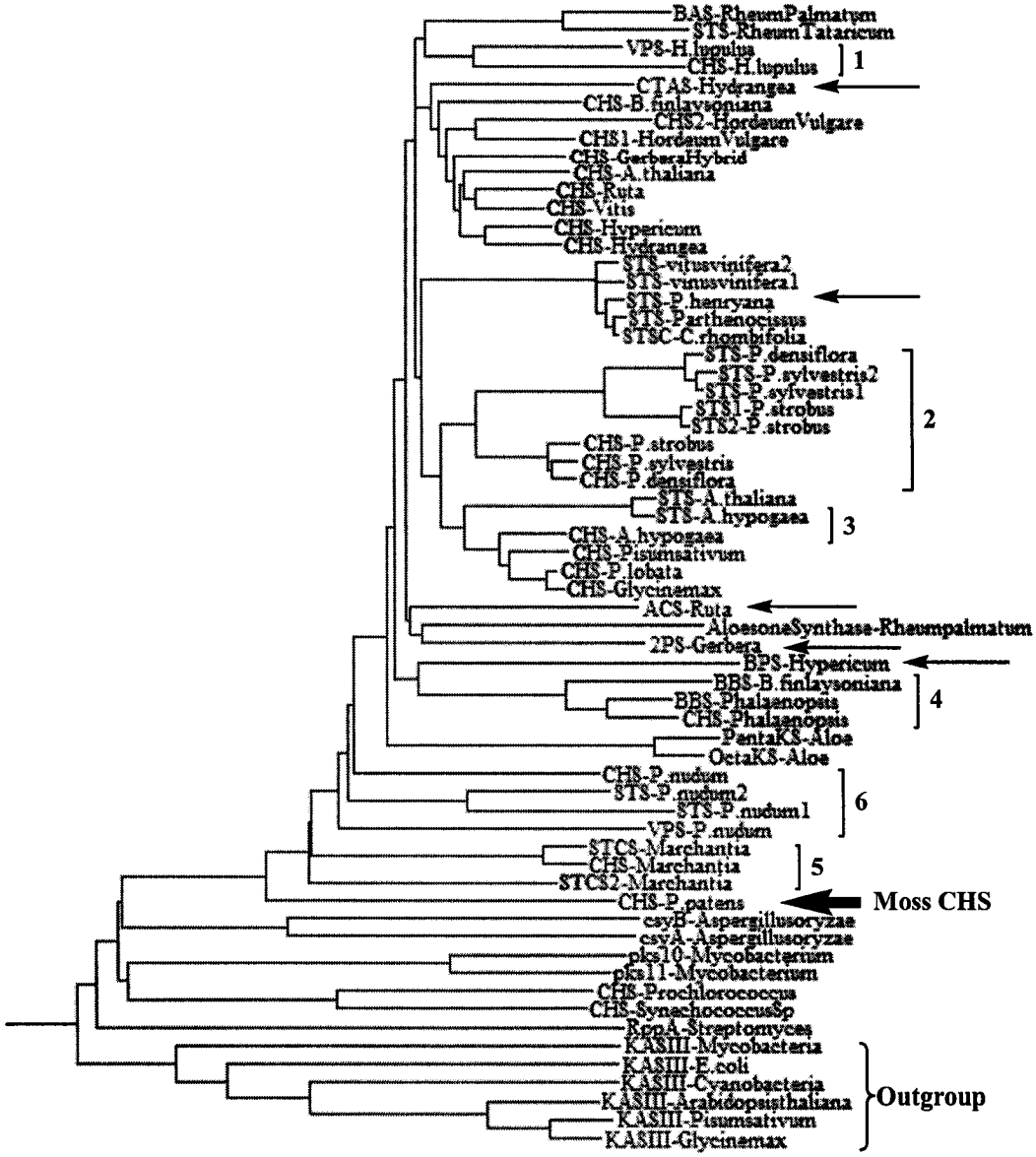


Figure 23. Phylogenetic analysis of the CHS superfamily using the neighbor-joining method. KAS III enzymes were used as outgroup.

enzymes such as *Vitis* STSs and *Gerbera* 2-pyrone synthase (2-PS) were not found in the same cluster with their related CHSs, but next to CHSs from unrelated species (marked with arrows in Fig. 23). These results suggest that the evolution pattern of the CHS superfamily may be more complex than previously realized. One possible evolutionary scenario that would explain the observed patterns is as follows. In some cases, there may have been gene duplication and functional diversification leading to STS and other CHS-like enzymes prior to plant speciation, and these multiple members would have been passed down to their descendants. Depending upon the environment and functional selection, some plant species have preserved the ancient STSs; thus, such STSs may appear unrelated to the independently evolved modern-day CHSs of the same species. On the other hand, other species may have lost the ancient STS genes during evolution, and after speciation, "reinvented" STSs (or other CHS-like enzymes) from their own CHSs. In both cases, the original proposition of "the independent evolution of STS from CHS" would hold true, but the timing of the duplication and diversification events relative to speciation would result in different clustering patterns in a phylogenetic tree. It is noted that these two scenarios are not mutually exclusive. Clearly, more sequence data of CHS and other CHS-like enzymes will be helpful to further define the evolutionary process.

In all phylogenetic analyses performed in this study, including amino acid sequences of all types of condensing enzymes from archaea, bacteria, fungi, and plants, the position of *P. patens* CHS in the phylogenetic tree was found to be consistent with its evolutionary position. Thus, *P. patens* CHS is located along with the two CHS superfamily proteins from the liverwort, *Marchantia*, at the base of the plant CHS

superfamily clade, in between the CHS-like gene products from microorganisms and the pteridophyte CHS superfamily enzymes (Fig. 24).

Condensing enzymes catalyze the formation of a carbon-carbon bond *via* a Claisen condensation reaction. These enzymes are often called the thiolase superfamily[107], to which the CHS superfamily, the KAS enzymes, and the KS domains of PKSs belong. As noted earlier, there are two types of condensing enzymes that catalyze mechanistically identical decarboxylative condensation reactions. KAS III enzymes and the CHS superfamily employ the Cys-His-Asn catalytic triad, while KAS I and II enzymes and the KS domains in PKSs utilize the Cys-His-His triad. Why these two groups of enzymes have a catalytic triad substitution, and their evolutionary relationship to one another remain unclear. An extensive phylogenetic analysis of these two types of condensing enzymes was aimed at providing insight into these questions. In the past, several reports on the phylogenetic analysis of the KS domains of type I PKSs[108], non-reducing KS domains from fungi[109], bacterial PKSs[110], and bacterial type III PKSs[111] appeared in the literature; however, these studies focused on only a certain type of condensing enzyme.

The results obtained in this study suggest that archaea KAS III enzymes[112] preceded all other condensing enzymes including KAS I and II enzymes, and that the C-H-H type active site may have evolved from the C-H-N type active site. Archaea possess unique membrane phospholipids of which hydrophobic side chains are generally isoprenoid derivatives instead of long chain fatty acids.[113] Thus, archaea do not have KAS I and II enzymes.[112] The possibility of horizontal gene transfer[114] of KAS III from bacteria to archaea is not supported by the data presented here. However, limited

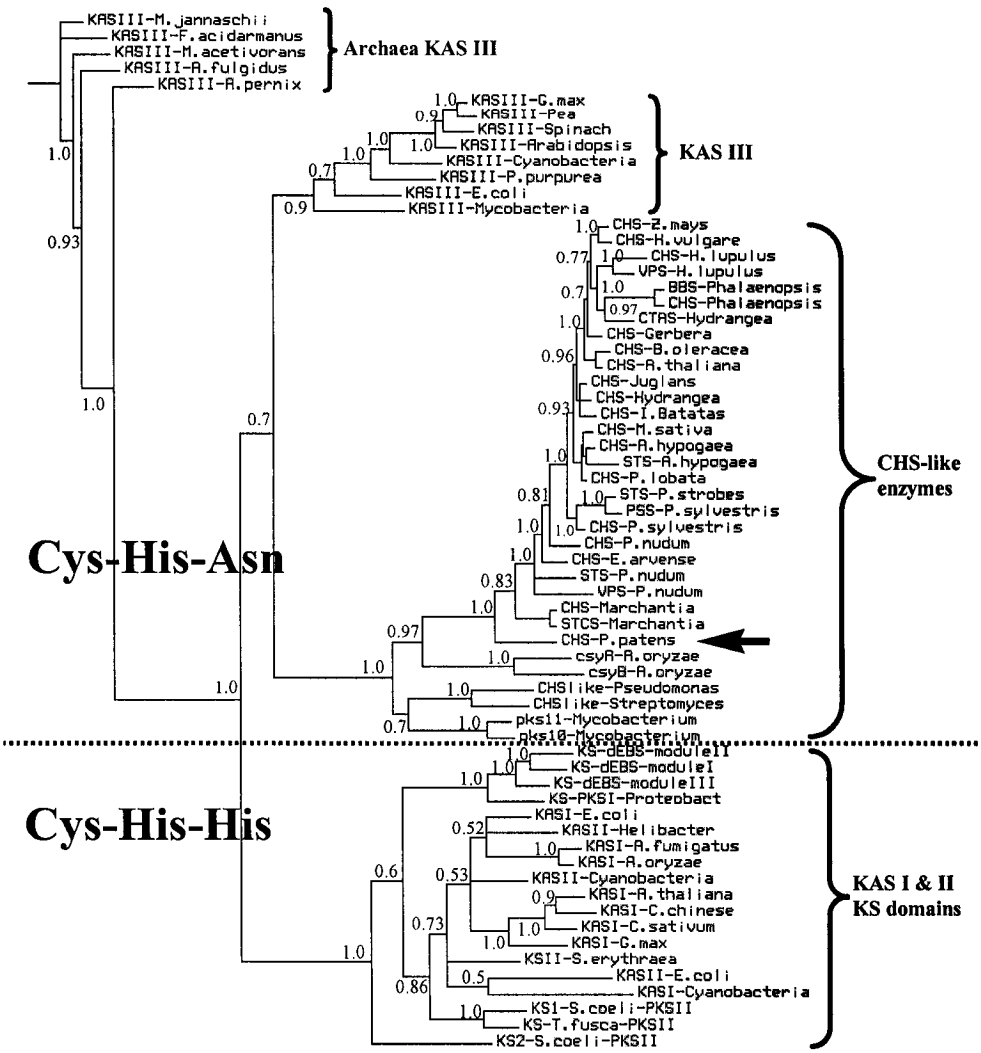


Figure 24. Phylogenetic analysis of condensing enzymes using the Bayesian method.

information on the KAS enzymes in archaea precludes a firm conclusion in this regard.

Interestingly, in the phylogenetic trees obtained in the study, KAS I and II, and the KS domains of PKSs formed an isolated cluster, while KAS III and the CHS superfamily formed another cluster. Thus, a line could be drawn dividing condensing enzymes into two parts, one with the C-H-H active site and the other with the C-H-N active site as shown in Fig. 24. This clear division of the condensing enzymes according to their active site types has never been reported in the literature.

A plausible evolutionary scenario can be put forward that agrees with the phylogenetic data obtained in this study. An ancestor condensing enzyme, which likely catalyzed a non-decarboxylative condensation (the Cys-His-Cys type), would first have emerged in the ancestor cells. Then the first decarboxylative condensing enzyme (later to become KAS III) that learned to couple the decarboxylation of the malonyl unit to the condensation reaction, which is more energetically favourable, would have evolved from the primitive non-decarboxylative condensing enzyme before the split of the archaea and bacteria lineages. Later, KAS I and II would have evolved from KAS III while changing their active site make-up to the Cys-His-His type for unknown reason(s). All these three types of KAS enzymes had played an essential role in the evolution of bacteria by providing means (membranes) to separate themselves from the surroundings.

The ancestor bacterial CHS-like enzymes would have been evolved from KAS III by acquiring the cyclization ability of polyketide intermediates. This would have been achieved in part by changing the gate-keeping SFGFG loop of the KAS III to the GFGPG loop that provides stereocontrol during the cyclization. Nonetheless, these ancestor CHS-like enzymes in bacteria still utilized simple acyl-CoA esters such as acetyl-CoA and

malonyl-CoA as the starter substrates. Then, the primitive CHS, which evolved from these ancestor CHS-like genes would have first appeared in primitive plants, liverworts and mosses, following the evolution of the phenylpropanoid pathway and plants' land invasion about 450 million years ago. The evolution of the phenylpropanoid pathway enabled the primitive CHS to adapt phenylpropanoid-CoA as a starter unit and to produce flavonoids that played a significant role as UV protector during the successful colonization of land by plants. All the other diverse members of the plant CHS superfamily have subsequently developed from CHS through gene duplication that may have occurred at different times throughout evolution. Functional diversification of the superfamily was achieved by start-CoA specialization, by changing the number of condensation reactions, and by utilizing different cyclization mechanisms to provide higher plants with chemical arsenals to survive in a more complicated environment.

