Emergence of a proton exchange-based isomerization and lactonization mechanism in the plant coumarin synthase COSY

Colin Y. Kim[1,2], Andrew J. Mitchell[1], David W. Kastner[2,3], Claire E. Albright[1], Michael Gutierrez[1], Christopher M. Glinkerman[1], Heather J. Kulik[3], Jing-Ke Weng[1,4]*

Author affiliations

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142, United States

[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, United States

[3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, United States

[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, United States

*Email: wengj@wi.mit.edu

Abstract

Plants contain rapidly evolving specialized metabolic enzymes to support the synthesis of a myriad of functionally diverse natural products. In the case of coumarin biosynthesis, a BAHD acyltransferase-family enzyme COSY was recently discovered in Arabidopsis that catalyzes coumarin formation from *o*-hydroxylated *trans*-hydroxycinnamoyl-CoA substrates. COSY is the first and only BAHD enzyme known to date that catalyzes an intramolecular acyl transfer reaction. Here we combine structural, biochemical, and computational approaches to investigate the mechanistic basis for the unique coumarin synthase activity of COSY. Comparative analyses of crystal structures of *Arabidopsis thaliana* COSY relative to other BAHD proteins reveal that COSY possesses an unconventional active-site configuration adapted to its specialized activity. Through deuterium exchange experiments, we discover a unique proton exchange mechanism at the α-carbon of the *o*-hydroxylated *trans*-hydroxycinnamoyl-CoA substrates during the catalytic cycle of COSY. Mutagenesis studies and quantum mechanical cluster modeling further support that this mechanism is key to COSY's ability to lower the activation energy of the *trans*-to-*cis* isomerization of the hydroxycinnamoyl-CoA substrates, a critical rate-limiting step leading to coumarin production. This study unveils the emergence of an unconventional catalytic mechanism mediated by a BAHD-family enzyme, and sheds light on the potential evolutionary origin of COSY and its recruitment to the evolutionarily new coumarin biosynthetic pathway in eudicots.

Introduction

Coumarins are 1,2-benzopyrone natural products found in many plants that serve diverse biological functions, including iron acquisition from the soil, plant defense against herbivores and pathogens, and regulation of plant-microbiome interactions[1–3]. In addition to their roles in plants' fitness, coumarins exhibit various pharmacological properties, such as antiviral[4], antibacterial[5] and anticoagulant activities[6], and represent an important pharmacophore in small-molecule drug development. Understanding genetic and biochemical mechanisms underlying coumarin biosynthesis in plants lays the foundation for expanding the utility of coumarins in agricultural and pharmaceutical applications[1,2].

Coumarins are biosynthetically derived from the plant phenylpropanoid pathway[2]. The *ortho*-hydroxylation of hydroxycinnamoyl-CoAs, catalyzed by Fe(II)- and 2-oxoglutarate-dependent dioxygenase Feruloyl-CoA 6′-Hydroxylase (F6′H), is a key step in coumarin biosynthesis, as it creates a branch point in phenylpropanoid metabolism, directing a portion of the hydroxycinnamoyl-CoA intermediates towards coumarin production[7]. The subsequent *trans*-to-*cis* isomerization and lactonization of the derived *o*-hydroxylated hydroxycinnamoyl-CoAs to form coumarins had been thought to occur spontaneously[8]. Recently, Vanholme et al. reported the identification of coumarin synthase (COSY) from *Arabidopsis thaliana*, which exhibits specific enzymatic activities that accelerate coumarin formation from *o*-hydroxylated *trans*-hydroxycinnamoyl-CoA substrates[8]. Consistent with COSY's *in vitro* biochemical activities, *A. thaliana cosy* mutants show reduced levels of scopoletin and sideretin,

which are major coumarins in Arabidopsis, and various iron-deficiency growth phenotypes when grown in iron-deficient soil[8].

