An Effective Strategy for Exploring Unknown Metabolic Pathways by Genome Mining

Dorianne A. Castillo,[†] Mariya D. Kolesnikova,[†] and Seiichi P. T. Matsuda*,[†,‡]

[†]Department of Chemistry and [‡]Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005, United States

S Supporting Information

ABSTRACT: Plants allocate an estimated 15−25% of their proteome to specialized metabolic pathways that remain largely uncharacterized. Here, we describe a genome mining strategy for exploring such unknown pathways and demonstrate this approach for triterpenoids by functionally characterizing three cytochrome P450s from *Arabidopsis thaliana*. Building on proven methods for characterizing oxidosqualene cyclases, we heterologously expressed in yeast known cyclases with candidate P450s chosen from gene clustering and microarray coexpression patterns. The yeast cultures produced mg/L amounts of plant metabolites in vivo without the complex phytochemical background of plant extracts. Despite this simplification, the product multiplicity and novelty overwhelmed analytical efforts by MS methods. HSQC analysis overcame this problem. Side-by-side HSQC comparisons of crude P450 extracts against a control resolved even minor P450 products among ∼100 other yeast metabolites spanning a dynamic range of >10 000:1. HSQC and GC−MS then jointly guided purification and structure determination by classical NMR methods. Including our present results for P450 oxidation of thalianol, arabidiol, and marneral, the metabolic fate for most of the major triterpene synthase products in *Arabidopsis* is now at least partially known.
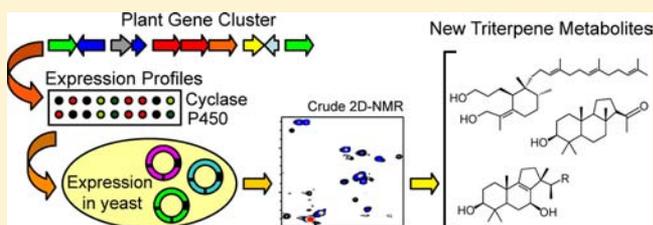
INTRODUCTION

Natural selection constructs in each plant species a versatile genetic arsenal for biosynthesizing specialized (secondary) metabolites that are accessible for defense and other needs of sessile organisms.[1−3] These natural products are collectively much greater in number than primary metabolites and structurally more diverse. Biosynthetic pathways of primary metabolism are well established, whereas the numerous pathways leading to specialized metabolites in plants are mostly unknown.[1]

Cytochrome P450s play a major role in specialized metabolism. The *Arabidopsis* genome contains ∼245 cytochrome P450s, of which only ∼20% have been functionally characterized.[4,5] Genomics can reliably identify most P450s of primary metabolism by homology because these genes are well conserved across families. However, elucidating specialized metabolic pathways is more difficult, and few of the estimated 200 *Arabidopsis* P450s of specialized metabolism have been characterized.[6] These few studies generally start from a known natural product as a basis for identifying the P450s. This approach depends on knowledge of the specialized metabolite, most of which are yet unidentified.[7] Filling this gap in metabolomics and functional genomics is the topic of this Article.

A large branch of specialized metabolism is terpenoid biosynthesis.[3] A few linear terpenes (C_{10}, C_{15}, C_{20}, C_{30}, and C_{40}) undergo cyclization, P450 oxidation, and other enzymatic reactions to make >40 000 terpenoids.[8,9] Within the C_{30} subset, oxidosqualene is cyclized to >100 triterpene skeletons,[10] which

are converted to >20 000 triterpenoids.[11,12] Genome mining[13] has been enormously successful in functionally characterizing oxidosqualene cyclases (OSCs) by heterologous expression.[14] In *Arabidopsis*, this approach[15,16] has identified far more OSC products (triterpenes) than have been found by direct analysis of plant material.[7b,17,18]

Extending genome mining to the subsequent metabolism of triterpenes to triterpenoids was envisioned a decade ago.[19,20] Since then, heterologous expression has been used to study glycyrrhizin biosynthesis[21] and other oxidative pathways[22] of pentacyclic triterpenes to known products that were identified mainly by GC−MS or LC−MS comparisons with authentic standards. Others[23,24] tackled the difficult problem of elucidating unknown pathways for P450 oxidation of two *Arabidopsis* triterpenes, thalianol and marneral. Their earlier discovery of coregulated gene clusters in plants[25] facilitated judicious choices of candidate P450s clustered with the thalianol and marnerol synthases. Gene overexpression demonstrated the P450 oxidations in planta. However, GC−MS gave only partial chemical structures of the novel products, a result underscoring the reliance of MS-based methods[26] on databases and authentic standards. Even with standards for the few known oxygenated derivatives of these unusual triterpene skeletons,[27−29] GC−MS would be challenged to identify the regio- and stereochemistry of oxidation unambiguously.

Building on this previous work and our experience in functionally characterizing OSCs,[13,15,16,30−33] we used heterologous expression and 2D NMR to develop an effective strategy for exploring unknown pathways. Relative to experiments in planta, heterologous expression[34] eliminated the background interference of extraneous phytochemicals and supplied ample amounts of P450 products for NMR analysis. HSQC then became a powerful tool for analyzing crude extracts, which often contain a variety of triterpenoids due to the product multiplicity[16,31,32] that is common in specialized metabolism. Heterologous expression and HSQC thus overcame major hurdles in exploring unknown metabolic pathways.

Here, we apply this genome mining strategy to study the enzymatic oxidation of three *Arabidopsis* OSC products: thalianol, arabidiol, and marneral. Each OSC was heterologously expressed with a coregulated *Arabidopsis* P450, which was functionally characterized by identifying the major oxidation products using GC−MS and 2D NMR. Our experimental design is directly applicable to the elucidation of other unknown P450 pathways of triterpenoid metabolism, where the novel isomeric products might confound the high-throughput methods of metabolomics[17b] and dereplication.[35] This general approach and its variations[36,37] may be useful in exploring how plants and other organisms creatively exploit the powerful genetic machinery available for specialized metabolism.[38]

■ RESULTS

Experimental Design. We studied the P450 oxidation of three triterpenes (thalianol, arabidiol, and marneral) by heterologous expression in vivo in yeast strains coexpressing the OSC, the candidate P450, and its redox partner ATR2 (an *Arabidopsis* NADPH-cytochrome P450 reductase). The yeast strains efficiently produced the foreign triterpene substrates in vivo, thus avoiding the complications of performing in vitro experiments. The resulting P450 metabolites were analyzed by GC−MS and 2D NMR.

More specifically, we selected candidate P450s based on gene cluster analysis[39] and coexpression[40] relative to each triterpene OSC (Figures S2 and S3), as done previously.[23,24] Recombinant yeast strains were constructed by transforming the JBY575 strain[41] with an OSC, a P450, and ATR2. Three analogous control strains were made without the *Arabidopsis* P450. The OSCs were available from previous work.[13,30,32] Two P450s (CYP705A1 and CYP71A16) were obtained as plasmids from resource centers. Another P450 (CYP708A2) and ATR2 were synthesized commercially with codon optimization. Determination of the correct protein sequence for ATR2 is described in text accompanying the sequence alignment shown in Figure S1. The three yeast strains and their controls were grown in liquid cultures. The putative P450 metabolites were obtained from cells as a crude ethanol extract comprising triterpenoids, triterpenes, yeast sterols, and other lipids.

An efficient assay for P450 metabolites in the crude extracts was needed for the many trial cultures for each of the several yeast strains. However, recognizing the metabolites as minor components in a complex mixture was challenging because authentic standards and spectral data were generally unavailable. Following others[23] in using a dereplication approach, we initially resorted to spectral comparisons against the substrate. For anticipated hydroxylations, GC−MS of trimethylsilyl (TMS) ethers showed peaks with the expected molecular ion but gave limited clues about the location and stereochemistry of

oxidation. These structural ambiguities would not likely be resolved by MS methods, including high-resolution LC−MS/MS.

We then turned to 2D NMR for mixture analysis.[42] Unlike 1D NMR, HSQC of the crude yeast extracts resolved many triterpenoid methyl signals. Often these and other correlated ¹H and ¹³C chemical shifts revealed structures of novel triterpenoid products prior to purification. Moreover, comparing HSQC spectra of crude P450 yeast extracts against the corresponding control distinguished P450 metabolite signals from the numerous unidentified peaks. These facile side-by-side or overlaid comparisons in Topspin software quickly identified even many minor (1%) metabolites in the crude extracts. GC−MS and NMR then guided chromatographic purification for classical structure determination.