4. Conclusion

Simple flavonoids (flavanones, flavones and flavanols) are found in bryophytes, and as such it is thought that CHS first appeared in bryophytes during the early evolution of land plants. Therefore, the *P. patens* CHS characterized in this study represents a modern day version of one of the oldest CHSs. *P. patens* CHS exhibited similar enzymatic characteristics as compared to higher plant CHSs. Together with the liverwort CHS-like genes, *P. patens* CHS provides the long-missing bridge in the phylogenetic tree between CHS-like enzymes from microorganisms and pteridophyte CHS superfamily enzymes. Based on the phylogenetic analysis, an evolutionary model is proposed for all the condensing enzymes, in which archaea KAS III serves as the ancestor for other condensing enzymes including KAS I and II, and the CHS superfamily.

Part II: Investigation of the cyclization mechanism of stilbene synthase

1. Introduction

1. 1. Different cyclization modes in polyketide synthase

Polyketides that are produced by the type III PKSs are highly diverse in chemical structure and bioactivities as a result of their various catalytic mechanisms. The type III PKSs accept several starter units, catalyze different steps of elongation, and have diverse cyclization mechanisms. Four different cyclization modes of lactonization, acylation, aldol reaction, and decarboxylative aldol reaction have been identified for the type III PKS-catalyzed cyclization reactions as shown in Fig. 25. During cyclization by lactonization, there are two possible mechanisms, in which oxygen is the nucleophile that attacks the target carbonyl carbon to form a new carbon-oxygen bond for both. It was believed that 6-methyl-4-hydroxy-2-pyrone in daisy[115] and hispidin in mushrooms[116] are the lactonization products. 6-Methyl-4-hydroxy-2-pyrone is produced from acetyl-CoA and two malonyl-CoA molecules by 2-pyrone synthase, while hispidin is synthesized from one molecule of caffeoyl-CoA and two molecules of malonyl-CoA by styrylpyrone synthase. In the chalcone type cyclization by acylation, CHS cyclizes the linear tetraketide precursors *via* an intramolecular Claisen condensation to form a new carbon-carbon bond.[26] The nondecarboxylative aldol type cyclization is involved in the biosynthesis of stilbenecarboxylic acid type compounds such as lunularic acid in liverworts[117] and hydrangenol in *Hydrangea* species.[118] On the other hand, the decarboxylative aldol type cyclization by STS in the biosynthesis of stilbenes eliminates carbon dioxide during the cyclization process.

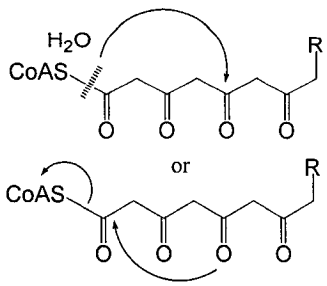
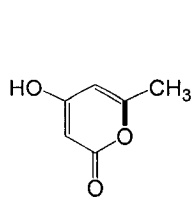
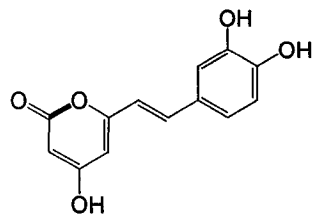
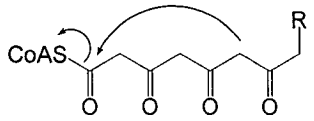
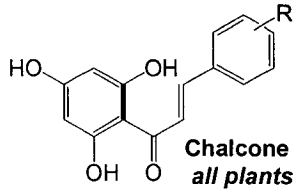
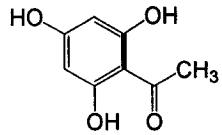
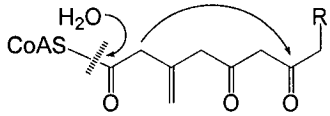
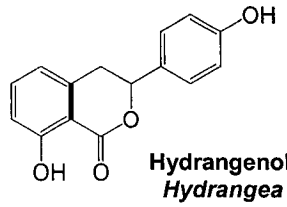
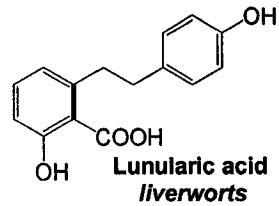
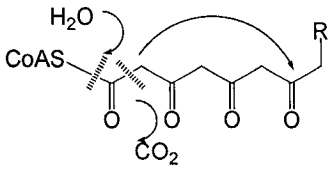
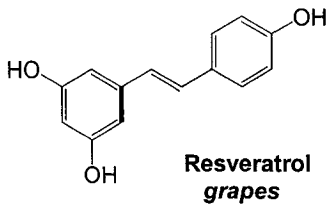
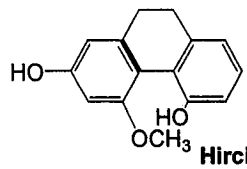
Cyclization modes	Mechanism	Examples
Lactonization		 **6-Methyl-4-hydroxy-2-pyrone** ***Gerbera hybrida***  **Hispidin** ***mushrooms***
Acylation (chalcone type)		 **Chalcone** ***all plants***  **Phloracetophenone** ***ginger***
Aldol reaction		 **Hydrangenol** ***Hydrangea***  **Lunularic acid** ***liverworts***
Decarboxylative Aldol reaction (Stilbene type)		 **Resveratrol** ***grapes***  **Hircinol** ***orchid***

Figure 25. Different cyclization reactions catalyzed by the type III PKSs and their corresponding products. The chemical bonds newly formed during cyclization are shown in thick lines.

1. 2. Different cyclization reactions catalyzed by chalcone and stilbene synthases

As representative members of the CHS superfamily, CHS and STS share common structural and functional features with the same catalytic mechanism for decarboxylative Claisen condensation. As shown in Fig. 26, both CHS and STS use a phenylpropanoid CoA-ester as their starter unit and then process three sequential condensation reactions with malonyl-CoA to form the linear polyketide intermediate. Subsequently, however, CHS and STS cyclize the tetraketide intermediate by entirely different mechanisms, connecting different carbon atoms of the tetraketide intermediate during the ring closure to form either chalcone or stilbene.[12]

The mechanism of the decarboxylative condensation has been extensively studied with CHS as a model enzyme.[12] However, the cyclization mechanism of the STS-catalyzed reaction is not well understood.[16] All STS enzymes that have been analyzed remove the terminal carboxyl group of the tetraketide intermediate. Therefore, the STS-catalyzed cyclization involves hydrolysis, decarboxylation, cyclization, and aromatization; however, the sequence of these chemical events is yet to be determined.[119, 120]

1. 3. Factors controlling the stilbene synthase type cyclization process

The active site structures of CHS, STS, and other condensing enzymes such as KAS share common features as shown in Fig. 9 (Part I).[29] The three-dimensional structure of alfalfa (*Medicago sativa*) CHS[15, 111] and peanut (*Arachis hypogaea*) STS[120, 122] revealed that the strictly conserved amino acid residues of CHS and STS, including His303, Asn336, Pro375, and Phe215 (amino acid numbering according to alfalfa CHS

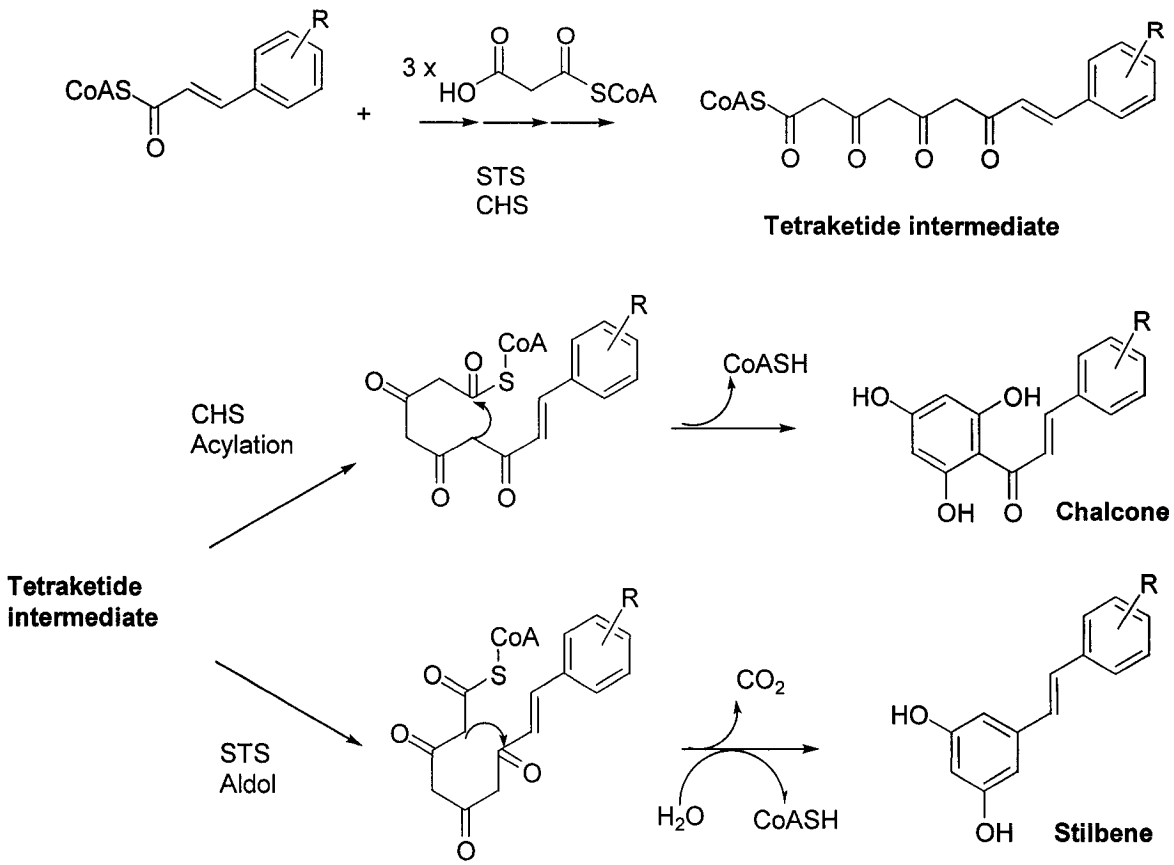


Figure 26. Different cyclization reactions catalyzed by chalcone and stilbene synthases.

and peanut STS), largely define the active sites of CHS and STS. In particular, the crystal structures of alfalfa CHS and peanut STS showed that the G^{372}FGPG loop serves as the "sidewall" of the active site, suggesting that it may play a significant role in determining the outcome of the cyclization reactions of CHS superfamily enzymes. Indeed, it was proposed, based on the mutagenic studies, that the GFGPG loop provides stereo-confinement during the folding process of the tetraketide intermediate.[29] As mentioned in Part I, the Pro375 to Gly mutant of CHS produced increased amounts of CTAL, the derailment product at the stage of the tetraketide (Fig. 7; Part I). This result indicated that the ability of the CHS mutant to catalyze the CHS type cyclization was severely impeded. In STS, the same mutation of Pro to Gly not only resulted in an increased production of both derailment products of BNY and CTAL, but this STS-P375G mutant also catalyzed the CHS-type cyclization.[29] These results suggested that different folding of the common tetraketide intermediate brings different sets of carbon atoms close in proximity in CHS and STS and this is facilitated by their different active site geometries.