COSY belongs to the BAHD acyltransferase family, which is named after the first four discovered enzymes of this family: benzyl alcohol O-acetyltransferase (BEAT)[9], anthocyanin O-hydroxycinnamoyl transferase (AHCT)[10], N-hydroxycinnamoyl/benzoyltransferase (HCBT)[11], and deacetylvindoline 4-O-acetyltransferase (DAT)[12]. Canonical BAHD acyltransferases catalyze the transfer of the acyl group from an acyl-CoA thioester donor molecule to an -OH or -NH_2-containing acyl acceptor molecule[13]. All canonical BAHD acyltransferases known to date contain a ubiquitously conserved catalytic histidine. In the ester-forming BAHDs, the catalytic histidine serves as a general base to deprotonate the -OH of the acyl acceptor substrate, priming it for nucleophilic attack on the acyl-CoA substrate[14]. Many BAHD acyltransferases also contain a highly conserved tryptophan residue in the active site, with its indolic nitrogen serving as an oxyanion hole to stabilize the tetrahedral intermediate formed after the initial nucleophilic attack[14]. With its specific role in coumarin biosynthesis, COSY is the first BAHD acyltransferase that uses a single substrate as both the acyl donor and acceptor to produce a lactone product. In addition, COSY may also facilitate *trans*-to-*cis* isomerization of *o*-hydroxylated hydroxycinnamoyl-CoAs prior to lactonization as part of its catalytic cycle, implicating a potential reconfiguration of the ancestral canonical BAHD acyltransferase catalytic machinery to accommodate its unique coumarin synthase activity[8].

To understand the mechanistic basis for the unconventional activity of COSY as a BAHD-family enzyme, we present several crystal structures of *A. thaliana* COSY (*At*COSY) in its apo and various ligand-bound forms. Through comparative structural analyses, site-directed mutagenesis, deuterium labeling, and quantum mechanical (QM) cluster modeling, we uncover the structural features and catalytic mechanism underlying COSY's unique coumarin synthase activity. Through phylogenomic analyses, we also query the potential evolutionary origin of COSY as a new catalyst in plant coumarin biosynthesis.

Results

Structural features of *At*COSY

To investigate the molecular basis for the unique catalytic activity of COSY, we first obtained the apo structure of *At*COSY at 1.9 Å resolution (Supplementary Fig. 1). Similar to other BAHD acyltransferases[14], *At*COSY displays pseudo-symmetric N-terminal (residues 1-181) and C-terminal (residues 230-451) domains, connected by a linker loop (residues 182-229) (Supplementary Fig. 1). The *At*COSY active site is located at the interface between the two pseudo-symmetric domains, and features the conserved catalytic residues, His161 and Trp371, stemming from the N- and C-terminal domains, respectively (Supplementary Fig. 1). Additionally, we observed a single Ca^{2+} ion that binds to the linker loop region and is coordinated by the main-chain carbonyl oxygens of Arg211 and Leu214 residues (Supplementary Fig. 1).

Next, we obtained product-bound states of *At*COSY with scopoletin and umbelliferone at 2.5 Å and 2.8Å resolutions, respectively (Fig. 1a, Supplementary Figs. 2 and 3). Scopoletin and umbelliferone are common coumarins found in plants both containing a hydroxy substituent at C7, with scopoletin possessing an extra methoxy substituent at C6. Both compounds adopt a similar pose in *At*COSY's active site coordinated by a shared set of active-site-lining residues, notably Phe40, Tyr42 and Cys164 (Fig. 1a, Supplementary Figs. 2, 3). The benzene ring of Phe40 forms a π–π stacking interaction with the O-containing heterocycle of scopoletin, while the side chains of Tyr42 and Cys164 form hydrogen bonds with the methoxy and hydroxy substituents of scopoletin, respectively (Supplementary Fig. 2). Multiple sequence alignment of three COSY orthologs together with three phylogenetically related Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) orthologs indicates that Phe40 is shared between the two orthologous groups, whereas Tyr42 and Cys164 are uniquely conserved among COSY orthologs (Fig. 1b, Supplementary Fig. 4). Nevertheless, close inspection of the *p*-coumaroyl-shikimate-bound *At*HCT structure (PDB ID: 5KJU)[14] revealed that Phe34 (corresponding to Phe40 in *At*COSY) is a surface-exposed residue not involved in substrate/product binding.