7β-Hydroxythalianol Is the Primary CYP708A2 Metabolite of Thalianol. CYP708A2 is known to metabolize thalianol.[23] To determine the location and configuration of this oxidation, we studied thalianol oxidation in JBY575 yeast transformed with thalianol synthase (*THAS1, At5g48010*), CYP708A2 (*At5g48000*), and ATR2 (Scheme 1). GC−MS

Scheme 1. Oxidation of Thalianol by CYP708A2 To Form 3 (Solid Arrow), Accompanied by Oxidation to 4 Either Directly by CYP708A2 or via 5 by a Yeast P450 (Dashed Arrows)
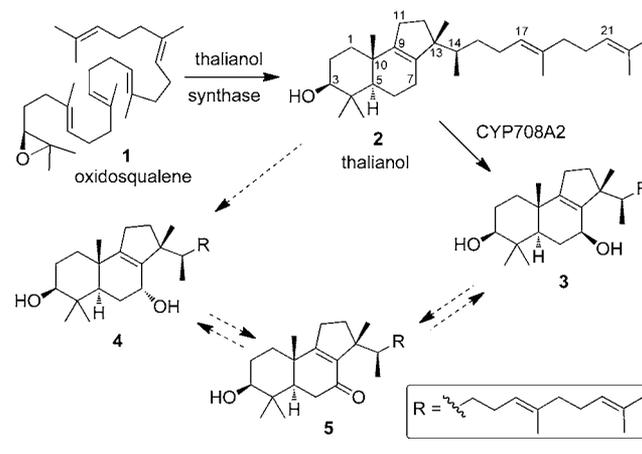
analysis of the crude product as TMS ethers (Figure 1A) showed peaks for thalianol (**2**; *m/z* 498) and two putative hydroxythalianols (**3** and **4**; *m/z* 586) that were absent in a control culture lacking CYP708A2. Comparison of the MS fragmentation of **3** and **4** versus **2** suggested that hydroxylation occurred in ring B or C, but the exact position was ambiguous. A previous in planta study of the same CYP708A2 oxidation[23] reached a similar conclusion based on GC−MS spectra resembling those of **3** and **4**. Reversed-phase HPLC (Figure 1B) provided purified samples of **3** and **4** that were analyzed by NMR (Figure 1C). 2D NMR correlations (Figure 2) established the location and configuration of hydroxylation for **3** and **4** as 7β and 7α, respectively. The minor product 7-ketothalianol (**5**) was partially purified by HPLC, with structure determination by 2D NMR. Additionally, one or more of some very minor triterpenoids observed by GC−MS might have originated from enzymatic oxidation of thalianol (Figures S5,S6).

Did CYP708A2 generate a mixture of **3**−**5** or a single 7-oxygenated product that was further metabolized by one of the
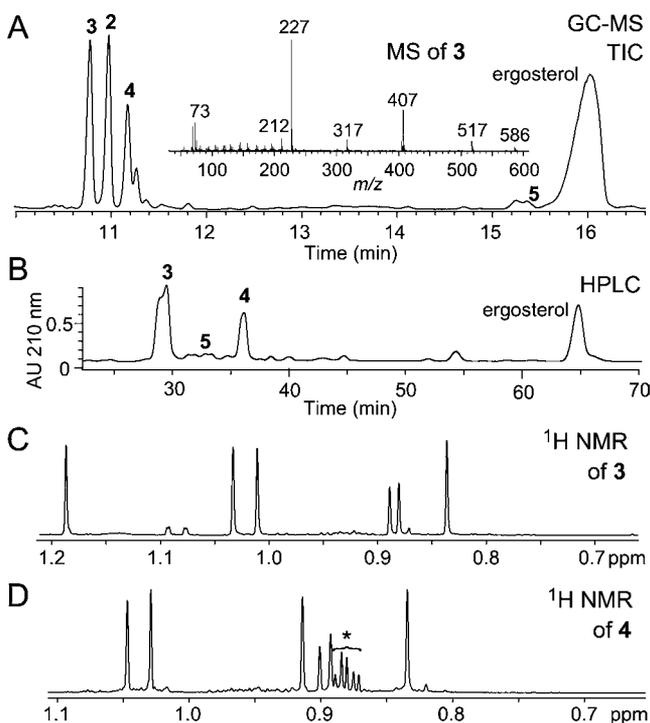
Figure 1. Oxidation of thalianol by CYP708A2. (A) GC−MS of the crude extract as the total ion chromatogram (TIC) and MS of **3**. (B) HPLC purification separated **3** and **4**. (C) Homogeneity of purified **3** and **4** (except for fatty acids, marked by *) as judged by ^{1}H NMR.
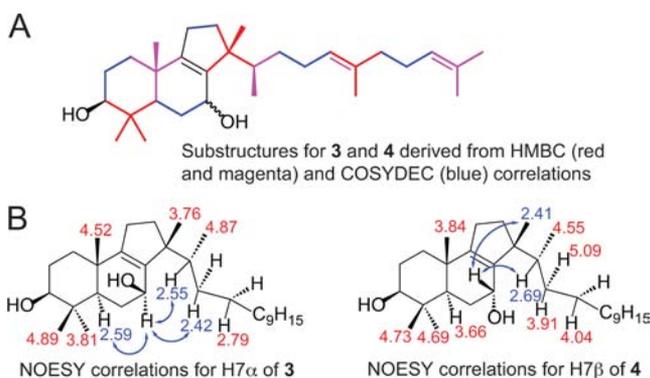
Figure 2. Classical 2D NMR confirmed the structure of **3** and **4**. (A) HMBC and COSYDEC correlations established hydroxylation at C7. (B) NOESY correlations established the C7 configurations. Distances from H7 to remote (red) and nearby (blue) hydrogen atoms are shown for the major side-chain conformation (B3LYP/6-31G* geometry).

seven *S. cerevisiae* P450s? To clarify this ambiguity, which is often encountered when P450s are heterologously expressed,[43] we took aliquots of a culture at ca. 11 h intervals and measured the relative amounts of **3** and **4** by GC−MS (Table S1). The ratio of **3**:**4** was ~10:1 at early time points and decreased to 2:1 at saturation. This trend is consistent with initial P450 oxidation to **3**, which is then metabolized further by oxidation to **5** and reduction to **4** by CYP708A2 and/or yeast enzymes.

CYP705A1 Cleaves the Arabidiol Side Chain to a C_{19} Product. Metabolism of arabidiol (**6**) was studied analogously in cultures of JBY575 expressing arabidiol synthase (*PEN1, At4g15340*), CYP705A1 (*At4g15330*), and *ATR2*. Flash chromatography of the saponified ethanol extract gave fractions

A−H, which were analyzed by GC−MS as TMS ethers. GC−MS of the main polar fraction (E) (Figures 3A and S9)
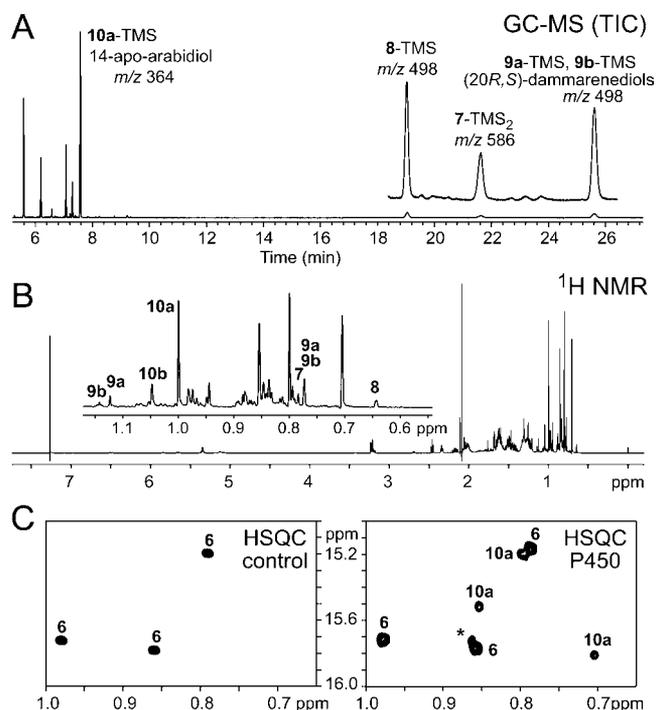
Figure 3. Oxidation of arabidiol by CYP705A1. (A) GC−MS and (B) 800 MHz ^{1}H NMR of chromatographic fraction E (saponified P450 metabolites; **7** and **8** proved to be families of isomers). (C) HSQC (600 MHz) comparison of control and P450 spectra of the unsaponified crude extracts; the asterisk denotes a minor P450 metabolite.

suggested side-chain oxidation of arabidiol to a secondary alcohol (**7**; m/z 586) and a nonisomeric tertiary alcohol (**8**; m/z 498) derivative. The third peak in the triterpenoid region (t_R 12−27 min) represented coeluting mono-TMS ethers of (20*R*)- and (20*S*)-dammarenediols (**9a** and **9b**), which are PEN1 cyclization byproducts rather than P450 metabolites. Had we relied on GC−MS or LC−MS analysis, we would have annotated CYP705A1 as a hydroxylase that makes side-chain oxygenated derivatives of arabidiol.