Earlier biomimetic studies on the cyclization of polyketide compounds showed that both Claisen and aldol type intramolecular cyclization reactions were readily achievable in solution over a range of pH values.[123, 124] These studies also demonstrated that the cyclization mode of linear polyketides in solution is determined by the presence or absence of an ester bond at the C1 carboxylate.[125] When C1 is part of an ester (or thioester) bond, the C6→C1 Claisen cyclization predominates. In contrast, the C2→C7 aldol cyclization is favoured when C1 exists as a free acid (Fig. 27). It seems reasonable to expect then that similar chemistry may operate during the CHS- and STS-catalyzed cyclization reactions. In the CHS-catalyzed cyclization, the C6 and C1 carbons

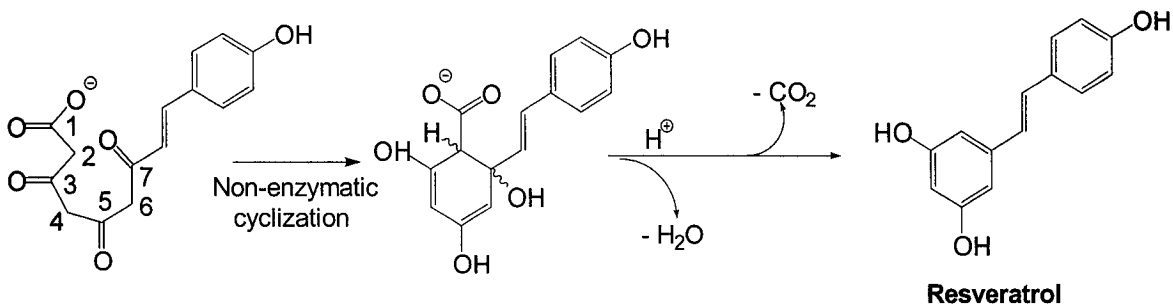


Figure 27. Non-enzymatic cyclization by aldol condensation producing stilbene in solution.

are connected by an intramolecular Claisen condensation with the CoA molecule acting as a leaving group. In contrast, during the STS-catalyzed cyclization (Fig. 26), the hydrolysis of the thioester bond prior to cyclization would prevent the competing CHS-type cyclization. Indeed, X-ray crystallographic studies of STS[120, 122] provided structural evidence for the hydrolysis reaction unique in the STS-type cyclization. In the crystal structures of STSs from both pine and peanut, there was a water molecule optimally situated at the enzyme active site, reminiscent of the activated water molecule in the active sites of type II *E. coli* thioesterases[126] (Fig. 28). Thus, the thioesterase-like hydrogen bond network (dubbed as the "aldol switch") allows Glu192, through Thr132, to activate a Ser338-positioned water molecule to a nucleophilic hydroxide anion, which initiates the hydrolysis reaction. On the other hand, such a hydrogen bond network is not possible in CHS due to a slightly shifted positioning of the water molecule in the active site.[15] As shown in Fig. 28, the distance between the water molecule and the hydroxyl group of Thr132 in CHS is one angstrom farther than that in the STS active site, thus prohibiting the hydrogen bond formation. Hence it was proposed that the precise positioning of the water molecule at the active site plays a key role in turning on the "aldol switch" system.

Although the presence of the "aldol switch" at the STS active site provided physical evidence for the hydrolysis reaction during the STS-catalyzed cyclization, it does not define when it occurs in the overall mechanism. In principle, there could be hydrolysis at several different stages during the ring formation, prior to or after the cyclization, or following aromatization. Inferring from the presence of natural products of stilbenecarboxylic acid skeleton, like lunularic acid from the liverwort *Lunularia cruciata*

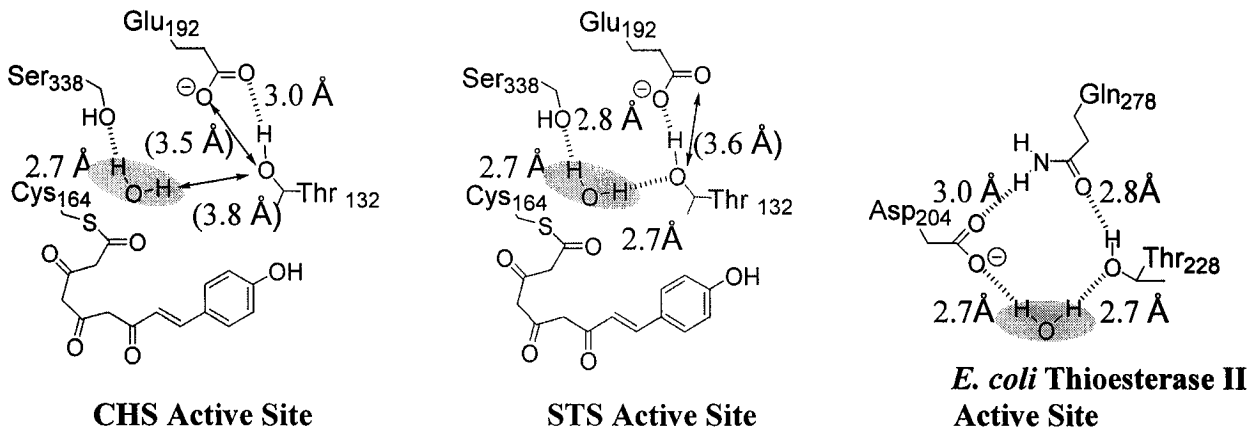


Figure 28. Proposed "aldol switch" hydrogen bond network at the STS active site.

and hydrangeic acid from *Hydrangea* (Fig. 25), Schroder had proposed that STS first produces stilbenecarboxylic acid as an intermediate, followed by the decarboxylation of the stilbenecarboxylic acid intermediate (Fig. 29; Pathway A).[12] However, the fact that stilbenecarboxylic acid had never been detected in the STS reaction mixture, made this mechanism less likely.[127] Later a Japanese group successfully showed that the intermediacy of stilbenecarboxylic acid in the STS reaction was not possible using an elegant isotope incorporation study.[127] When the STS reaction was carried out in the presence of [d_2]-malonyl-CoA and D_2O, resveratrol labelled with three deuterium atoms (resveratrol-d_3) was detected indicating that there was cyclization prior to ring formation (Pathway B in Fig. 29). If the STS-catalyzed reaction had proceeded *via* a stilbene carboxylic acid, resveratrol-d_3 would have not been formed (Pathway A in Fig. 29).

It should also be noted that pathway B depicted in Fig. 29 is one of several plausible reaction pathways that can lead to the deuterium labelled resveratrol. One plausible mechanism was proposed by Noel based on the three-dimensional structure of STS and the solution chemistry of polyketides.[123, 124, 125] The prominent feature of Noel's mechanism is that it includes cyclic intermediates produced by a nondecarboxylative aldol condensation as shown in Fig. 30 (A and B).[120] Based on the 3-D modeling of both linear and cyclized intermediates in the STS active site, it was further predicted that the initial cyclized intermediate derived from the nondecarboxylative aldol cyclization is most likely to have 2*S*, 7*R* stereochemistry. Thus, the proposed mechanism predicts that the conversion of the tetraketide intermediate to resveratrol by STS proceeds as (1) the hydrolysis of the thioester bond to give *p*-coumaroyltriacetic acid (CTAA), (2) C2→C7 aldol cyclization to form a cyclic intermediate, and (3) decarboxylation and

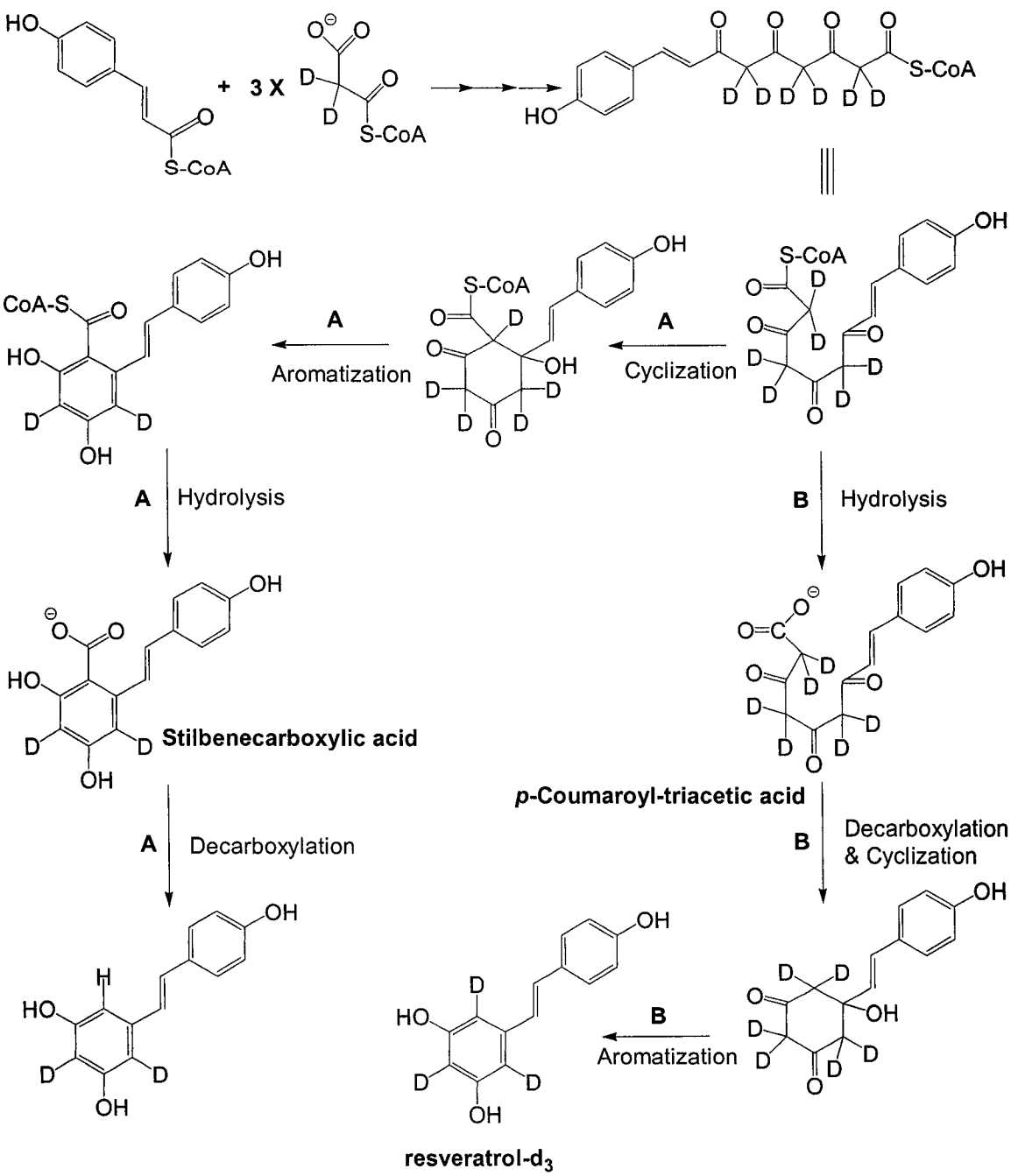


Figure 29. Reaction schemes leading to the deuterium labeled resveratrol (resveratrol-d_3).