All previously published canonical BAHD acyltransferase structures harbor two substrate-binding pockets that accommodate binding of acyl donors and acyl acceptors, respectively[14–16]. However, when the scopoletin-bound *At*COSY structure is superpositioned with the *p*-coumaroylshikimate-bound *At*HCT structure (RMSD = 2.034

Å)[14], it is evident that the accessible acyl-acceptor-binding pocket of *At*HCT is occluded by an active-site loop (residues 392-400) in *At*COSY (Figs. 1b,c). *At*COSY's confined active-site configuration demonstrates the structural need for specialized coumarin synthase activity, which involves a single *o*-hydroxylated hydroxycinnamoyl-CoA substrate containing both the acyl donor and acceptor moieties (e.g., 6-hydroxyferuloyl-CoA; 6OHFCoA).

Despite multiple attempts of soaking and co-crystallization of *At*COSY with various substrate mimetics, including *p*-coumaroyl-CoA, feruloyl-CoA, dihydro-*p*-coumaroyl-CoA and cinnamoyl-CoA, we did not yield substrate-bound crystal structures. Instead, we were able to obtain the structure of *At*COSY in complex with free CoA, which occupies the conventional BAHD family CoA binding site when superimposed with *p*-coumaroyl-CoA-bound *At*HCT (PDB: 5KJT) (Supplementary Fig. 5). We noticed that a single oxidation of Cys308 to sulfenic acid was present in the CoA-bound structure, suggesting that it may play a role as a redox-gate for the proper substrate entry (Supplementary Fig. 6a). Guided by both the CoA-bound and scopoletin-bound *At*COSY structures, we modeled the *trans*-6OHFCoA substrate in the COSY active site (Fig. 2d). This reconstructed pose of *trans*-6OHFCoA indicates that residues Phe40, Tyr42, His161, Cys164 and Trp371 engage close contacts with the substrate, and are likely involved in specific substrate binding and catalysis (Fig. 2d).

Structure-guided biochemical characterization of *At*COSY

To enable functional characterization of COSY, we established a coupled assay system under light where the COSY substrate, 6OHFCoA or 2-hydroxy-*p*-coumaroyl CoA (2OHpCCoA), is first generated from its corresponding free acid, 6OHFA or 2-hydroxy-*p*-coumaric acid (2OHpCA), using *A. thaliana* 4-coumarate-CoA ligase 1 (*At*4CL1) in the presence of free CoA, ATP and Mg^{2+}. After 3 min of preincubation, *At*COSY is then added. The reaction proceeds for an additional 15 min before it is quenched by urea, followed by immediate analysis by LC-MS (Fig. 2a, Supplementary Figs. 7,8, Methods). We observed substantial spontaneous formation of coumarin from 6OHFCoA or 2OHpCCoA in the absence of *At*COSY, while addition of *At*COSY resulted in significant acceleration of coumarin production (Figs. 2b,c, Supplementary Fig. 7), consistent with the previously reported observation[8]. Since we observed binding of a Ca^{2+} ion in the *At*COSY structure, we tested whether the coumarin synthase activity may be dependent upon Ca^{2+} or other divalent metals. Nevertheless, we observed no significant difference in coumarin synthase activity whether or not Ca^{2+} is present in the assay buffer (Supplementary Fig. 9). Similarly, addition of Zn^{2+}, Ni^{2+}, Mn^{2+}, Co^{2+}, Fe^{2+} or Cu^{2+} to the assay buffer had little impact on the coumarin synthase activity (Supplementary Fig. 10), suggesting the observed calcium-binding site may be an artifact of crystallization. Moreover, based on our observation of Cys308 oxidation in the CoA-bound structure, we generated the C308A and C308S mutants, and assayed their activities in buffers containing a concentration gradient of hydrogen peroxide (Supplementary Figs. 6b,8). Even so, varying redox environments had little effect on coumarin synthase activity for wild-type (WT), C308A, and C308S enzymes

(Supplementary Fig. 6b). Thus, C308 oxidation is likely a crystallography artifact in our solved structures.