Anticipating a mixture containing comparable amounts of **7**, **8**, and **9a**/**9b**, we proceeded with NMR analysis. Unexpectedly, 1D NMR (Figure 3B) showed one major component with only four methyl singlets and none of the usual side-chain resonances of tricyclic triterpenoids. 2D NMR revealed that the side chain had been cleaved at the C14−C15 bond, leaving a C_{19} fragment (Scheme 2).[44] We informally denote this 13*R* C_{19} methyl ketone (**10a**) as 14-apo-arabidiol (borrowing from carotenoid nomenclature). SciFinder substructure searching uncovered no remotely similar compounds except for 14-hydroxy analogues of nonbiological origin with different stereochemistry in the ABC ring system.[45]

Among the minor products in the saponified mixture was the 13*S* epimer **10b**, which represents partial epimerization during saponification.[46] Subsequent experiments confirmed that only **10a** is formed when workup excluded saponification. Side-by-side HSQC comparison of crude extracts from the CYP705A1 construct against a control lacking the P450 (Figure 3C) showed **10a**, **7**, and **8** only in the CYP705A1 extracts, whereas

Scheme 2. Cleavage of the Arabidiol Side Chain To Form 10a
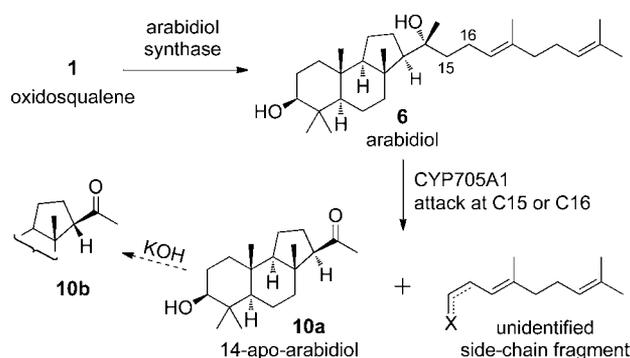
both cultures contained arabidiol and at lower levels dammarenediols, other PEN1 byproducts, and yeast sterols.

Why did we initially overlook 14-apo-arabidiol in the GC–MS results? Not anticipating a C_{19} product, we did not exhaustively investigate the early fatty-acid region of the chromatogram. Even after the NMR identification of **10a** and **10b**, locating their GC–MS peaks was not trivial. Both epimers were present as mixtures of mono- and bis-TMS ethers, and their MS fragmentation showed only weak ions for losses of CH_3 and TMSOH. Their abundant low-mass ions (Figure S8) resembled mass spectra of early eluting lipids commonly observed in yeast extracts more than ions of tricyclic triterpenoids (Figures S6,S7). After this humbling experience, we routinely surveyed P450 product profiles by analyzing crude extracts using both HSQC and GC–MS.

CYP71A16 Hydroxylates Marneral Selectively at C23. Cultures of JBY575 transformed with marneral synthase (*MRN1*, *At5g42600*), CYP71A16 (*MRNO*, *At5g42590*), and *ATR2* (*At4g30210*) were extracted, chromatographed, and analyzed spectrally as in the CYP708A2 and CYP705A1 studies. In this case, the mixture of P450 metabolites, unoxidized triterpenes, yeast sterols, and other lipids was further complicated by the presence of the OSC product as a combination of aldehyde marneral (**11a**) and alcohol marnerol (**11b**) (Scheme 3). This duality, which has been noted in both

Scheme 3. Marneral Hydroxylation by CYP71A16
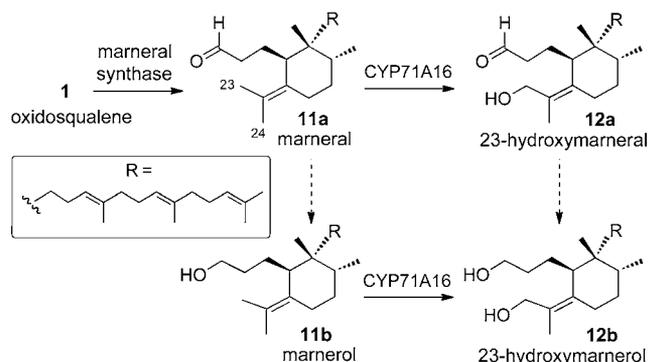
plant[24,47] and yeast[24,30] extracts, extended to the P450 metabolites (**12a** and **12b**). The aldehyde reduction might arise from yeast metabolism and/or workup with or without saponification. Another analytical complication was TMS derivatization, which converted each aldehyde moiety into a

mixture of three species, the underivatized aldehyde and the *E* and *Z* TMS enol ethers.

GC–MS and HSQC were used to compare the triterpenoid profile in CYP71A16 cultures and control cultures lacking the P450. GC–MS (Figure 4A) showed TMS derivatives of a
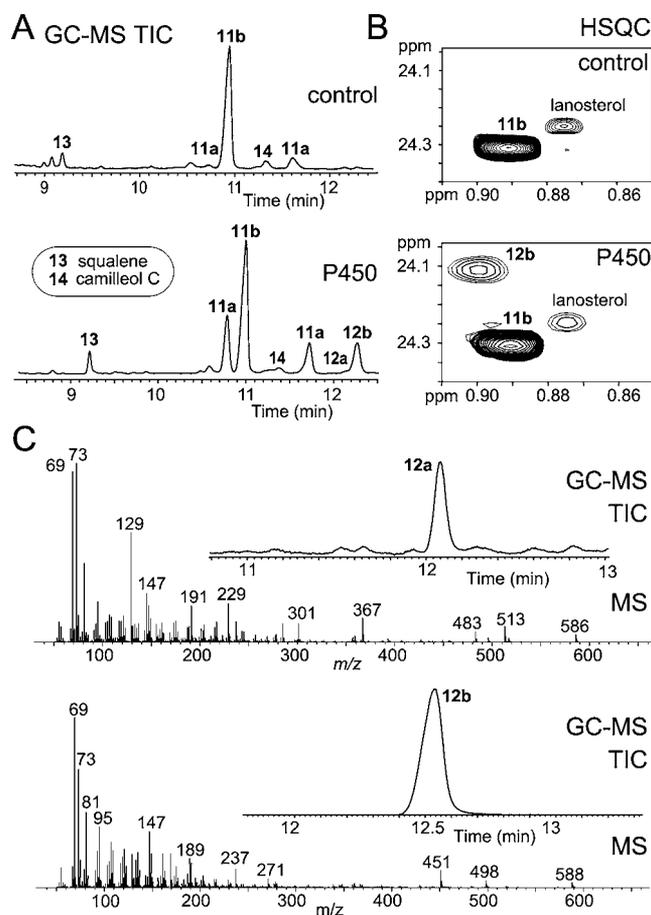
Figure 4. Hydroxylation of marneral at C23 by CYP71A16. (A) GC–MS and (B) HSQC comparison of crude extracts from cultures with ("P450") and without ("control") CYP71A16. The second GC–MS peak for **11a** corresponds to a TMS enol ether of the aldehyde. (C) GC–MS of the CYP71A16 products **12a** and **12b** after HPLC purification.

hydroxymarnerol product (**12b**; m/z 588) and a small hydroxymarneral peak (**12a**) that were unique to the P450 culture. Both cultures contained marnerol (**11b**; m/z 500) and marneral (**11a**; m/z 498 underivatized). HSQC, which avoided the complications of TMS derivatization, cleanly showed the same pattern (Figure 4B) and indicated hydroxylation at C23 or C24. 1D NMR was used to detect and identify trace amounts of 3- and 23-aldehyde species, which gave distinctive chemical shifts in the 9.5–10.5 ppm region.