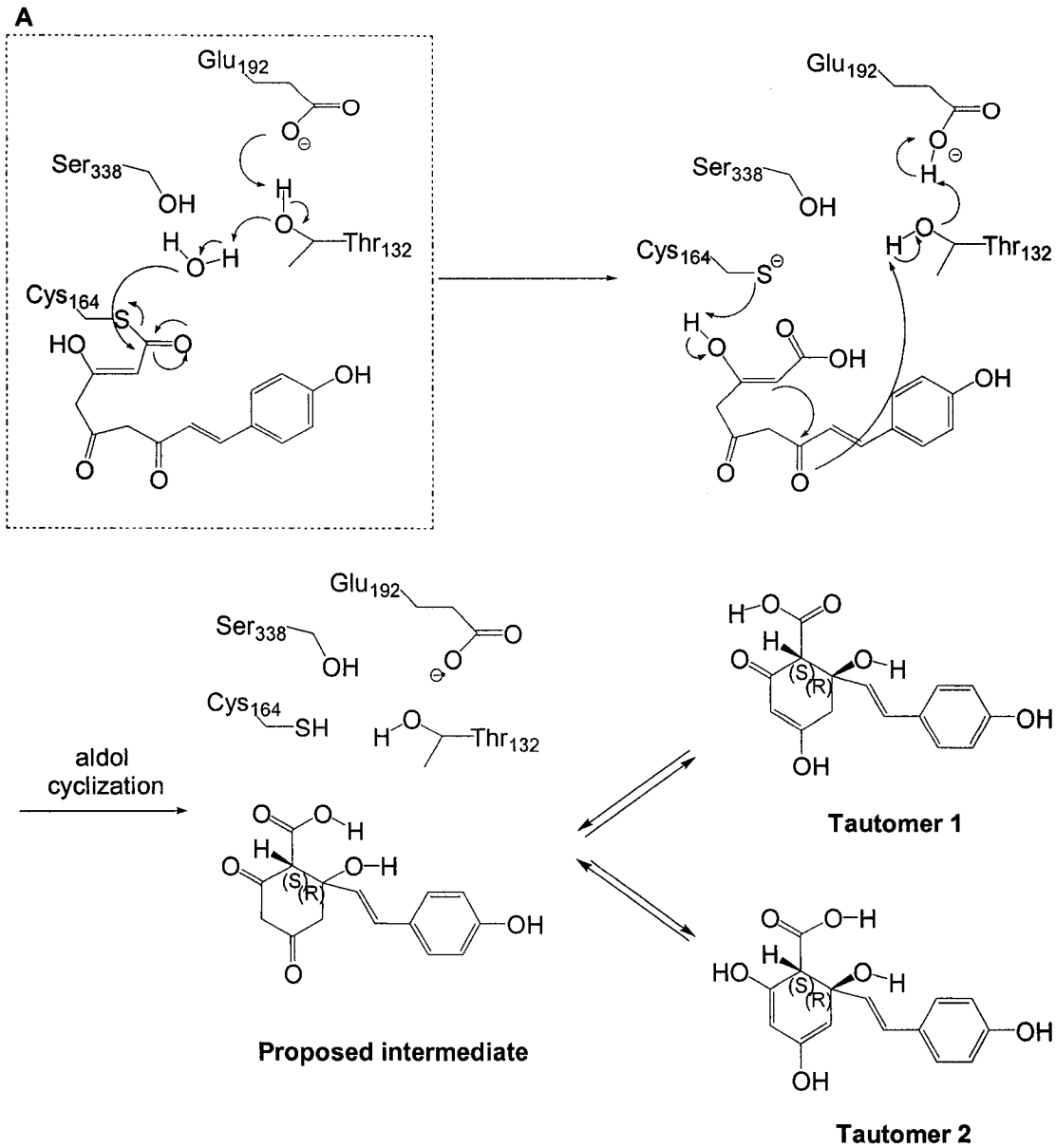


Figure 30. The proposed reaction mechanism of stilbene synthase-catalyzed cyclization. (Reproduced from reference 120 with permission).

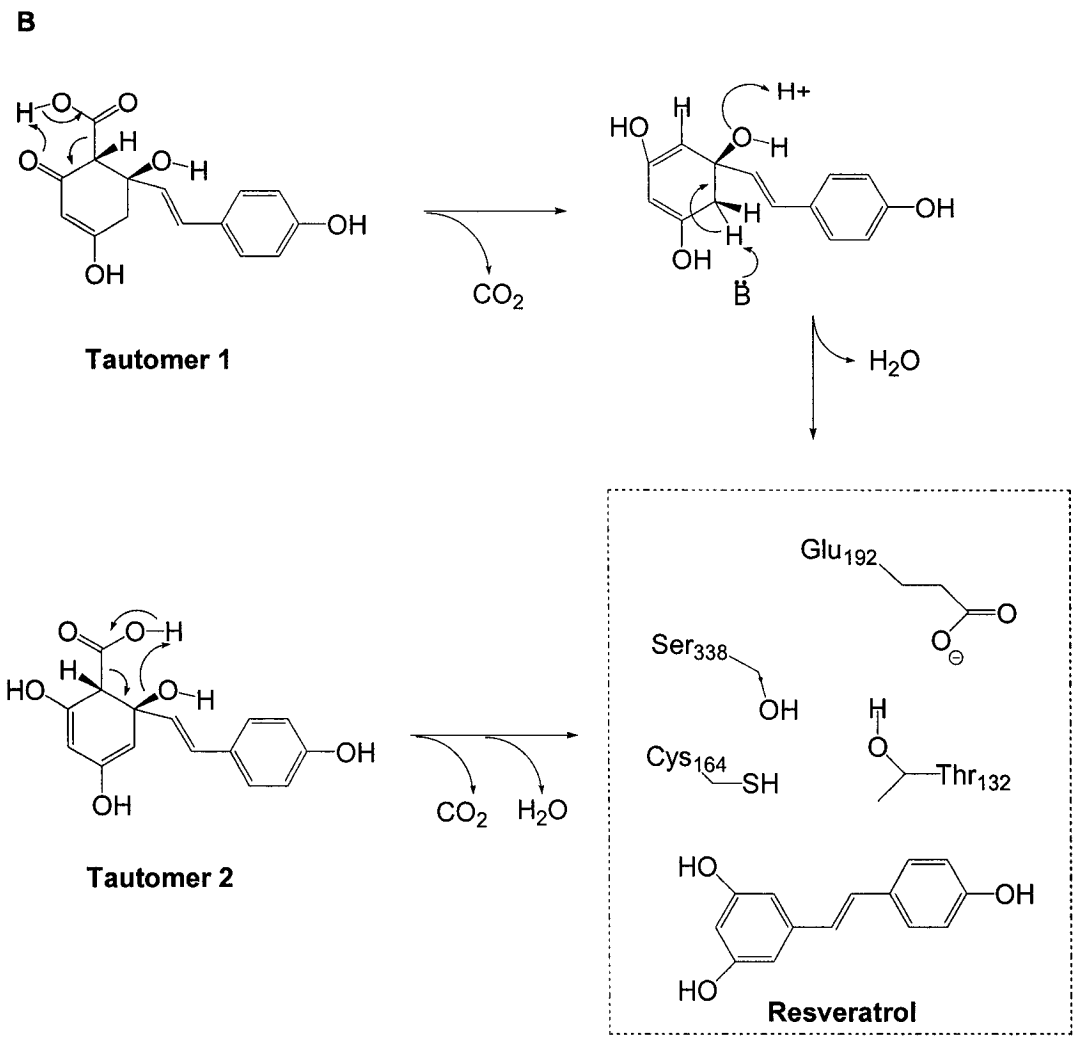


Figure 30. (Continued)

dehydration to complete the new ring formation.

There are other equally plausible mechanisms (Fig. 31). In pathway A, aldol cyclization precedes hydrolysis, which is followed by decarboxylation/dehydration. In pathway B, on the other hand, there would be hydrolysis of the thioester bond first, but the resulting CTAA intermediate would undergo decarboxylative aldol cyclization to give a cyclic intermediate, which would readily be converted to resveratrol by dehydration/aromatization.

1. 4. The intermediacy of *p*-coumaroyltriacetic acid

A common feature of Noel's mechanism (Fig. 30) and pathway B in Fig. 31 is the involvement of CTAA as a reaction intermediate. The fact that *in vitro* STS produces much smaller amounts of CTAL, which is the lactonized form of CTAA, than CHS (CTAL : resveratrol = 1 : 8 by STS *vs* CTAL: naringenin chalcone = 1 : 3 by CHS)[29] suggested that CTAA might be an intermediate of the STS reaction while being a dead-end derailment product in the CHS reaction. However, direct evidence for the intermediacy of CTAA in the STS reaction has been lacking.

p-Coumaroyltriacetic acid synthase (CTAS), a homologue of chalcone synthase from *Hydrangea macrophylla*[129], was used to produce CTAL in this study. Unlike CHS and STS, CTAS catalyzes three consecutive decarboxylative condensations of *p*-coumaroyl-CoA and malonyl-CoA molecules but does not catalyze any type of cyclization reaction, thus producing the linear tetraketide as the main product. When the CTAS reaction mixture, or the CHS and STS reaction mixtures, is acidified,

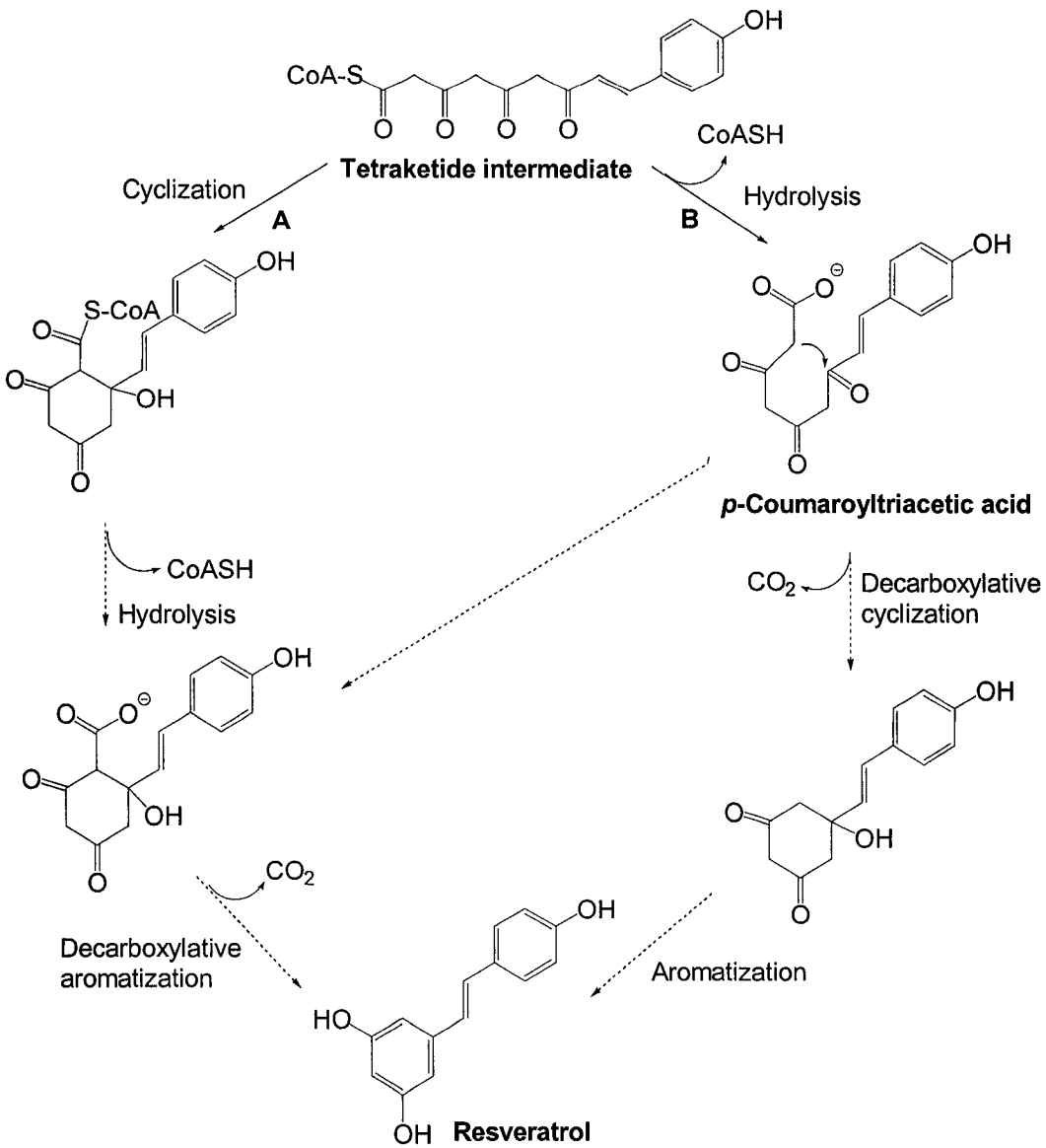


Figure 31. Possible mechanisms of the stilbene synthase reaction.

CTAL can be extracted by organic solvents. However, when the extraction is carried out without acidification, the amounts of CTAL detected from the CTAS reaction mixture were drastically reduced. This suggested that CTAA, which is water soluble, may be the true product of CTAS, and that the interconversion of CTAA and CTAL is pH-dependent as shown in Fig. 32 (see below for more details). To demonstrate the intermediacy of CTAA in the STS reaction, the direct conversion of CTAA to resveratrol by STS was explored.