The ubiquitously conserved catalytic histidine is critical for BAHD acyltransferase activities[17]. Therefore, to investigate the role of this residue in *At*COSY, we generated the H161A and H161Q mutants and assayed their activities using 6OHFCoA as the substrate. Surprisingly, the H161A and H161Q mutants exhibit WT-level coumarin synthase activity (Fig. 2e, Supplementary Figs. 8,11, Supplementary Table 2). Although these results cannot rule out the possibility that His161 may still serve as a base that deprotonates the 6-OH group of 6OHFCoA as in other BAHD acyltransferases, His161 seems to be dispensable for the overall coumarin synthase activity, suggesting that 6-OH deprotonation is likely not a rate-limiting step in the full catalytic cycle of COSY.

Next, we examined the role of Trp371, which is the other highly conserved catalytic residue among many BAHD acyltransferases that serves as the oxyanion hole to stabilize the tetrahedral intermediate[14]. Four Trp371 mutants, W371H, W371A, W371V and W371M, were generated and assayed against 6OHFCoA as the substrate. Whereas the W371H mutant shows WT-level coumarin synthase activity, the W371A, W371V and W371M mutants exhibit significantly reduced, but not completely abolished activities (Fig. 2e, Supplementary Figs. 8,11, Supplementary Table 2). These results suggest that W371 plays an important but nonessential role in the catalytic cycle of COSY. W371 likely contributes to the catalytic cycle of COSY through its indole nitrogen, which can be substituted by the nitrogen-containing imidazole in the W371H mutant.

We then examined the roles of four additional active-site-lining residues: Phe40, Tyr42 and Cys164, which make close contact with the phenolic portion of COSY substrates and coumarin products, and Gly166, the backbone nitrogen of which is in hydrogen bonding distance to C2 of the 6OHFCoA substrate (Figs. 1a, 2d). We generated the F40T and C164A single mutants, as these substitutions are observed between *At*COSY and *At*HCT structures at the corresponding positions (Supplementary Figs. 8,12). Additionally, we generated the Y42F mutant with the intention to abolish the hydrogen bond formed between the Tyr42 *p*-hydroxy and the methoxy group of scopoletin. To test the importance of the hydrogen bonding of Gly166 backbone nitrogen with C2, we also generated the G166A and G166L mutants (Supplementary Fig. 8). Furthermore, we constructed two double mutants F40T/Y42S and Y42F/H161A to perturb the COSY active site (Supplementary Fig. 8). The relative activities of these mutant enzymes were assayed against the 6OHFCoA substrate. Whereas the F40T, C164A, and F40T/Y42S mutants show WT-level coumarin synthase activities, the Y42F mutant shows a modest decrease in activity (Fig. 2e, Supplementary Fig. 11, Supplementary Table 2). These results suggest that none of the three residues are essential for the coumarin synthase activity of COSY despite their direct interactions with the substrate and product in the active site. The G166A mutant exhibits WT-level activity. In contrast, the steric bulk of leucine in the G166L mutant likely disfavors hydrogen bonding capability of its backbone nitrogen, leading to a significant decrease in coumarin synthase activity (Fig. 2e, Supplementary Fig. 11, Supplementary Table 2). The Y42F/H161A double mutant did show a complete loss of coumarin synthase activity

(Fig. 2e, Supplementary Fig. 11, Supplementary Table 2), although this is likely due to overall impaired structural stability of this mutant. It was noted that the Y42F/H161A mutant was difficult to express and purify compared to WT *At*COSY and other mutants.