Flash chromatography on silica gel and reversed-phase HPLC furnished samples of **12a** and **12b** (Figure 4C). 2D NMR showed the hydroxylation to be at the *Z* allylic methyl (C23). ^{1}H NMR data reported[28] for **12b** were consistent with our NMR results but insufficient for unambiguous structure identification due to the modest chemical-shift precision (±0.01 ppm) and lack of ^{13}C NMR data. Very minor unidentified side-chain oxygenated marnerol derivatives, possibly CYP71A16 metabolites, were observed in GC–MS and HSQC spectra (Figures S10,S11). We detected only traces

(<0.1%) of 23-aldehyde species, such as the marnerol derivative, for which NMR data have been reported.[28] Thus, CYP71A16 hydroxylates marnerol/marneral in high accuracy specifically at C23.

HSQC Gives a Comprehensive Profile of Substances in Crude Extracts. Although 2D NMR is used mainly for structure determination of purified compounds, we found HSQC to be valuable for analyzing crude extracts from yeast cultures (Figures 3C and 4B). These figures showed only the upper intensity levels for ~0.1% of the total HSQC spectral area. Digging deeper and more broadly into a similar HSQC spectrum (Figure 5) revealed a multitude of minor P450 metabolites, OSC triterpene byproducts, yeast sterols, and other lipids. Some spectral regions were degraded by artifacts from major peaks (sinc ripples from FID truncation, peak shoulders from imperfect shimming, and sidebands from adiabatic decoupling), but >95% of the spectral area showed peaks spanning a dynamic range of >17 000:1 (Figure 5E). This typical HSQC of a crude yeast extract contained thousands of peaks representing ~100 compounds ranging from <1 μg to >1 mg.

The expanded HSQC snapshots in Figure 5A and B illustrate the resolution of many upfield methyl signals of minor sterols and triterpenoids. Comparing their correlated ^{1}H and ^{13}C chemical shifts with literature data can reveal the location of double bonds and functional groups in the tricyclic ring system and nearby side chain. The H-5α signals (Figure 5D) gave complementary structural insights. HMBC and COSYDEC spectra can augment the HSQC results but are more appropriate for simpler mixtures due to their greater signal overlap and lower sensitivity.

Figure 5 illustrates the power and versatility of HSQC analyses. HSQC can be used classically with other 2D NMR methods for determining the structure of unknown substances. The samples need not be pure. For example, the structure determination of **10a** and **10b** was performed on a 4:1 mixture containing many minor triterpene metabolites (Figures 3B and S20–S22). HSQC can also be used to identify known compounds by matching pairs of ^{1}H–^{13}C correlated chemical shifts with literature data. This identification is highly reliable when a good match (±0.002 ppm for ^{1}H, ±0.02 ppm for ^{13}C) is obtained for distinctive signals representing C–H pairs that span the triterpenoid structure. The sterols and triterpenes in Figure 5E were identified by this method. The utility of HSQC for structure determination, quantification, and compound identification in complex mixtures is well-known.[42]

Figure 5. HSQC spectrum of crude extracts from yeast expressing PEN1, ATR2, CYP705A1, and CYP705A2. The latter P450 generates little or no additional arabidiol oxidation products, as judged by comparison with the HSQC spectrum of Figure 4B. (A and B) Expansions showing methyl singlets for major metabolites (**6**, **10a**), unidentified minor CYP705A1 metabolites (E1–E6, F1–F2, and G1), PEN1 cyclization byproducts (B1–B5), and yeast sterols (B6–B7). Peak labels correspond to flash chromatography fractions B–G; P450 metabolites are underlined. (C) Overview of the upfield methyl region. (D) Expansion showing the H-5α signals (doublets in F$_2$, actual chemical-shift scale in F$_1$). (E) Positive HSQC peak intensities span a dynamic range of >17 000:1. Negative peak intensities cover a similar dynamic range. Identity of peaks B1–B7: B1, 14-epithalianol; B2, (13S)-malabarica-14(27),17,21-trienol; B3, (13S)-malabarica-14E,17,21-trienol; B4, 13R epimer of B2; B5, 13R epimer of B3; B6, lanosterol; B7, 4,4-dimethylcholesta-8,24-dienol.

■ DISCUSSION

This work demonstrates the power of combining genome mining, heterologous expression, and HSQC analysis to elucidate unknown biosynthetic pathways and identify their novel products. We illustrate this strategy by functionally characterizing triterpene oxidation by three *Arabidopsis* P450s and rigorously determining the structures of the triterpenoid metabolites. We show that CYP708A2 oxidizes thalianol to 7β-hydroxythalianol (**3**), that CYP705A1 cleaves the side chain of arabidiol to give a C$_{19}$ methyl ketone (**10a**), and that CYP71A16 hydroxylates an allylic methyl of marneral/marnerol to 23-hydroxymarneral/23-hydroxymarnerol (**12a/12b**). The three P450 oxidations illustrate diverse outcomes: hydroxylation of an allylic CH$_2$ in the ring system, side-chain cleavage resembling mammalian pregnenolone synthesis, and hydroxylation of an allylic methyl. Side-chain hydroxylation and desaturation are other potential outcomes that were not observed as significant products.

The hydroxylation of marneral/marnerol at C23 by CYP71A16 resembles marneral/marnerol metabolism in the iris family.[29] Whereas the iris pathway leads to fragrant irones, the pathway in *Arabidopsis* obviously does not. The seeming coincidence of 23-hydroxylation in these distant taxa merely reflects the broad distribution of P450 families in angiosperms[48] and the limited number of triterpene skeletons and common oxidation sites. *Arabidopsis* and iris evidently recruited and modified a CYP71 gene for the same commonplace

hydroxylation of the same triterpene skeleton. Such convergent evolution is unsurprising in specialized metabolism.[1,49]

The side-chain cleavage of arabidiol is uncommon among P450 oxidations.[4a,50] The mechanism could resemble that of pregnenolone synthesis from cholesterol,[51] which involves sequential P450 hydroxylations at C22 and C20, followed by cleavage with loss of isocaproaldehyde. Another class of cleavage mechanisms (Figure S4) is exemplified by the P450 oxidation of nerolidol to 4,8-dimethyl-1,3,7-nonatriene (DMNT) and presumably but-1-en-3-one.[52] DMNT and its C_{16} homosesquiterpene analogue TMTT are observed in *Arabidopsis* leaf[6a] and root.[52] TMTT is synthesized in leaf from (*E,E*)-geranyllinalool by CYP82G1 oxidative cleavage, which can also efficiently convert nerolidol to DMNT.[6a] Interestingly, the arabidiol side chain is an exact substructure of nerolidol. Moreover, oxidative cleavage of nerolidol by CYP82G1 and of arabidiol by CYP705A1 occurs at identical positions, removing the same C_{11} substructure and (presumably) converting the tertiary alcohols to methyl ketones. CYP82G1 and CYP705A1 are in the same CYP71 clan. If DMNT is cleaved from arabidiol, this would represent an interesting variant of convergent evolution in specialized plant metabolism.[1,49] Is the target arabidiol metabolite for *Arabidopsis* 14-apo-arabidiol or the unidentified side-chain fragment (perhaps DMNT)?

Several recurrent themes pervade this work: the greater difficulty of exploring unknown versus partially known metabolic pathways; the power of heterologous expression in vivo to generate large amounts of metabolite free of a complex phytochemical background (tempered by the potential for generating artifacts of yeast metabolism); the formation of minor enzymatic products, as commonly observed in specialized metabolism; the challenge of detecting the often novel metabolites during product isolation; the importance of flexibility in selecting isolation methods ranging from chemical transformation to the many variants of solvent extraction, with or without saponification; the need for analytical methods that are not dependent on authentic standards; and the value of using MS and HSQC analysis concurrently at all stages of purification.