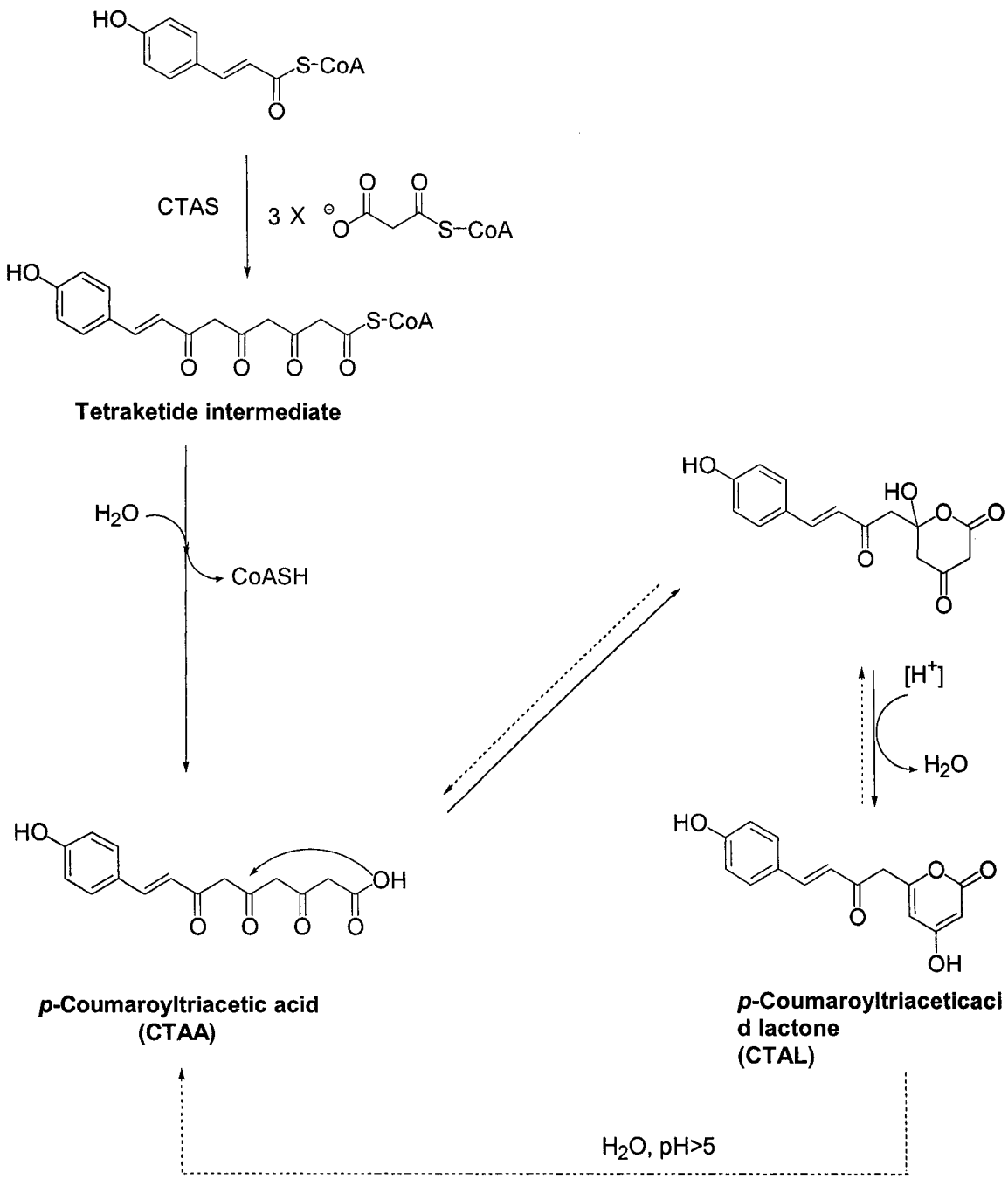


Figure 32. The reaction catalyzed by *p*-coumaroyltriacetic acid synthase (CTAS) to produce the *p*-coumaroyltriacetic acid.

2. Materials and Experimental Procedures

2. 1. Overexpression and purification

Pueraria lobata CHS[29], *Arachis hypogaea* STS[29], and *Hydrangea macrophylla* var. *thunbergii* CTAS[101] were overexpressed in *E. coli* AD494 as thioredoxin fusion proteins and purified as previously described (Part I, Section 2).

2. 2. Conversion of *p*-coumaroyltriacetic acid to resveratrol by stilbene synthase

The CTAS reaction was performed with the appropriate amount of enzyme (~50 μg), *p*-coumaroyl-CoA (100 μM) and [2-^{14}C] malonyl-CoA (17 μM) in 1 mL of 0.1 M KPi buffer (pH 7.2) containing 1 mM DTT, 10% glycerol, and 0.1 % Triton X-100 at 37 °C for 1 hr.[82] The reaction mixture was acidified by addition of 50 μL of 1 N HCl and extracted with 2 mL of ethyl acetate.

The solvent was evaporated, and thus obtained CTAL (4 nmol) was added to the STS reaction mixture (40 μL) containing sufficient amounts of STS (5 ~ 10 μg) in 0.1 M KPi buffer (pH 7.2) containing 0.1 % Triton X-100. A portion (5 μL) of the reaction mixture was withdrawn at different time intervals and analyzed by RP-TLC as described earlier (Part I, Section 2). A parallel reaction was also carried out with CHS and CTAL under the same conditions.

2. 3. The effect of pH and substrate concentration on the conversion of CTAA to resveratrol by STS

To investigate the pH dependence of the enzymatic conversion of CTAA to resveratrol by STS, the standard reactions of CHS, STS and CTAS were first conducted under different pH values (6, 7, and 8); and then the same amounts of CTAL produced by CTAS were added to the STS solution to a final concentration of 100 μM in 0.1 M KPi buffer at the three pH values. After incubation at 37 °C for 2 h, a portion (5 μL) of each reaction mixture was analyzed by RP-TLC. Also, to determine the dependence of the STS reaction on the concentration of CTAL, the enzyme reaction was performed at five different concentrations of CTAL ranging from 50 to 300 μM.

3. Results and Discussion

3. 1. The intermediacy of CTAA in the STS reaction

As shown in Fig. 33, acidification of the reaction mixture (pH<5) with HCl (1 N) at the end of enzyme reaction greatly enhanced the amounts of derailment products CTAL and BNY, detected by radio-TLC for both STS and CTAS reactions. These results suggest that the majority of these derailment compounds are present in the reaction mixture as free carboxylic acids at neutral pH, and thus are poorly extractable by ethyl acetate. In the CHS and STS reactions, these derailment compounds are formed by the non-enzymatic lactonization process in solution. Although the mechanistic details are yet to be elucidated, CTAS is thought to produce the linear CTAA, which is then converted to CTAL non-enzymatically as in the CHS and STS reactions. A similar pH-dependence of extractable derailed lactones has been reported.[101] As mentioned earlier, CTAL is produced as a derailment product by the non-enzymatic hydrolysis of the growing polyketide chain at the stage of tetraketide intermediate during the *in vitro* CHS and STS reactions (Fig. 7; Part I).[29, 82, 101] It was thought that CTAA (a triketide acid) and CTAL (a β-keto-2-pyrone) are in equilibrium and interconvertible under appropriate conditions. CTAA is conceptually converted to CTAL by cyclization and dehydration *via* a hemiketal intermediate, while CTAL can be converted to CTAA either by hydration and retro-aldol reaction or by hydrolysis of the lactone ring (Fig. 32). The conversion of CTAA to CTAL is acid-catalyzed; thus, it was expected that the conversion of CTAL to CTAA might be favored under non-acidic conditions.

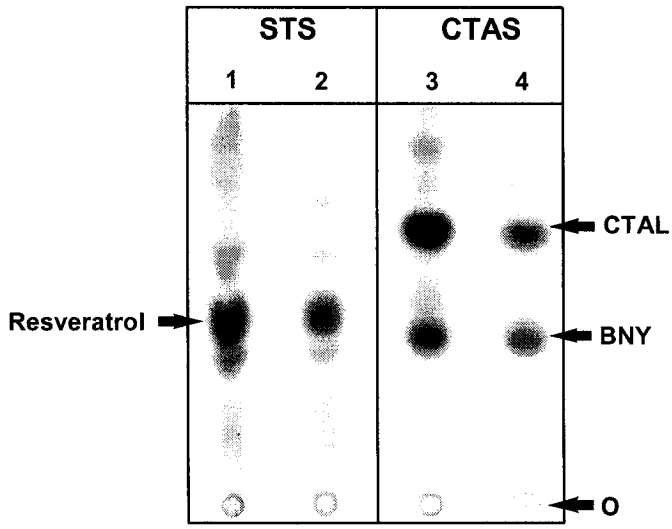


Figure 33. Radio-thin layer chromatogram of STS, and CTAS reactions. Lanes 1 and 3: standard reactions with acidification prior to ethyl acetate extraction; Lanes 2 and 4: parallel reactions without acidification. o: TLC origin.

In a preliminary study to assess whether STS accepts CTAA as a substrate and cyclizes it to resveratrol, CTAS was first reacted with *p*-coumaroyl-CoA and malonyl-CoA, and then STS was added directly to the mixture that should contain CTAA as well as CTAL. The amounts of CTAL detected in the reaction mixture (Fig. 34) that had undergone sequential CTAS and STS reactions decreased as compared to the amounts of CTAL in the CTAS reaction alone. Furthermore, CTAL seemed to be quantitatively converted to resveratrol by STS, suggesting that STS may indeed accept CTAA as a substrate and catalyze the ring formation to give resveratrol. However, this preliminary experiment could not completely exclude the possibility that STS might have converted the remaining *p*-coumaroyl-CoA and malonyl-CoA in the reaction mixture to resveratrol. Hence, STS was then reacted with isolated CTAL.

At the end of the CTAS reaction, all the products were converted to lactones (CTAL and BNY) by acid treatment and separated from the remaining CoA ester substrates by ethyl acetate extraction. After the solvent was evaporated, the lactones were redissolved in a reaction mixture containing STS. Then at different time intervals, a portion of the reaction mixture was withdrawn, acidified, and analyzed by RP-TLC. This time course study (Fig. 35) showed that resveratrol was produced by STS at the expense of CTAL. These results clearly demonstrated that CTAA, which is derived from CTAL, was indeed accepted as substrate and cyclized to resveratrol by STS. It should be noted that BNY remained unchanged throughout the reaction. The corresponding acid of BNY is a diketo acid that can not undergo intramolecular cyclization reaction.

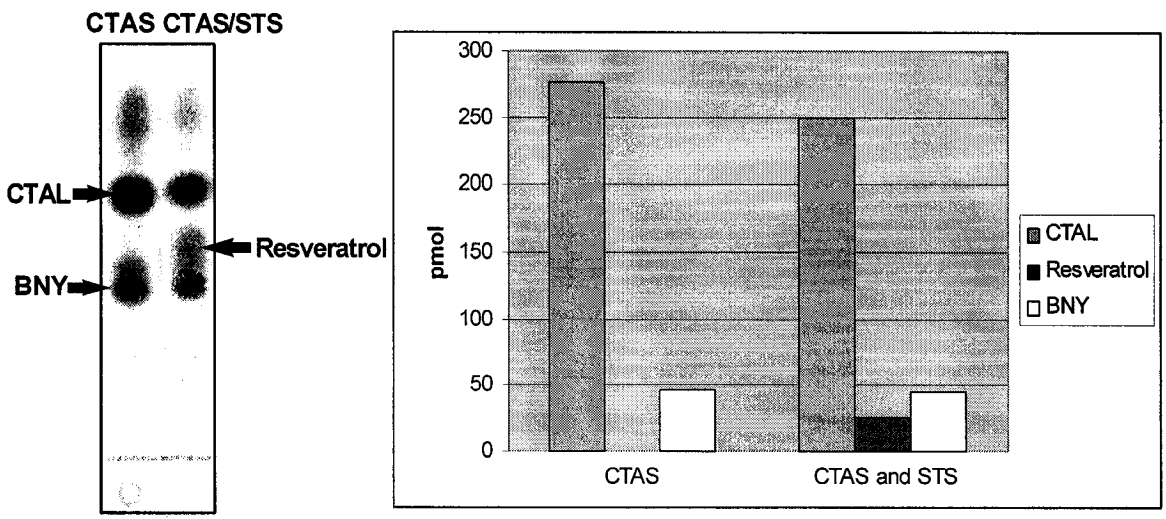


Figure 34. Analysis of the individual CTAS reaction and STS/CTAS reaction in combination.