In search for other key catalytic residues of *At*COSY, we examined the various ligand-bound crystal structures for conformational changes of residues surrounding the active site. Indeed, we observed conformational differences of an active-site loop containing Tyr373 and Leu374 between the CoA-bound and scopoletin-bound *At*COSY structures. In the CoA-bound structure, Tyr373 exhibits π-π stacking interaction with the adenine portion of CoA causing Leu374 to shift towards C2 of 6OHFCoA, relative to its conformation in the scopoletin-bound structure (Fig. 2f, Supplementary Fig. 13). To test the role of this dynamic loop region, we generated the Y373A, Y373F, and L374A mutants and examined their activities against 6OHFCoA (Supplementary Fig. 8). The Y373A and Y373F mutants exhibit WT-level coumarin synthase activities, whereas the L374A mutant shows a significant decrease in activity similar to that of the W371A mutant (Fig. 2e, Supplementary Fig. 11, Supplementary Table 2). These observations indicate that the π-π stacking interaction of Tyr373 with CoA does not contribute significantly to the coumarin synthase activity, while Leu374 on this dynamic loop region is likely involved in catalysis, albeit in a nonessential manner.

Altogether, our site-directed mutagenesis analyses of *At*COSY did not pinpoint specific residues that are absolutely required for the coumarin synthase activity. This is in stark contrast to conventional BAHD acyltransferases and other distantly related acyltransferases which require a critical basic residue to initiate the catalytic cycle as

well as additional key catalytic residues for stabilizing the tetrahedral intermediates or leaving groups[14,18,19]. We next examined the role of the uniquely observed acyl-acceptor binding site occlusion in *At*COSY by constructing a mutant with its active-site loop occupying this space swapped with the corresponding region of *At*HCT based on sequence and structural alignment (Figs. 1b,c, Supplementary Fig. 8). This 'Loop-swap' mutant exhibits a significant decrease in coumarin synthase activity when assayed against the 6OHFCoA substrate, while harboring no *p*-coumaroyl-shikimate-producing-activity when assayed for *At*HCT activity (Fig. 2e, Supplementary Figs. 11, 14, Supplementary Table 2). With this observation and given that acyl acceptor deprotonation does not seem to constitute the primary energy barrier for coumarin formation from *o*-hydroxylated hydroxycinnamoyl-CoAs, we hypothesized that the main catalytic mechanism of COSY is to employ the overall geometry and dynamism of its active site to facilitate substrate *trans*-to-*cis* isomerization, and thus bring the reactive groups closer to trigger lactonization.

Mechanistic basis for substrate *trans*-to-*cis* isomerization catalyzed by COSY

Vanholme et al. proposed a mechanism for the *trans*-to-*cis* isomerization of *o*-hydroxylated hydroxycinnamoyl-CoA substrates through intramolecular electron delocalization prior to lactonization[8]. However, we were intrigued by a previous study on scopoletin and scopolin biosynthesis in *Manihot esculenta* (cassava) showing that *in planta* feeding of *E*-cinnamic-2-d_1 acid resulted in a hydrogen/deuterium (H/D) exchange at C2 of the derived hydroxycoumarins[20]. This observation led us to

hypothesize that enzyme-catalyzed coumarin formation may involve acid-base catalysis at C2, which was not depicted by the mechanism proposed by Vanholme et al[8].