HSQC analysis of crude mixtures proved to be a powerful tool for exploring unknown pathways. With the ability to disperse signals in two dimensions over an intensity range of >10 000:1, HSQC resolved many minor triterpenoids that were obscured by peak overlap in GC−MS. Differential 2D NMR analyses[42f] were especially useful in comparing crude extracts of cultures with and without a candidate P450 gene. Side-by-side HSQC comparisons in Topspin software smoothly identified metabolite peaks absent in the control spectrum. Structural insights were gleaned from comparisons against the known chemical shifts of the triterpene substrate, which are reproducible to ±0.001 ppm for ^{1}H and ±0.01 ppm for ^{13}C NMR.[16,18,31−33]

Our results provide the first experimental support for gene clustering proposed for arabidiol metabolism[4a,23,39b] and add to existing support[23,24] for thalianol and marneral clusters. Can these clusters be exploited in silico to predict the step order and length of an entire metabolic pathway, such as helvolic acid biosynthesis in fungi?[33] Despite strong evidence from gene coexpression, gene clustering, parallel observations in planta,[23] and reaction efficiency with high regio- and stereoselectivity in recombinant yeast, might different P450s with compatible spatial and temporal expression perform the triterpene oxidations in planta? Simple answers may be elusive

considering the promiscuity of specialized metabolism.[53] Even in well-established pathways of primary mammalian metabolism, such ambiguities can be difficult to resolve.[54] Identification of final metabolic products and their biological roles would provide a needed framework for evaluating the many divergent experimental results and determining how closely the pathways correlate to gene clusters in *Arabidopsis*.

The recent completion of functional characterization for the 13 *Arabidopsis* OSCs[16] indicates 13 triterpenes as major OSC products (Figure 6): cycloartenol (**13**) derived from CAS1 cyclization, lupanediol (**14**) and lupeol (**17**) from LUP1, β-amyrin (**16**) from LUP2 and LUP4, camelliol C (**18**) from LUP3, tirucall-7,24-dienol (**19**) and isotirucallol (**20**) from LUP5, arabidiol (**6**) from PEN1, baruol (**21**) from PEN2, **19** again from PEN3, thalianol (**2**) from PEN4, marneral (**11a**)
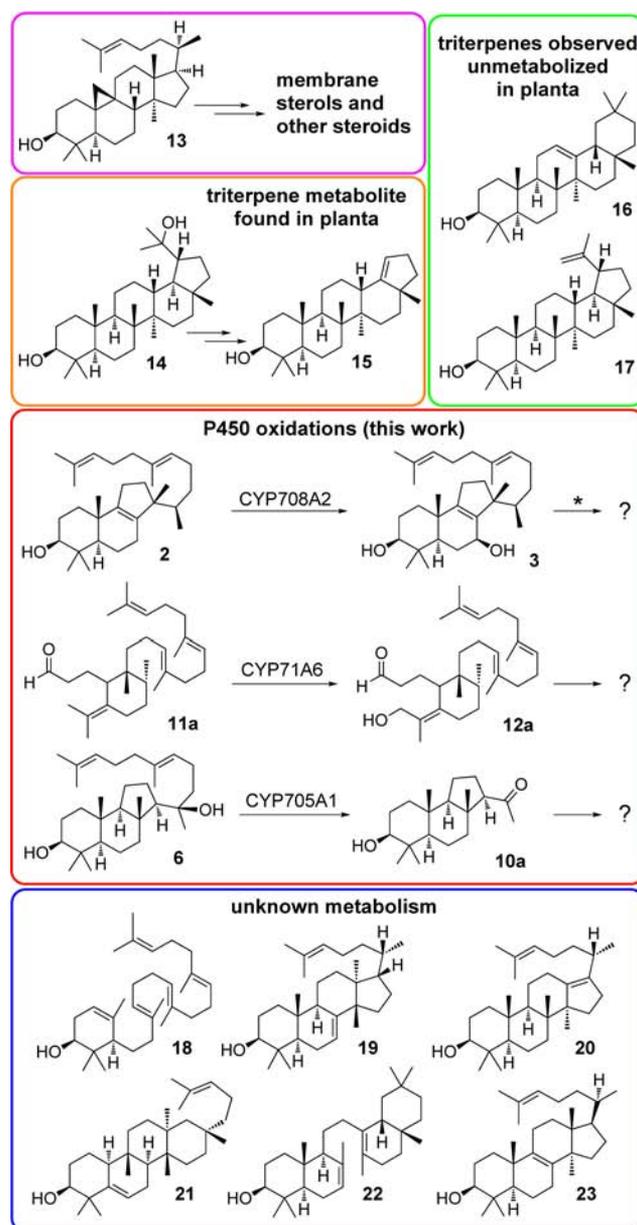
Figure 6. Currently known metabolic fate of the primary products of triterpene synthases (OSCs) in *Arabidopsis*. *CYP705A5 appears to metabolize the product of CYP708A2[23] (i.e., **3**) and is separately reported to be involved in flavonoid biosynthesis.[55]

from PEN5, *seco-β*-amyrin (**22**) from PEN6, and lanosterol (**23**) from LSS1. Cycloartenol is further metabolized to plant membrane sterols and lupanediol to trinorlupeol (**15**).[18] *β*-Amyrin and lupeol are found unmodified in stem wax, silique, and bud.[18] With our results for P450 oxidation of thalianol, arabidiol, and marneral/marnerol, the metabolic fate of most of the primary products of *Arabidopsis* OSCs is now at least partially known. This emerging picture is consistent with observations[7b,17a,18] suggesting that *Arabidopsis* does not oxidize the nonsteroidal triterpenes that it produces constitutively.

■ CONCLUSIONS

A substantial portion of plant genomes is dedicated to specialized metabolic pathways that are taxonomically narrow, fast evolving,[56] tissue specific,[1] catalytically promiscuous,[53] and often silent.[38] These pathways contain a wealth of information about selective pressures, defense mechanisms, P450 function, and plant metabolomes. However, specialized pathways are barely explored. Genomic analysis usually[57] fails because the narrowly distributed pathways lack true orthologs. High-throughput MS-based metabolomics often[57] struggles to identify novel specialized metabolites, whose structural diversity can defy the largest MS databases and collections of authentic standards. Traditional natural products isolation[42a] can solve structures but overlooks silent metabolism and rarely connects chemical structures with systems biology.[58,59] We demonstrate an alternative strategy for exploring specialized pathways that incorporates elements of genomics (to optimize heterologous expression), metabolomics (for complementary[60] HSQC and MS dereplication of crude extracts), and natural products methods (for isolation and final structure elucidation). This strategy was successful in all three systems that were studied. Continued efforts should ultimately provide a comprehensive accounting of *Arabidopsis* triterpenoids. Applied more generally to exploring unknown metabolic pathways, this strategy may begin to fill a large knowledge gap in metabolomics and functional genomics.

■ EXPERIMENTAL SECTION

Subcloning and Yeast Transformation. *CYP705A1* and *CYP71A16* were obtained from the Riken Bioresource Center and the *Arabidopsis* Biological Resource Center (ABRC), respectively. *CYP708A2* and *ATR2* were synthesized and codon optimized for yeast by GenScript. These genes were subcloned into yeast expression vectors pRS424Gal, pRS423Gal, or pRS426Gal. Construction of similar plasmids has been described for thalianol,[13] arabidiol,[32] and marneral[30] synthases (OSCs). Appropriate sets of OSC, P450, and ATR2 plasmids were then used sequentially to transform *Saccharomyces cerevisiae* strain JBY575 (*MATa ura3−52 trpl-A63 leu2−3,112 his3-A200 ade2 Gal+*)[41] by the lithium acetate method. The resulting transformants (JBY575[pGCF4.0, pDCR5.1, pDCR4.1], JBY575-[pMDK3.6, pDCR9.3, pDCR4.1], and JBY575[pXQ11.2, pDCR8.3, pDCR4.1]) were used to study P450 metabolism of thalianol, arabidiol, and marneral heterologously in vivo. Analogous control strains lacking the foreign P450 were prepared similarly. Specific details for constructing each transformant are given in the Supporting Information.