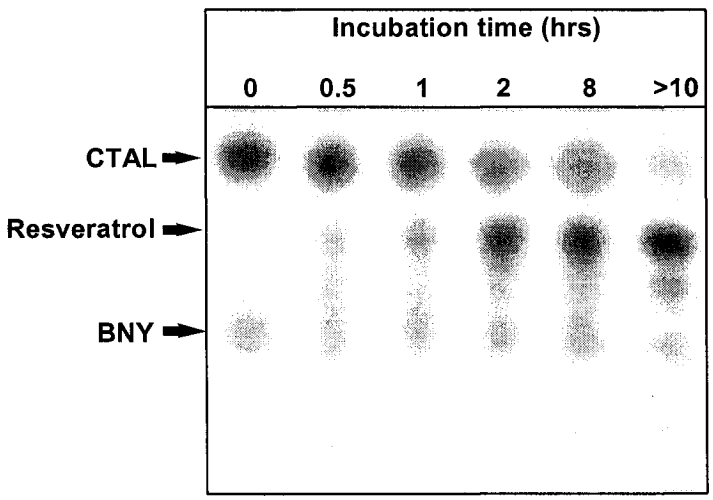


Figure 35. Radio-TLC of the time-course study of STS reacted with CTAL. In lane 6, more STS was added after 8 hrs incubation and the reaction mixture was further incubated for additional 2 hrs.

As expected, CHS was unable to cyclize CTAA to naringenin chalcone due to its different cyclization mechanism. A similar time-course study using CHS and CTAL as substrate (Fig. 36) showed that absolutely no naringenin chalcone was produced under identical conditions. Instead, a trace amount of resveratrol was detected after a prolonged incubation with CHS (less than 5% of the amount of resveratrol produced by a parallel STS reaction under the same conditions), due to the cross-reaction activity of CHS (resveratrol production by CHS).[82]

3. 2. The effects of pH and substrate concentration on the STS-catalyzed cyclization of CTAA

pH Dependence and kinetic studies of the conversion of CTAA to resveratrol by STS were used to further verify the enzymatic nature of the resveratrol production from CTAA by STS and to gain more information on the cyclization process. The analysis of the pH dependence (Fig. 37) revealed that the standard STS reaction was pH-dependent in that the resveratrol production was highest at pH 7 (Panel A). The results obtained at pH 6 and 8 were informative since the resveratrol production from the CoA ester substrates (*p*-coumaroyl-CoA and malonyl-CoA) was comparable at pH 6 and 8. In contrast, the resveratrol production from CTAL was barely detectable at pH 6, while a fair amount of resveratrol (20 % as compared to that of pH 7) was produced at pH 8. This sharp decrease in the resveratrol production from CTAL at pH 6 most likely resulted from little CTAL to CTAA conversion, a requisite for STS cyclization. The observation that less resveratrol was produced from CTAL at pH 8 as compared to pH 7 may be explained by the lower activity of STS at pH 8, even though the concentration

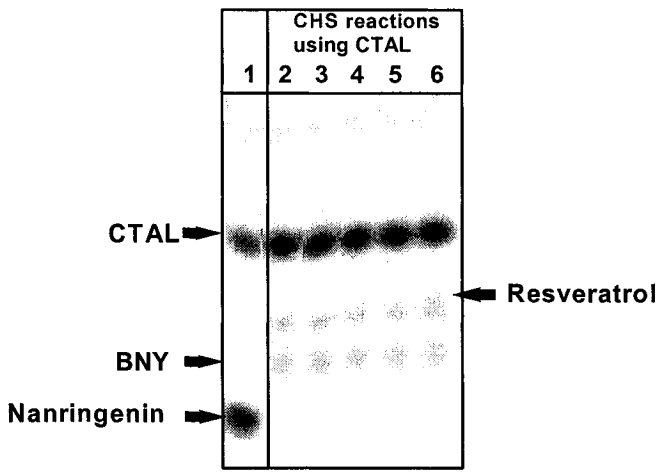


Figure 36. Radio-TLC of the CHS reactions with CTAL. (1: Standard CHS reaction; 2~6: time course studies of the CHS reactions with CTAL from t = 0 to t > 5 hrs.)

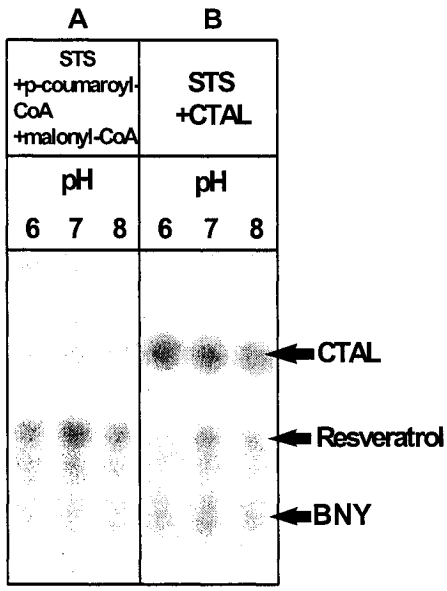


Figure 37. Radio-TLC of the STS and CTAS reactions with (**A**) the CoA esters as substrates, and (**B**) CTAL as a substrate at different pH values.

of available CTAA should have been higher at pH 8 than at pH 7 (Fig. 37).

Fig. 38 shows that the resveratrol production from CTAL was substrate concentration-dependent. The reaction velocity (v_o) was at its half-maximum when the initial concentration of CTAL was 200 μM. The relative amounts of CTAL and CTAA present in the reaction mixture could not be accurately determined. However, an estimation can be made as follows. When the CTAS reaction mixture was acidified prior to extraction, the amounts of CTAL detected increased 4-fold as compared to extraction at neutral pH. If one assumes that the conversion of CTAA to CTAL after acidification is complete and that the efficiencies of ethyl acetate extraction of CTAL and CTAA are 100 % and 0 %, respectively; then it can be estimated that their ratio (CTAA : CTAL) is 4 under the standard reaction conditions (pH 7.2). Therefore, when 200 μM of CTAL is added to a reaction mixture at pH 7.2, the concentrations of CTAA and CTAL at equilibrium will be 150 μM and 50 μM, respectively. Therefore, the apparent K_m value for CTAA under these reaction conditions was estimated to be close to 150 μM. This relatively high $K_{m(app)}$ value indicates that CTAA is a weak substrate compared to the natural CoA ester substrates. The CoA moiety p-coumaroyl-CoA and malonyl-CoA plays a significant role in the substrate binding to CHS and STS.[129] Therefore, the absence of the CoA moiety in CTAA should reduce the binding affinity of this substrate for the enzyme.

Interestingly, a similar cyclization of a triketide acid to a stilbene had been reported in the literature.[130] In 1969, in a study on the biosynthesis of a stilbene, pinosylvin, Hillis and Ishikura observed that synthetic cinnamoyl triacetic acid was

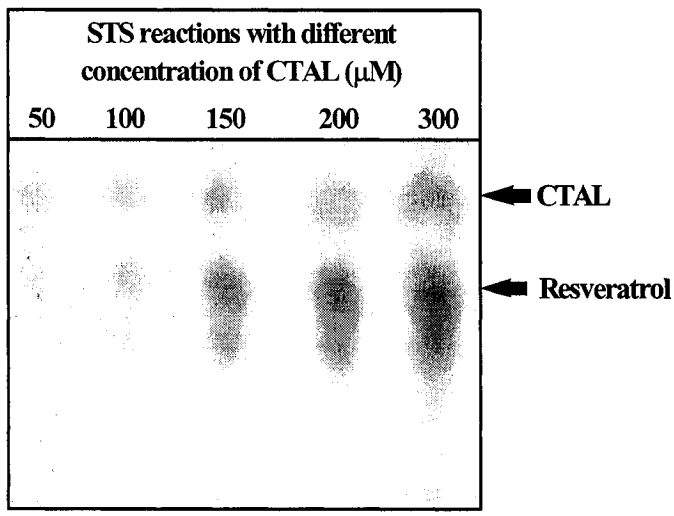


Figure 38. Radio-TLC of the STS reaction with different concentrations of CTAL.

converted to pinosylvin by crude enzyme extracts of the leaves of an *Eucaryptus* species shown to contain pinosylvin (Fig. 39). Neither buffer alone nor heat-treated enzyme preparation catalyzed the pinosylvin formation. Although the authors wrongly postulated that cinnamoyltriacetic acid was cyclized to pinosylvin *via* pinosylvic acid (a stilbene carboxylic acid), theirs was the first report that suggested the intermediacy of a triketo acid in the STS reaction.

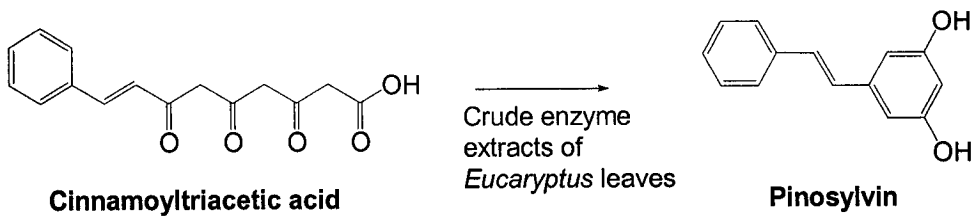


Figure 39. The cyclization of cinnamoyltriacetic acid to pinosylvin by enzyme extracts of *Eucaryptus* leaves.

4. Conclusion

This study successfully demonstrated, for the first time, that STS indeed accepts CTAA as substrate and is capable of converting it to resveratrol enzymatically. Therefore, CTAA is a reaction intermediate in the STS reaction and the tetraketide-CoA ester must be hydrolyzed prior to the cyclization, (pathway B, Fig. 40). The mechanism of how CTAA is converted to resveratrol, either cyclization followed by decarboxylation and aromatization (Noel's proposed mechanism) or decarboxylative cyclization and aromatization, still remains to be elucidated.

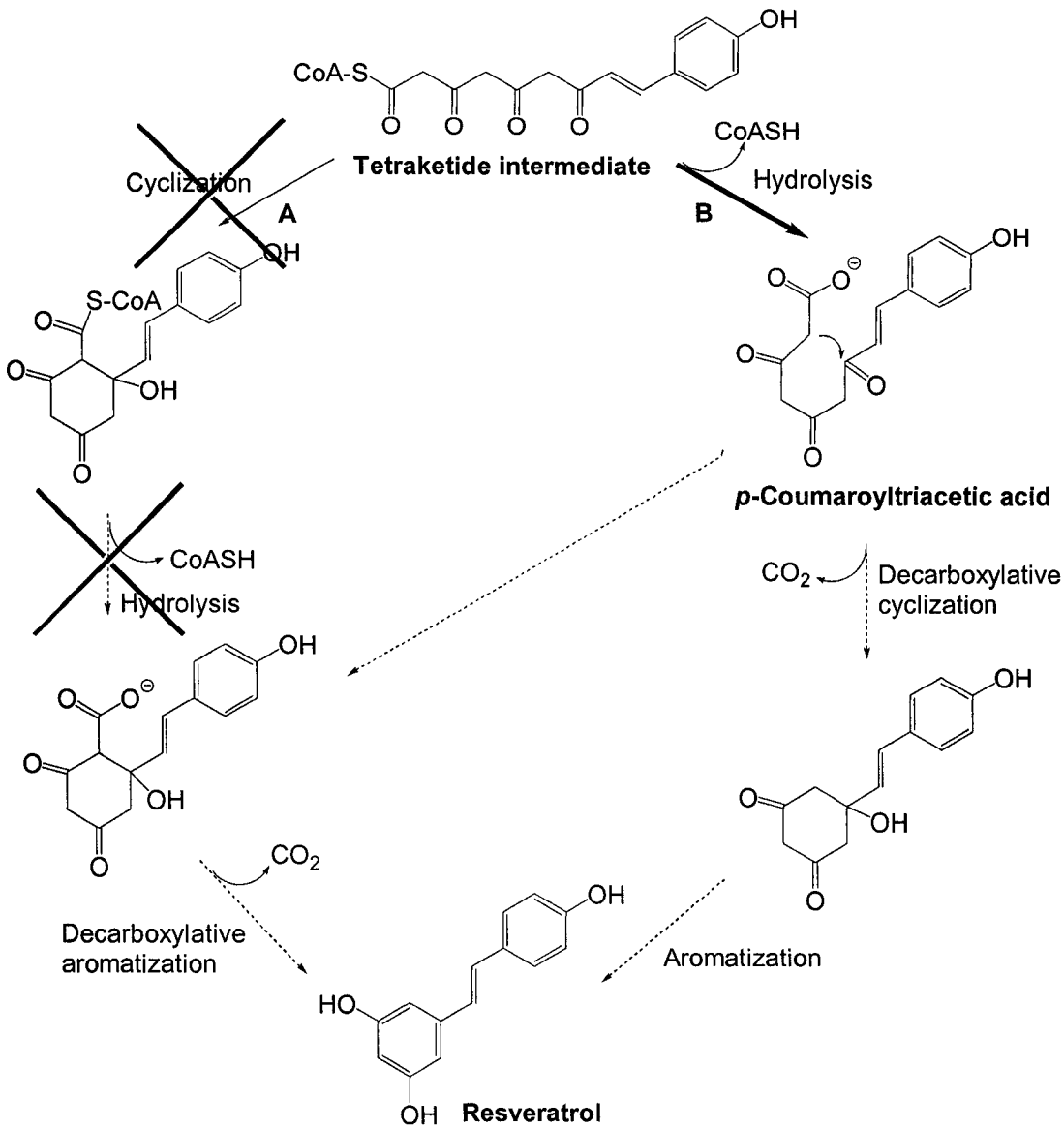


Figure 40. Possible cyclization mechanisms of stilbene synthase including the confirmed first step.