To test this hypothesis, we synthesized D2-6OHFA with two deuteriums installed at C2 and C6, respectively, and used it as the hydroxycinnamic acid input in the *At*4CL1-*At*COSY coupled assay (Fig. 3a, Supplementary Figs. 15-17). In the control experiment without *At*COSY, spontaneous coumarin formation from D2-6OHFCoA yielded scopoletin showing the predominant ion of 193.04747 (m/z); **4**, suggesting the retention of both deuteriums (Figs. 3b,d, Supplementary Fig. 18). On the contrary, the COSY-catalyzed reaction yielded scopoletin showing the predominant ion of 192.04114 (m/z); **5**, suggesting the exchange of one deuterium with hydrogen at C2 (Figs. 3b,d). Presumably, the H/D exchange entails protonation of C2 followed by deprotonation or dedeuteronation. Interestingly, we noticed that 89.7% of the scopoletin product formed in the COSY-catalyzed reaction was **5**, which is much higher than 50%—the theoretical ratio if deprotonation or dedeuteronation occurs non-selectively, not considering kinetic isotope effect (KIE). When KIE is considered, since the zero-point energy of a C-H bond is higher than that of a C-D bond, deprotonation of the C-H bond at C2 would be favored if this occurred non-selectively. However, our data showed the opposite is true. This result suggests that the second-half deprotonation step (dedeuteronation in the case of deuterium-labeled precursor) occurs in a stereoselective manner in the COSY active site (Figs. 3d, 4). Additionally, this also indicates that the first-half protonation step of C2 occurs stereoselectively at the *si* face of the olefin, as non-selective protonation would result in a lower H/D exchange rate (Figs. 3d, 4).

To further test the C2 proton exchange mechanism mediated by COSY, we performed the reciprocal H/D exchange experiment by assaying *At*COSY against the 2OHpCCoA substrate in 66% D_2O. Consistent with the H/D exchange seen in the previous experiment using deuterium-labeled substrate, we observed significant exchange of C2 hydrogen with deuterium from D_2O in the resultant umbelliferone product (Supplementary Fig. 19). The presence of deuteron at C2 is further confirmed by NMR analysis of the purified umbelliferone product from the enzyme assay (Supplementary Fig. 20). In contrast, spontaneous umbelliferone formation from 2OHpCCoA in 66% D_2O in the absence of *At*COSY showed no H/D exchange (Supplementary Fig. 19).

To test whether COSY is responsible for the acid-base chemistry that occurs at C2 of coumarin biosynthesis *in vivo*, we grew WT *A. thaliana* (Col-0) and *the cosy* mutant (SALK_080878) on MS media containing 50% D_2O, and measured H/D exchange rate in the major coumarin product of Arabidopsis, scopoletin. Consistent with the results from the *in vitro* H/D exchange experiments, scopoletin produced in WT plants shows much higher deuterium labeling than that of the *cosy* mutant (Supplementary Fig. 21), supporting that *At*COSY is responsible for catalyzing the C2 acid-base chemistry in coumarin production *in vivo*.

We hypothesized that the COSY-dependent proton exchange phenomenon at C2 of coumarin biosynthesis is directly linked to COSY's catalytic function in mediating the *trans*-to-*cis* isomerization of the *o*-hydroxylated hydroxycinnamoyl-CoA substrate. To identify residues that may be involved in the acid-base catalysis at C2, we assayed

various *At*COSY mutants against D2-6OHFACoA, and measured the relative H/D exchange rate in the resultant scopoletin product (Fig. 3c, Supplementary Fig. 18). Among the mutants tested, the W371A, G166L, and 'Loop-swap' mutants show a major decrease of H/D exchange in scopoletin compared to that of the WT enzyme (Fig. 3c, Supplementary Fig. 18), which is correlated with their reduced overall COSY activity (Fig. 2e). Trp371 is positioned close to C2 of the reconstructed *trans*-6OHFACoA substrate in the active site of the CoA-bound *At*COSY structure (Fig. 2d), and therefore may play a role in proton exchange at C2 through a water molecule coordinated by its indole amine. The backbone nitrogen of Gly166 is close to C2 of 6OHFCoA and may engage in the hydrogen bonding network that stabilizes critical reaction intermediates in the catalytic cycle. The decrease in H/D-exchange rate observed in the 'Loop-swap' mutant is likely due to the overall alteration of the active-site geometry. Although the Y42F/H161A double mutant showed minimal H/D exchange barely above the no-*At*COSY control, we believe this is most likely due to its nearly abolished enzyme activity caused by structural instability (Fig. 3c, Supplementary Fig. 18). Among all mutants tested, the L374A mutant presents a unique case. Whereas the H/D exchange was completely abolished, the mutant enzyme still retains appreciable level of coumarin synthase activity, indicating a specific role of Leu374 in facilitating the proton exchange mechanism (Figs 3b-d, Supplementary Fig. 18).