Yeast Culture and Product Isolation. Each JBY575 transformant was selected on synthetic complete medium lacking uracil, tryptophan, and histidine, supplemented with 2% glucose, and solidified with 1.5% agar. Cells from the resulting colonies were grown in 10 mL liquid cultures in the above medium, followed by inoculation into 100 mL batches of inducing medium (the same medium with galactose in place of glucose). At saturation the cells were harvested by centrifugation

and extracted with ethanol aided by sonication or bead beating. After further centrifugation to remove cell debris, the ethanol fraction was collected by decanting and evaporated to a residue that was partitioned between water and ether. The ether layers were evaporated to a "crude extract", which was analyzed by GC−MS and HSQC for triterpenoid content. The extract was then purified chromatographically, with characterization of individual fractions by GC−MS and NMR. In an alternative isolation method used occasionally, the cell pellet was extracted by a modified Folch method (2:1 CH_2Cl_2−methanol), followed by partitioning of the evaporated residue between MTBE and water, with or without saponification (10% KOH in 80:20 ethanol−water at 70 °C for 1 h). Control cultures lacking the P450 plasmid were grown under conditions and workup procedures identical to those used for the corresponding P450 strain.

All cultures were grown at 30 °C with shaking at 250 rpm in Erlenmeyer flasks loosely capped with aluminum foil. The experiments were repeated many times at different scales; 100 mL cultures in 250 mL Erlenmeyer flasks gave consistently good production of oxygenated triterpenoids and yeast (1.5−2 g cells/100 mL). Larger cultures (e.g., 250 mL in 1 L flasks) gave lower amounts of oxygenated products relative to OSC products, perhaps because of insufficient aeration.

Chromatographic Purification. The "crude" or Folch extracts were subjected to flash chromatography using Strata solid-phase extraction (SPE) cartridges (1, 2, or 5 g of silica gel; 55 μm particle size, 70 Å pore size) from Phenomenex (Torrance, CA) in a standard SPE vacuum manifold. This purification was occasionally supplemented by preparative TLC on silica gel 60 glass-backed plates (250 μm layer). Often further purification was done by preparative HPLC on an Agilent 1100 system using UV detection at 210 nm, a 250 mm × 21.2 mm C_{18} column, and a linear gradient of 85−100% methanol in water at ca. 23 °C.

GC−MS Analysis. GC−MS was done on an Agilent 6890 GC interfaced to a 5973N MS detector and containing a Restek 30 m × 0.25 mm i.d. Rxi-35Sil column (35% phenyl−65% methylpolysiloxane, 0.25 μm film thickness). Triterpenoid samples were derivatized as TMS ethers by heating 1−5% of the sample in 60 μL of 2:1 pyridine−bis(trimethylsilyl)trifluoroacetamide for 1 h. After pulsed splitless injection of this mixture (2 μL) at 280 °C, the oven temperature was held at 110 °C for 1 min, increased at 30 °C/min to 270 °C, and then increased at 0.5 °C/min to 280 °C (held 8 min). The helium carrier gas was maintained at a constant flow of 2.7 mL/min. Electron-impact MS data were collected in full-scan mode (*m/z* 50−650) and analyzed with Agilent Chemstation software.

NMR Analysis. NMR spectra were acquired at 25 °C in $CDCl_3$ solution on 600−800 MHz instruments equipped with a cryogenic probe. Referencing was to internal tetramethylsilane. Spectra were processed with Topspin 3 software. 2D NMR spectra were obtained with standard pulse sequences, usually with echo-antiecho phase encoding, shaped pulses, and adiabatic decoupling (HSQC). HSQC, HMBC, COSYDEC (F_1-decoupled COSY), and NOESY spectra were acquired using the following respective parameters: acquisition time, 0.2−0.3, 0.5, 0.5, 0.5 s; d1 delay, 1.2, 1.1, 1.0, 1.0 s; pulse sequence, hsqcedetgpsp.3, hmbcetgpl2nd, cosyjdqf, noesyetgp (or near equivalents in vnmrj); F_1 spectral width, 30, 70, 1.4, 1.4 ppm; o2p (center of F_1 window), 56, 45, 2.8, 2.8 ppm; number of FIDs, 512−1024, 400−1024, 332, or 444 (600 or 800 MHz with d6 = 0.2 s), 500−800. COSYDEC spectra were acquired with a d6 constant-time period of 0.2 s (sometimes also 0.3 s). The NOESY mixing time was 0.5 s. 2D spectra were processed with 4-fold linear prediction in F_1 and $\pi/2.5$-shifted qsine apodization in F_2 and F_1 (HSQC, HMBC, NOESY) using 4k and 2k points (si). The t1away macro (70−90%) was used as needed to attenuate t1 noise. Reduced F_1 windows (resulting in limited spectral aliasing) and 256−512 t1 increments allowed measurement of chemical shifts to ca. ±0.01 ppm (^{13}C) or ±0.001 ppm (^{1}H). The F_1 ppm labels in figures were adjusted to minimize confusion from aliasing.

Dynamic range in HSQC spectra was estimated from ratios of peak heights. Manual peak picking was used to choose the largest positive and negative signals (2 516 671 792 and −2 549 601 544, terminal

CH_3 and CH_2 peaks of long-chain impurities) and the smallest signals that were distinguishable from noise. A small positive peak (144 480) was chosen in the weakest clearly defined olefinic doublet (10 Hz coupling) among a group of similar doublets.

Isolation of Thalianol Oxidation Products. The crude extract of a preliminary 100 mL culture of JBY575[pGCF4.0, pDCR5.1, pDCR4.1] showed likely thalianol metabolites[23] by GC−MS. Purification by preparative TLC (developed with 1:2 ether/hexanes) gave the major metabolite **3**, which was tentatively identified as 7β-hydroxythalianol by NMR. The minor oxygenated thalianols were investigated with a larger culture (5 × 100 mL). The crude ethanol extract was subjected to flash chromatography (1 g silica SPE cartridge; elution with gradients of ether/hexanes). The 1:1 ether−hexanes fractions contained **3** and two minor compounds. As indicated in Figure 4C, preparative reversed-phase HPLC separated the three oxygenated thalianols. 2D NMR suggested the two minor compounds to be the 7α-hydroxy and 7-keto thalianols **4** and **5**. A third experiment (17 × 100 mL) using the same extraction and purification procedures gave ample amounts of **3**, **4**, and **5** for rigorous structure determination by HSQC, HMBC, COSYDEC, and NOESY (see the Supporting Information). Cultures of the control strain (lacking pDCR5.1) showed no evidence of **3**, **4**, or **5** by GC−MS.

Isolation of Arabidiol Oxidation Products. The crude ethanol extract from combined 8 × 100 mL cultures of JBY575[pMDK3.6, pDCR9.3, pDCR4.1] was saponified, followed by GC−MS analysis. Flash chromatography on a 2 g silica SPE cartridge with ether−hexanes gradients gave fractions A−H (listed as (% ether in hexanes), triterpenoid content): A (5% ether), squalene; B (15−40% ether), PEN1 products, including 14-epithalianol; C (50% ether), yeast metabolites, arabidiol (**6**); D (50% ether), **6**, 14-apo-arabidiol epimers (**10a**, **10b**), an unidentified side-chain oxygenated malabaricatrienol, ergosterol; E (100% ether), **7**, **8**, **9a**, **9b**, **10a**, **10b**, unknowns (see Figure 5); F, G, and H (100% ether): continuing elution of side-chain oxygenated tricyclic triterpenes (see Figure S9).

The main polar fraction (E) contained **10a** as the major arabidiol metabolite with some contamination of **10b** ascribed to partial epimerization of **10a** during saponification. To obtain a full profile of arabidiol oxidation products, we repeated the experiment (10 × 100 mL) but did not saponify the crude ethanol extracts. GC−MS and HSQC analysis showed no formation of **10b**. GC−MS of five independent control cultures for each of the saponified and nonsaponified experiments showed no formation of **10a** or other P450 products. Similarly, HSQC spectra of the two combined sets of control extracts each showed no **10a** (Figure 2C). Multiple control experiments were motivated by the 4-fold variability in GC−MS triterpenoid levels observed among different P450 cultures.