REFERENCES

1. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **1991**, 252, 675-679.
2. Hopwood, D. A., Sherman, D. H. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **1990**, 24, 37-66.
3. Donadio, S., Katz, L. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **1992**, 111, 51-60.
4. Beck, J., Ripka, S., Siegner, A., Schiltz, E., Schweizer, E. The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. *Eur. J. Biochem.* **1990**, 192, 487-498.
5. Rudd, B. A., Hopwood, D. A. Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **1979**, 114, 35-43.
6. Yue, S., Motamedi, H., Wendt-Pienkowski, E., Hutchinson, C. R. Anthracycline metabolites of tetracenomycin C-nonproducing *Streptomyces glaucescens* mutants. *J. Bacteriol.* **1986**, 167, 581-586.
7. McDaniel, R., Eberta-Khosla, S., Fu, H., Hopwood, D. A., Khosla, C. Engineered biosynthesis of novel polyketides: influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 11542-11546.

8. Robinson, J. A. Polyketide synthase complexes: their structure and function in antibiotic biosynthesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1991**, 332, 104-114.

9. Holak, T. A., Nilges, M., Prestegard, J. H., Gronenborn, A. M., Clore, G. M. Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry. *Eur. J. Biochem.* **1988**, 175, 9-15.

10. Xu, G. Y., Tam, A., Lin, L., Hixon, J., Fritz, C. C., Powers, R. Solution structure of *B. subtilis* acyl carrier protein. *Structure* **2001**, 9, 277-287.

11. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A., Simpson T. J. Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **1997**, 36, 6000-6008.

12. Schroder, J., Schroder, G. Stilbene and chalcone synthases: related enzymes with key functions in plant-specific pathways. *Z. Naturforsch.* **1990**, 45, 1-8.

13. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, 7, 285-295.

14. Qiu, X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khanderkar, S., Lonsdale, J., Abdel-Meguid, S. S. Crystal structure of beta-ketoacyl-acyl carrier protein synthase III. A key condensing enzyme in bacterial fatty acid biosynthesis. *J. Biol. Chem.* **1999**, 274, 36465-36471.

15. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., Noel, J. P. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **1999**, 6, 775-784.

16. Austin, M. B., Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **2003**, 20, 79-110.

17. Schroder, J. Probing plant polyketide biosynthesis. *Nat. Struct. Biol.* **1999**, 6, 714-716.

18. Lukacin, R., Springob, K., Urbanke, C., Ernwein, C., Schroder, G., Schroder, J., Matern, U. Native acridone synthases I and II from *Ruta graveolens* L. form homodimers. *FEBS Lett.* **1999**, 448, 135-140.

19. Lukacin, R., Schreiner, S., Matern, U. Transformation of acridone synthase to chalcone synthase. *FEBS Lett.* **2001**, 508, 413-417.

20. 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.

21. Paniego, N. B., Zuurbier, K.W., Fung, S.Y., van der Heijden, R., Scheffer, J. J., Verpoorte, R. Phlorisovalerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. *Eur. J. Biochem.* **1999**, 262, 612-616.

22. Borejsza-Wysocki, W., Hrazdina, G. Aromatic polyketide synthases purification, characterization, and antibody development to benzalacetone synthase from raspberry fruits). *Plant Physiol.* **1996**, 110, 791-799.

23. Beckert, C., Horn, C., Schnitzler, J.-P., Lehning, A., Heller, W., Veit, M. Styrylpyrone biosynthesis in *Equisetum arvense*. *Phytochemistry* **1997**, 44, 275-283.

24. Nambudiri, A. M., Vance, C. P., Towers, G. H. Styrylpyrone biosynthesis in *Polyporus hispidus*. II. Enzymic hydroxylation of *p*-coumaric acid and bis-noryangonin. *Biochim. Biophys. Acta.* **1974**, 343, 148-155.

25. Moore, B. S., Hertweck, C., Hopke, J. N., Izumikawa, M., Kalaitzis, J. A., Nilsen, G., O'Hare, T., Piel, J., Shipley, P. R., Xiang, L., Austin, M. B., Noel, J. P. Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. *J. Nat. Prod.* **2002**, 65, 1956-1962.

26. Kreuzaler, F., Hahlbrock, K. Enzymic synthesis of an aromatic ring from acetate units. Partial purification and some properties of flavanone synthase from cell-suspension cultures of *Petroselinum hortense*. *Eur. J. Biochem.* **1975**, 56, 205-213.

27. Jez, J. M., Noel, J. P. 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, 39640-39646.

28. Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A., Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **2000**, 39, 890-902.

29. Suh, D.-Y., Fukuma, K., Kagami, J., Yamazaki, Y., Shibuya, M., Ebizuka, Y., Sankawa, U. Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Biochem. J.* **2000**, 350, 229-235.

30. Raudnitz, H. Concerning a pigment commonly attributed to the presence of leuco-anthocyanin. *Science* **1958**, 128, 782.

31. Kreuzaler, F., Hahlbrock, K. Enzymatic synthesis of aromatic compounds in higher plants: formation of naringenin (5,7,4'-trihydroxyflavanone) from *p*-coumaroyl coenzyme A and malonyl coenzyme A. *FEBS Lett.* **1972**, 28, 69-72.

32. Hagmann, M. L., Heller, W., Grisebach, H. Induction and characterization of a microsomal flavonoid 3'-hydroxylase from parsley cell cultures. *Eur. J. Biochem.* **1983**, 134, 547-554.

33. Mol, J. N. M., Robbinst, M. P., Dixon, R. A., Veltkamp, E. Spontaneous and enzymic rearrangement of naringenin chalcone to flavanone. *Phytochemistry* **1985**, 24, 2267-2269.

34. Erdmann, R., Kunau, W. H. Purification and immunolocalization of the peroxisomal 3-oxoacyl-CoA thiolase from *Saccharomyces cerevisiae*. *Yeast* **1994**, 10, 1173-1182.

35. Suh, D.-Y., Kagami, J., Fukuma, K., Sankawa, U. Evidence for catalytic cysteine-histidine dyad in chalcone synthase. *Biochem. Biophys. Res. Commun.* **2000**, 275, 725-730.

36. Rupert, C. S. Repair of ultraviolet damage in cellular DNA. *J. Cell Comp. Physiol.* **1961**, 58, 57-68.

37. Witkin, E. M. Ultraviolet-induced mutation and DNA repair. *Annu. Rev. Microbiol.* **1969**, 23, 487-514.

38. Pirozynski, K. A., Malloch, D. W. The origin of land plants: a matter of mycotrophism. *Biosystems* **1975**, 6, 153-164.

39. Stafford, H. A. Flavonoid evolution: An enzymic approach. *Plant Physiol.* **1991**, 96, 680-685.

40. Carefoot, T. H., Karentz, D., Pennings, S. C., Young, C. L. Distribution of mycosporine-like amino acids in the sea hare *Aplysia dactylomela*: effect of diet on amounts and types sequestered over time in tissues and spawn. *Comp. Biochem. Physiol. C. Toxicol. Pharmacol.* **2000**, 126, 91-104.

41. Garcia-Pichel, F., Castenholz, R. W. Occurrence of UV-absorbing, mycosporine-like compounds among cyanobacterial isolates and an estimate of their screening capacity. *Appl. Environ. Microbiol.* **1993**, 59, 163-169.

42. Suh, H. J., Lee, H. W., Jung, J. Mycosporine glycine protects biological systems against photodynamic damage by quenching singlet oxygen with a high efficiency. *Photochem. Photobiol.* **2003**, 78, 109-113.

43. Markham, K. R. Bryophyte flavonoids, their structures, distribution, and evolutionary significance. *Proc. Phytochem. Soc.* **1990**, 29, 143-159.

44. Dirson, R., Groson, R., Seris, J. Prolyl 4-hydroxylase from *in vitro* cell cultures. *J. Plant Physiol.* **1990**, 136, 444-450.

45. Kubitzki, K. Phenylpropanoid metabolism in relation to land plant origin and diversification. *J. Plant Physiol.* **1987**, 131, 17-24.

46. Swain, T. The evolution of flavonoids. In V. Cody, E. Middleton, Jr., J. B. Harborne, ed.; Alan R. Liss: New York, 1986; p 1-42.

47. Havir, E. A., Hanson, K. R. L-Phenylalanine ammonia-lyase. II. Mechanism and kinetic properties of the enzyme from potato tubers. *Biochemistry* **1968**, 7, 1904-1914.

48. Rupprich, N., Kindl, H. Stilbene synthases and stilbenecarboxylate synthases, I. Enzymatic synthesis of 3,5,4-trihydroxystilbene from *p*-coumaroyl coenzyme A and malonyl coenzyme A. *Hoppe Seylers Z. Physiol. Chem.* **1978**, 359, 165-172.

49. Tropf, S., Lanz, T., Rensing, S. A., 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.

50. Heath, R. J., Rock, C. O. The Claisen condensation in biology. *Nat. Prod. Rep.* **2002**, 19, 581-596.

51. Rock, C. O., Cronan, J. E. *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim. Biophys. Acta.* **1996**, 1302, 1-16.

52. Arnvig, M. K., McGuire, J. N., von Wettstein-Knowles, P. Acyl carrier protein (ACP) inhibition and other differences between beta-ketoacyl synthase (KAS) I and II. *Biochem. Soc. Trans.* **2000**, 28, 607-610.

53. Lai, C. Y., Cronan, J. E. Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J. Biol. Chem.* **2003**, 278, 51494-51503.

54. Mathieu, M., Modis, Y., Zeelen, J. P., Engel, C. K., Abagyan, R. A., Ahlberg, A., Rasmussen, B., Lamzin, V. S., Kuman, W. H., Wierenga, R. K. The 1.8 Å crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. *J. Mol. Biol.* **1997**, 273, 714-728.

55. Schulz, H., Staack, H. 3-Ketoacyl-CoA-thiolase with broad chain length specificity from pig heart muscle. *Methods Enzymol.* **1981**, 71, 398-403.

56. Balasubramaniam, S., Goldstein, J. L., Brown, M. S. Regulation of cholesterol synthesis in rat adrenal gland through coordinate control of 3-hydroxy-3-methylglutaryl coenzyme A synthase and reductase activities. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, 74, 1421-1425.

57. Masamune, S., Palmer, M. A. J., Gamboni, R., Thompson, S., Davis, J. T., Williams, S. F., Peoples, O. P., Sinskey, A. J., Walsh, C. T. Bio-Claisen condensation catalyzed by thiolase from *Zoogloea ramigera*. Active site cysteine residues. *J. Am. Chem. Soc.* **1989**, 111, 1879-1881.

58. Modis, Y., Wierenga, R. K. A biosynthetic thiolase in complex with a reaction intermediate: the crystal structure provides new insights into the catalytic mechanism. *Structure* **1999**, 7, 1279-1290.

59. Modis, Y., Wierenga, R. K. Crystallographic analysis of the reaction pathway of *Zoogloea ramigera* biosynthetic thiolase. *J. Mol. Biol.* **2000**, 297, 1171-1182.

60. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Lindquist, Y., Larsen, S. The X-ray crystal structure of beta-ketoacyl [acyl carrier protein] synthase I. *FEBS Lett.* **1999**, 460, 46-52.

61. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., Lindqvist, Y. Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J.* **1998**, 17, 1183-1191.

62. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Larsen, S. Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **2001**, 9, 233-243.

63. Witkowski, A., Joshi, A. K., Lindqvist, Y., Smith, S. Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry* **1999**, 38, 11643-11650.

64. Devereux, R., Loeblich, A. R., Fox, G. F. Higher plant origins and the phylogeny of green algae. *J. Mol. Evol.* **1990**, 31, 18-24.

65. Hori, H., Lim, B. L., Osawa, S. Evolution of green plants as deduced from 5S rRNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 820-823.

66. Gemmell, A. R. Studies in the Bryophyta. I. The influence of sexual mechanism on varietal production and distribution of British Musci. *New Phytol.* **1950**, 49, 64-71.

67. Farrar, D. R. Gametophytes of four tropical fern genera reproducing independently of their sporophytes in the southern Appalachians. *Science* **1967**, 155, 1266-1267.

68. Farrar, D. R. Species and evolution in asexually reproducing independent fern gametophytes. *Syst. Bot.* **1990**, 15, 98-111.

69. Chapman, R. L., Buchheim, M. A. Green algae and the evolution of land plants: inferences from nuclear-encoded rRNA gene sequences. *Biosystems* **1992**, 28, 127-137.

70. Cove, D. J., Kammerer, W., Knight, C. D., Leech, M. J., Martin, C. R., Wang, T. L. Developmental genetic studies of the moss, *Physcomitrella patens*. *Symp. Soc. Exp. Biol.* **1991**, 45, 31-43.

71. Haig, D., Wilczek, A., Sexual conflict and the alternation of haploid and diploid generations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2006**, 361, 335-343.

72. Schaefer, D. G., Zryd, J. P. Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* **1997**, 11, 1195-1206.

73. Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., Hasebe, M. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 8007-8012.

74. Lang, D., Eisinger, J., Reski, R., Rensing, S. A. Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism in mosses. *Plant Biol.* **2005**, 7, 238-250.

75. Cove, D. The moss *Physcomitrella patens*. *Annu. Rev. Genet.* **2005**, 39, 339-358.

76. Basile, A., Giordano, S., Lopez-Saez, J. A., Cobianchi, R. C. Antibacterial activity of pure flavonoids isolated from mosses. *Phytochemistry* **1999**, 52, 1479-1482.

77. Yamazaki, Y., Suh, D.-Y., 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.

78. 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.

79. Huang, J.-X., Qu, L.-J., Yang, J., Yin, H., Gu, H.-Y. A preliminary study on the origin and evolution of chalcone synthase (CHS) gene in angiosperms. *Acta Botanica Sinica* **2004**, 46, 10-19.

80. Ashton, N. W., Schulze, A., Hall, P., Bandurski, R. S. Estimation of indole-3-acetic acid in gametophytes of the moss, *Physcomitrella patens*. *Planta* **1985**, 164, 142-144.

81. Nakajima, O., Akiyama, T., Hakamatsuka, T., Shibuya, M., Noguchi, H., Ebizuka, Y., Sankawa, U. Isolation, sequence and bacterial expression of a cDNA for chalcone synthase from the cultured cells of *Pueraria lobata*. *Chem. Pharm. Bull.* **1991**, 39, 1911-1913.

82. Yamaguchi, T., Kurosaki, F., Suh, D.-Y., Sankawa, U., Nishioka, M., Akiyama, T., Shibuya, M., Ebizuka, Y. Cross-reaction of chalcone synthase and stilbene synthase overexpressed in *Escherichia coli*. *FEBS Lett.* **1999**, 460, 457-461.

83. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, 72, 248-254.

84. Wolf, B., Lausarot, P. M., Lesnaw, J. A., Reichmann, M. E. Preparation of polymeric protein markers and an investigation of their behavior in sodium

dodecyl sulfate-polyacrylamide gel electrophoresis. *Biochim. Biophys. Acta* **1970**, 200, 180-183.

85. Stockigt, J., Zenk, M. H. Chemical syntheses and properties of hydroxycinnamoyl-Coenzyme A derivatives. *Z. Naturforsch.* **1975**, 30, 352-358.

86. Thompson, J. D., Higgins, D. G., Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, 22, 4673-4680.

87. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, 25, 4876-4882.

88. Huelsenbeck, J. P., Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, 17, 754-755.

89. Page, R. D. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **1996**, 12, 357-358.

90. Kumar, S., Tamura, K., Nei, M. Intergrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **2004**, 5, 150-163.

91. Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., Scheele, G. A. Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **1987**, 6, 43-48.

92. Sawant, S. V., Kiran, K., Singh, P. K., Tuli, R. Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants. *Plant Physiol.* **2001**, 126, 1630-1636.

93. Rothnie, H. M., Reid, J., Hohn, T. The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. *EMBO J.* **1994**, 13, 2200-2210.

94. Loke, J. C., Stahlberg, E. A., Strenski, D. G., Haas, B. J., Wood, P. C., Li, Q. Q. Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures. *Plant Physiol.* **2005**, 138, 1457-1468.

95. Alexandrov, N. N., Troukhan, M. E., Brover, V. V., Tatarinova, T., Flavell, R. B., Feldmann, K. A. Features of Arabidopsis genes and genome discovered using full-length cDNAs. *Plant Mol. Biol.* **2006**, 60, 69-85.

96. Jez, J. M., Bowman, M. E., Noel, J. P. Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 5319-5324.

97. Liu, B. Y., 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.

98. Abe, I., Takahashi, Y., Lou, W., Noguchi, H. Enzymatic formation of unnatural novel polyketides from alternate starter and nonphysiological extension substrate by chalcone synthase. *Org. Lett.* **2003**, 5, 1277-1280.

99. Samappito, S., Page, J. E., Schmidt, J., De-Eknamkul, W., Kutchan, T. M. Molecular characterization of root-specific chalcone synthases from *Cassia alata*. *Planta* **2002**, 216, 64-71.

100. Ito, M., Ichinose, Y., Kato, H., Shiraishi, T., Yamada, T. Molecular evolution and functional relevance of the chalcone synthase genes of pea. *Mol. Gen. Genet.* **1997**, 255, 28-37.

101. Durbin, M. L., McCaig, B., Clegg, M. T. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol. Biol.* **2000**, 42, 79-92.

102. Matsumura, H., Watanabe, S., Harada, K., Senda, M., Akada, S., Kawasaki, S., Dubouzet, E. G., Minaka, N., Takahashi, R. Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean. *Theor. Appl. Genet.* **2005**, 110, 1203-1209.

103. Durbin, M. L., Jr. Learn, G. H., Huttley, G. A., Clegg, M. T. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 3338-3342.

104. Helariutta, Y., Kotilainen, M., Elomaa, P., Kalkkinen, N., Bremer, K., Teeri, T. H., Albert, V. A. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: Evolution with substrate change and catalytic simplification. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 9033-9038.

105. Yang, J., Huang, J., Gu, H., Zhong, Y., Yang, Z. Duplication and adaptive evolution of the chalcone synthase genes of dendranthema (Asteraceae). *Mol. Biol. Evol.* **2002**, 19, 1752-1759.

106. Oberholzer, V., Durbin, M., Clegg, M. T. Comparative genomics of chalcone synthase and Myb genes in the grass family. *Genes Genet. Syst.* **2000**, 75, 1-16.

107. Haapalainen, A. M., Merilainen, G., Wierenga, R. K. The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem. Sci.* **2006**, 31, 64-71.

108. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C., Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 15670-15675.

109. Schmitt, I., Martin, M. P., Kautz, S., Lumbsch, H. T. Diversity of non-reducing polyketide synthase genes in the Pertusariales (lichenized Ascomycota): A phylogenetic perspective. *Phytochemistry* **2005**, 66, 1241-1253.

110. Jenke-Kodama, H., Sandmann, A., Muller, R., Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **2005**, 22, 2027-2039.

111. Gross, F., Luniak, N., Perlova, O., Gaitatzis, N., Jenke-Kodama, H., Gerth, K., Gottschalk, D., Dittmann, E., Muller, R. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **2006**, 185, 28-38.

112. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L.,

Geoghagen, N. S., Venter, J. C. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannachii*. *Science* **1996**, 273, 1058-1073.

113. Petero, J., Lopez-Garcia, P., Moreira, D. Ancestral lipid biosynthesis and early membrane evolution. *Trend Biochem. Sci.* **2004**, 29, 469-477.

114. Syvanen, M. Molecular clocks and evolutionary relationships: possible distortions due to horizontal gene flow. *J. Mol. Evol.* **1987**, 26, 16-23.

115. Helariutta,Y., Elomaa, P., Kotilainen, M., Griesbach, R. J., Schroder, J., Teeri, T. H. Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (Asteraceae). *Plant Mol. Biol.* **1995**, 28, 47-60.

116. Beckert, C., Horn, C., Schnitzler, J.-P., Lehning, A., Heller, W., Veit, M. Styrylpyrone biosynthesis in *Equisetum arvense*. *Phytochemistry* **1996**, 44, 275-283.

117. Wurzel, G., Becker, H., Eicher, T., Tiefensee, K. Molluscicidal properties of constituents from the liverwort *Ricciocarpos natans* and of synthetic lunularic acid derivatives. *Planta Med.* **1990**, 56, 444-445.

118. Eckermann, C., Schroder, G., Eckermann, S., Strack, D., Schmidt, J., Schneider, B., Schroder, J. Stilbenecarboxylate biosynthesis: a new function in the family of chalcone synthase-related proteins. *Phytochemistry* **2003**, 62, 271-286.

119. Tropf, S., Karcher, B., Schroder, G., Schroder, J. Reaction mechanisms of homodimeric plant polyketide synthase (stilbenes and chalcone synthase). A single active site for the condensing reaction is sufficient for synthesis of

stilbenes, chalcones, and 6'-deoxychalcones. *J. Biol. Chem.* **1995**, 270, 7922-7928.

120. Austin, M. B., Bowman, M. E., Ferrer, J. L., Schroder, J., Noel, J. P. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. *Chem. Biol.* **2004**, 11, 1179-1194.

121. Jez, J. M., Austin, M. B., Ferrer, J., Bowman, M. E., Schroder, J., Noel, J. P. Structural control of polyketide formation in plant-specific polyketide synthases. *Chem. Biol.* **2000**, 7, 919-930.

122. Shomura, Y., Torayama, I., Suh, D. Y., Xiang, T., Kita, A., Sankawa, U., Miki, K. Crystal structure of stilbene synthase from *Arachis hypogaea*. *Proteins* **2005**, 60, 803-806.

123. Harris, T. M., Carney, R. L. Biogenetically modeled synthesis of β-resorcylic acids. *J. Am. Chem. Soc.* **1966**, 88, 2053-2054.

124. Harris, T. M., Carney, R. L. Synthesis of 3,5,7-triketo acids and esters and their cyclizations to resorcinol and phloroglucinol derivatives. Models of biosynthesis of phenolic compounds. *J. Am. Chem. Soc.* **1967**, 89, 6734-6741.

125. Harris, T. M., Harris, C. M. Biomimetic syntheses of aromatic polyketide metabolites. *Pure Appl. Chem.* **1986**, 58, 283-294.

126. Li, J., Derewenda, U., Dauter, Z., Smith, S., Derewenda, Z. S. Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme. *Nat. Struct. Biol.* **2000**, 7, 555-559.

127. Shibuya, M., Nishioka, M., Sankawa, U., Ebizuka, Y. Incorporation of three deuterium atoms excludes intermediacy of stilbenecarboxylic acid in stilbene synthase reaction. *Tetrahedron Lett.* **2002**, 43, 5071-5074.

128. Akiyama, T., Shibuya, M., Liu, H. M., Ebizuka, Y. *p*-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. *Eur. J. Biochem.* **1999**, 263, 834-839.

129. Oguro, S., Akashi, T., Ayabe, S., Noguchi, H., Abe, I. Probing biosynthesis of plant polyketides with synthetic *N*-acetylcysteamine thioesters. *Biochem. Biophys. Res. Commun.* **2004**, 325, 561-567.

130. Hillis, W. E., Ishikura, N. An enzyme from *Eucalyptus* which converts cinnamoyltriacetic acid into pinosylvin. *Phytochemistry* **1969**, 8, 1079-1088.