Based on new insights gained from the deuterium labeling experiments, we propose a revised catalytic mechanism for COSY. His161 acts as a nonessential base that promotes the deprotonation of the *o*-hydroxy of 6OHFCoA, leading to electron

delocalization through the conjugated double bond system involving C2. The dynamic loop containing Leu374 helps position the substrate for the stereospecific protonation of C2 by a water molecule (Fig. 4a). The geometric orientation of this catalytic water molecule was confirmed by four replicates of 250 ns molecular dynamics (MD) simulations with the *At*COSY complex in the reactant state (Supplementary Fig. 22). The *s-trans* conformation of the substrate intermediate then undergoes rotary conformational change around the transient C2-C3 single bond to arrive at the *s-cis* conformation. Next, Trp371 mediates a stereospecific deprotonation of C2 through a coordinated hydroxide, and in turn elicits lactonization and the formation of the tetrahedral intermediate with its oxyanion stabilized by the indole amine of Trp371. The catalytic cycle is completed by the release of CoA and the concomitant production of scopoletin. We reason that a key aspect of the catalytic function of COSY is to lower the energy barrier for the *trans*-to-*cis* isomerization of its substrate by enabling the alternative acid-base chemistry at C2 which does not happen in the uncatalyzed spontaneous reaction.

Catalytic mechanism evaluation by density functional theory QM calculations

To further investigate the mechanistic role of key residues and the energetic feasibility of the proposed mechanism, we constructed large QM cluster models. Using the constructed QM models, we performed density functional theory (DFT) calculations on all proposed intermediates at the B3LYP/6-31G* level of theory after confirming energetics were not sensitive to functional choice (Methods and Supplementary Table

3). The large QM cluster models include all residues in contact with the scopoletin and CoA portions of the substrate (Fig. 4b). The cluster model was constructed from the scopoletin and CoA-bound crystal structures (Supplementary Fig. 23). Before calculating the energetics, we identified five stepwise intermediates between the reactant, 6OHFCoA, and the products, scopoletin and CoA (Fig. 4a). Evaluation of the QM-optimized geometries for each intermediate revealed that the *s-cis* isomer is the highest-energy intermediate with an energy of 27.6 kcal/mol relative to the reactant state (R) (Supplementary Table 3). Conversely, the product has the lowest relative energy of -22.8 kcal/mol (Fig. 4c). We also observe that the intermediate following the initial deprotonation via His161 (IM_1) is energetically uphill by only 6.6 kcal/mol. The low thermodynamic cost is in line with our mutagenesis studies, which suggest that the formation of IM_1 is not rate-limiting. To evaluate whether there would be a kinetic barrier to the deprotonation of 6OHFCoA, we performed a constrained geometry scan of the deprotonation across the N⋯H reaction coordinate. We found that the dissociation of the hydrogen is effectively barrierless and not the rate-limiting step of the mechanism (Supplementary Fig. 24). These findings support our observed biochemical data that His161 is dispensable for the overall coumarin synthase activity.