Isolation of Marneral Oxidation Products. GC−MS analysis of crude extracts from a trial 100 mL culture of JBY575[pXQ11.2, pDCR8.3, pDCR4.1] showed mainly marnerol, marneral, and a hydroxymarnerol. The Folch extract from a larger experiment (18 × 100 mL) was purified on a 5 g silica SPE cartridge using ether/hexanes gradients. Marnerol and other MRN1 products (eluting with 10−15% ether) were well separated from traces (<1%) of putative side-chain oxygenated marnerols (eluting with 30% ether) and from the hydroxymarnerol (eluting with 100% ether). Preparative reversed-phase HPLC provided the purified hydroxymarnerol, which was identified by 2D NMR as the 23-hydroxy derivative **12b**. In another experiment (10 × 100 mL), the crude ethanol extract was purified on an SPE column (CH_2Cl_2 elution, followed by gradients of ether/hexanes). Preparative reversed-phase HPLC then gave purified samples of **11a** and **11b**, **12a**, and **12b**. The minor and major 23-hydroxy metabolites **12a** and **12b** were characterized by GC−MS (Figure 4C) and 2D NMR (see the Supporting Information). Metabolites **12a** and **12b** were undetectable in GC−MS and HSQC (Figure 3A and B) of several control cultures.

NMR Signal Assignments. ^{13}C and ^1H NMR assignments for **3**, **4**, **5**, **10a**, **10b**, **12a**, and **12b** are presented in Tables S2,S3. As described in section 12 of the Supporting Information, triterpenoid structures were elucidated from 2D NMR correlations, comparisons with reported NMR chemical shifts for triterpenes,[13,16,30−33] and

compatibility with quantum mechanical predictions of NMR chemical shifts (Tables S5−S7).

ASSOCIATED CONTENT

S **Supporting Information**

Details of subcloning and yeast transformations, alignment and evaluation of coding sequences for ATR2, microarray data showing coexpression of OSCs and P450s, possible mechanisms for P450 cleavage of arabidiol, tables of NMR chemical shifts and their quantum mechanical predictions, and figures of NMR and mass spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

matsuda@rice.edu

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Pichersky, E.; Gang, D. R. *Trends Plant Sci.* **2000**, *5*, 439−445.

(2) Weng, J. K.; Philippe, R. N.; Noel, J. P. *Science* **2012**, *336*, 1667−1670.

(3) Croteau, R.; Kutchan, T. M.; Lewis, N. G. In *Biochemistry and Molecular Biology of Plants*; Buchanan, B., Gruissem, W., Jones, R., Eds.; American Society of Plant Physiologists: Rockville, MD, 2000; pp 1250−1318.

(4) (a) Bak, S.; Beisson, F.; Bishop, G.; Hamberger, B.; Höfer, R.; Paquette, S.; Werck-Reichhart, D. *Cytochromes P450*; The Arabidopsis Book, American Society of Plant Biologists: Rockville, MD, 2011; doi: 10.1199/tab.0144. (b) Schuler, M. A.; Duan, H.; Bilgin, M.; Ali, S. *Phytochem. Rev.* **2006**, *5*, 205−237. (c) Nelson, D. R. *Hum. Genomics* **2009**, *4*, 59−65.

(5) The number of *Arabidopsis* P450s excludes pseudogenes. Estimates of characterized P450s are derived from literature reports and the annotated list of *Arabidopsis* P450s at www-ibmp.u-strasbg.fr/~CYPedia/index.html.

(6) Leading references for some *Arabidopsis* P450s linked to specific natural products. Homoterpenes: (a) Lee, S.; Badieyan, S.; Bevan, D. R.; Herde, M.; Gatz, C.; Tholl, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21205−21210. Camalexins: (b) Beets, C. A.; Huang, J. C.; Madala, N. E.; Dubery, I. *Planta* **2012**, *236*, 261−272. Arabidopyrones: (c) Weng, J. K.; Li, Y.; Mo, H.; Chapple, C. *Science* **2012**, *337*, 960−964. Glucosinolates: (d) Chen, S.; Glawischnig, E.; Jorgensen, K.; Naur, P.; Jorgensen, B.; Olsen, C. E.; Hansen, C. H.; Rasmussen, H.; Pickett, J. A.; Halkier, B. A. *Plant J.* **2003**, *33*, 923−937. Many early steps of specialized metabolism (e.g., oxidosqualene formation) are shared with biosynthetic pathways that are broadly distributed among plants and constitutively expressed. Examples of known P450 oxidations in early, shared polypropanoid biosynthesis: (e) Pan, Y.; Michael, T. P.; Hudson, M. E.; Kay, S. A.; Chory, J.; Schuler, M. A. *Plant Physiol.* **2009**, *150*, 858−878.

(7) (a) von Roepenack-Lahaye, E.; Degenkolb, T.; Zerjeski, M.; Franz, M.; Roth, U.; Wessjohann, L.; Schmidt, J.; Scheel, D.; Clemens, S. *Plant Physiol.* **2004**, *134*, 548−559. (b) D'Auria, J. C.; Gershenzon, J.

Curr. Opin. Plant Biol. **2005**, *8*, 308−316. The number of secondary metabolites has markedly increased since the 2005 census of 170, due mainly to LC−MS/MS identification of phenylpropanoids, glucosinolates, and other polar substances. However, few new P450 metabolites of terpenes and triterpenes have been discovered in *Arabidopsis*.

(8) Hamberger, B.; Bak, S. *Philos. Trans. R. Soc., B* **2013**, *368*, 20120426.

(9) Misawa, N. *Curr. Opin. Biotechnol.* **2011**, *22*, 627−633.

(10) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261−291.

(11) Sheng, H.; Sun, H. *Nat. Prod. Rep.* **2011**, *28*, 543−593.

(12) Hill, R. A.; Connolly, J. D. *Nat. Prod. Rep.* **2012**, *29*, 780−818.

(13) Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678−5679.

(14) Xue, Z.; Duan, L.; Liu, D.; Guo, J.; Ge, S.; Dicks, J.; O'Máille, P.; Osbourn, A.; Qi, X. *New Phytol.* **2012**, *193*, 1022−1038.

(15) Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305−314.

(16) Morlacchi, P.; Wilson, W. K.; Xiong, Q.; Bhaduri, A.; Sttivend, D.; Kolesnikova, M. D.; Matsuda, S. P. T. *Org. Lett.* **2009**, *11*, 2627−2630.

(17) (a) Lytovchenko, A.; Beleggia, R.; Schauer, N.; Isaacson, T.; Leuendorf, J. E.; Hellmann, H.; Rose, J. K.; Fernie, A. R. *Plant Methods* **2009**, *5*, 4. (b) Quanbeck, S. M.; Brachova, L.; Campbell, A. A.; Guan, X.; Perera, A.; He, K.; Rhee, S. Y.; Bais, P.; Dickerson, J. A.; Dixon, P.; Wohlgemuth, G.; Fiehn, O.; Barkan, L.; Lange, I.; Lange, B. M.; Lee, I.; Cortes, D.; Salazar, C.; Shuman, J.; Shulaev, V.; Huhman, D. V.; Sumner, L. W.; Roth, M. R.; Welti, R.; Ilarslan, H.; Wurtele, E. S.; Nikolau, B. J. *Front. Plant Sci.* **2012**, *3*, 15.

(18) Shan, H.; Wilson, W. K.; Phillips, D. R.; Bartel, B.; Matsuda, S. P. T. *Org. Lett.* **2008**, *10*, 1897−1900.

(19) Suzuki, H.; Achnine, L.; Xu, R.; Matsuda, S. P. T.; Dixon, R. A. *Plant J.* **2002**, *32*, 1033−1048.

(20) Haralampidis, K.; Trojanowska, M.; Osbourn, A. E. *Adv. Biochem. Eng. Biotechnol.* **2002**, *75*, 31−49.

(21) Seki, H.; Sawai, S.; Ohyama, K.; Mizutani, M.; Ohnishi, T.; Sudo, H.; Fukushima, E. O.; Akashi, T.; Aoki, T.; Saito, K.; Muranaka, T. *Plant Cell* **2011**, *23*, 4112−4123.

(22) (a) Shibuya, M.; Hoshino, M.; Katsube, Y.; Hayashi, H.; Kushiro, T.; Ebizuka, Y. *FEBS J.* **2006**, *273*, 948−959. (b) Fukushima, E. O.; Seki, H.; Ohyama, K.; Ono, E.; Umemoto, N.; Mizutani, M.; Saito, K.; Muranaka, T. *Plant Cell Physiol.* **2011**, *52*, 2050−2061. (c) Han, J. Y.; Hwang, H. S.; Choi, S. W.; Kim, H. J.; Choi, Y. E. *Plant Cell Physiol.* **2012**, *53*, 1535−1545. (d) Huang, L.; Li, J.; Ye, H.; Li, C.; Wang, H.; Liu, B.; Zhang, Y. *Planta* **2012**, *236*, 1571−1581.

(23) Field, B.; Osbourn, A. E. *Science* **2008**, *320*, 543−547.

(24) Field, B.; Fiston-Lavier, A. S.; Kemen, A.; Geisler, K.; Quesneville, H.; Osbourn, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16116−16121.

(25) Qi, X.; Bakht, S.; Leggett, M.; Maxwell, C.; Melton, R.; Osbourn, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8233−8238.

(26) Recent advances in LC−MS have overcome some of the dependence on authentic standards: Wu, H.; Guo, J.; Chen, S.; Liu, X.; Zhou, Y.; Zhang, X.; Xu, X. *J. Pharm. Biomed. Anal.* **2013**, *72*, 267−291.

(27) Domingo, V.; Arteaga, J. F.; Quilez del Moral, J. F.; Barrero, A. F. *Nat. Prod. Rep.* **2009**, *26*, 115−134.

(28) Ritzdorf, I.; Bartels, M.; Kerp, B.; Kasel, T.; Klonowski, S.; Marner, F.-J. *Phytochemistry* **1999**, *50*, 995−1003.

(29) Marner, F. J. *Curr. Org. Chem.* **1997**, *1*, 153−186.

(30) Xiong, Q.; Wilson, W. K.; Matsuda, S. P. T. *Angew. Chem., Int. Ed.* **2006**, *45*, 1285−1288.

(31) Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Kolesnikova, M. D.; Onak, C. S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2007**, *129*, 11213−11222.

(32) Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183−2186.

(33) Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Ivanova, Y.; Smith, M. L.; May, G. S.; Matsuda, S. P. T. *Org. Lett.* **2009**, *11*, 1241−1244.

(34) Duan, H.; Schuler, M. A. *Phytochem. Rev.* **2006**, *5*, 507−523.

(35) Lang, G.; Mayhudin, N. A.; Mitova, M. I.; Sun, L.; van der Sar, S.; Blunt, J. W.; Cole, A. L.; Ellis, G.; Laatsch, H.; Munro, M. H. *J. Nat. Prod.* **2008**, *71*, 1595−1599.

(36) Heterologous expression in *E. coli* to study sesquiterpene biosynthesis: Harada, H.; Shindo, K.; Iki, K.; Teraoka, A.; Okamoto, S.; Yu, F.; Hattan, J.; Utsumi, R.; Misawa, N. *Appl. Microbiol. Biotechnol.* **2011**, *90*, 467−476.

(37) Sainsbury, F.; Saxena, P.; Geisler, K.; Osbourn, A.; Lomonossoff, G. P. *Methods Enzymol.* **2012**, *517*, 185−202.

(38) Lewinsohn, E.; Gijzen, M. *Plant Sci.* **2009**, *176*, 161−169.

(39) (a) Osbourn, A.; Papadopoulou, K. K.; Qi, X.; Field, B.; Wegel, E. *Methods Enzymol.* **2012**, *517*, 113−138. (b) Wada, M.; Takahashi, H.; Altaf-Ul-Amin, M.; Nakamura, K.; Hirai, M. Y.; Ohta, D.; Kanaya, S. *Gene* **2012**, *503*, 56−64.

(40) Ehlting, J.; Provart, N. J.; Werck-Reichhart, D. *Biochem. Soc. Trans.* **2006**, *34*, 1192−1198.

(41) Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2211−2215.

(42) Review: (a) Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S. *Acc. Chem. Res.* **2012**, *45*, 288−297. HSQC of an *Arabidopsis* extract: (b) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385−9390. Quantifying triterpenoids in plant extracts by HSQC: (c) Kontogianni, V. G.; Exarchou, V.; Troganis, A.; Gerothanassis, I. P. *Anal. Chim. Acta* **2009**, *635*, 188−195. Quantitative HSQC of mixtures: (d) Hu, K.; Westler, W. M.; Markley, J. L. *J. Am. Chem. Soc.* **2011**, *133*, 1662−1665. Covariance 2D NMR in mixture analysis: (e) Zhang, F.; Bruschweiler-Li, L.; Bruschweiler, R. *J. Am. Chem. Soc.* **2010**, *132*, 16922−16927. Differential 2D NMR mixture analysis to establish biological function: (f) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7708−7713.

(43) Cankar, K.; van Houwelingen, A.; Bosch, D.; Sonke, T.; Bouwmeester, H.; Beekwilder, J. *FEBS Lett.* **2011**, *585*, 178−182.

(44) 2D NMR of this mixture of **7**, **8**, **9a**, **9b**, **10a**, and **10b** resolved all of the ^{1}H and ^{13}C NMR signals of **10a** and **10b** and gave correlations sufficient for their signal assignments and definitive structure determination (Tables S2−S5).

(45) Corey, E. J.; Cheng, H. *Tetrahedron Lett.* **1996**, *37*, 2709−2712.

(46) Epimer **10a** was slightly favored in the Boltzmann distribution based on B3PW91/6-311G(2d,p)//B3LYP/6-31G* relative energies: **10a**, C13−C14 rotamers, 0.09 and 0.96 kcal/mol; **10b**, C13−C14 rotamers, 0.00 and 4.08 kcal/mol; predicted ratio of epimers/rotamers: 41.7, 9.6; 48.6, 0.0.

(47) Go, Y. S.; Lee, S. B.; Kim, H. J.; Kim, J.; Park, H. Y.; Kim, J. K.; Shibata, K.; Yokota, T.; Ohyama, K.; Muranaka, T.; Arseniyadis, S.; Suh, M. C. *Plant J.* **2012**, *72*, 791−804.

(48) Mizutani, M. *Biol. Pharm. Bull.* **2012**, *35*, 824−832.

(49) Pichersky, E.; Lewinsohn, E. *Annu. Rev. Plant Biol.* **2011**, *62*, 549−566.

(50) Guengerich, F. P. *Chem. Res. Toxicol.* **2001**, *14*, 611−650.

(51) Strushkevich, N.; MacKenzie, F.; Cherkesova, T.; Grabovec, I.; Usanov, S.; Park, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10139−10143.

(52) Tholl, D.; Sohrabi, R.; Huh, J. H.; Lee, S. *Phytochemistry* **2011**, *72*, 1635−1646.

(53) Weng, J. K.; Noel, J. P. *Cold Spring Harbor Symp. Quant. Biol.* **2012**, DOI: 10.1101/sqb.2012.77.014787.

(54) Wassif, C. A.; Brownson, K. E.; Sterner, A. L.; Forlino, A.; Zerfas, P. M.; Wilson, W. K.; Starost, M. F.; Porter, F. D. *Hum. Mol. Genet.* **2007**, *16*, 1176−1187.

(55) Withers, J. C.; Shipp, M. J.; Rupasinghe, S. G.; Sukumar, P.; Schuler, M. A.; Muday, G. K.; Wyatt, S. E. *Am. J. Bot.* **2013**, *100*, 183−193.

(56) Zust, T.; Heichinger, C.; Grossniklaus, U.; Harrington, R.; Kliebenstein, D. J.; Turnbull, L. A. *Science* **2012**, *338*, 116−119.

(57) Counterexamples include combining chemical genomics and metabolomics to elucidate unknown pathways. For leading references, see: Bocobza, S.; Willmitzer, L.; Raikhel, N. V.; Aharoni, A. *Plant Physiol.* **2012**, *160*, 1160−1163.

(58) Kliebenstein, D. J. *Annu. Rev. Phytopathol.* **2012**, *50*, 155−173.

(59) Mintz-Oron, S.; Meir, S.; Malitsky, S.; Ruppin, E.; Aharoni, A.; Shlomi, T. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 339−344.

(60) HSQC establishes stereochemistry and carbon connectivity, while MS supplies information about heteroatoms. Although their relative strengths differ between hydrogen-poor polypropanoids and stereochemically rich terpenoids, neither method is effective alone, as illustrated in a phenylpropanoid study: Xu, Y.-J.; Foubert, K.; Dhooghe, L.; Lemiere, F.; Maregesi, S.; Coleman, C. M.; Zou, Y.; Ferreira, D.; Apers, S.; Pieters, L. *Phytochemistry* **2012**, *79*, 121−128.