We next sought to computationally evaluate the barrier height of the *s-trans*-to-*s-cis* (IM_2-to-IM_3) isomerization and the effect of having an sp^2 or sp^3 carbon at C2. While our experimental results predict C2 protonation to be an essential preparative step to isomerization, a previously proposed mechanism suggests that isomerization may occur with the conjugated substrate in IM_1 due to electron

delocalization[8] (Supplementary Fig. 25a). Thus, we calculated barrier heights for the previously proposed IM_1-to-IM_4 isomerization and our proposed IM_3-to-IM_4 isomerization. The constrained geometry scans of the substrates in implicit solvent with dielectric constant of water show that isomerization without C2 protonation leads to a large energetic cost of 37.9 kcal/mol in addition to the 6.6 kcal/mol energy of IM_1 (Supplementary Figs. 25b,c). Conversely, C2 protonation results in a modest thermodynamic penalty of 6.3 kcal/mol with no additional activation barrier (Supplementary Figs. 25b,c). Moreover, in three replicates of 250 ns MD simulation of *At*COSY with the sp^3-substrate starting at a dihedral angle of 180º for the C1-C2-C3-C4 dihedral, which isomerized to a dihedral of approximately 80º immediately during the equilibration phase, whereas the sp^2-substrate failed to isomerize (Supplementary Fig. 26). Upon isomerization, the proposed IM_4-to-IM_5 lactonization can occur subsequently as showcased by the downhill energetics in the constrained geometry scan of the C-O bond in implicit solvent (Supplementary Fig 27). These results indicate that protonation leads to a more favorable transition and that the *s-trans*-to-*s-cis* isomerization is likely rate-limiting (Fig. 4c). We suspected that the favorability of the protonated system is due to lower double-bond character than the conjugated species. To quantify the differences in the C2-C3 bond for both species, we evaluated the Mayer bond order, a QM measurement of bonding. The Mayer bond order analysis reveals significantly stronger bond strength for the conjugated species (1.39) than the C2-protonated species (0.77). The Mayer bond orders also correlate well with their respective bond lengths of 1.38 Å

and 1.49 Å, suggesting significant double bond character that would disfavor isomerization of the conjugated species.

To better understand how the protein environment stabilizes the high-energy IM_3 intermediate, we performed hydrogen bond analysis by identifying bond critical points (BCP) from the quantum theory of atoms in molecules (QTAIM)[21,22]. The analysis reveals that a strong hydrogen bond between His161 and the substrate stabilizes the *s-cis* conformation by -20.1 kcal/mol. Additionally, a hydrogen bond that is only present in the *s-cis* isomer, between the backbone of Gly166 and the substrate, stabilizes the conformation by -8.4 kcal/mol (Supplementary Fig. 28). We also identified strong hydrogen bonds that stabilize the hydroxide species essential to the acid-base chemistry (Supplementary Fig. 28). The hydrogen bond between Thr369 and the hydroxide contributes -22.0 kcal/mol, and the hydrogen bond with Trp371 contributes -20.3 kcal/mol. The positioning of these strong hydrogen bonds allow for the necessary orientation of the hydroxide for both the IM_1-IM_2 and IM_2-IM_3 transitions (Supplementary Fig. 29). Although the *s-trans*-to-*s-cis* isomerization is the highest-energy transition, our results suggest that the COSY active site has evolved to stabilize the high-energy intermediate.

Based on our observations that the COSY active site is key for stabilizing the highest-energy intermediate, we sought to understand how the protein environment affects the favorability of the products relative to the reactant. Our calculations show that the total energetic contribution of all hydrogen bonds between the ligands and the protein environment is greater in the product state (-73.5 kcal/mol) than in the reactant

state (-51.7 kcal/mol) (Supplementary Fig. 30). To further evaluate the role of COSY in favoring product formation, we sought to compute the energetic favorability of CoA release without the protein environment. Geometry optimization of the reactant and products, without the protein environment, reveals that product formation is only energetically favorable by -2.6 kcal/mol, further illustrating the importance of the protein environment in accelerating scopoletin formation (Supplementary Fig. 31). Moreover, we explored the dispensability of His161 as a catalytic base and calculated the reaction energetics for two alternative initial steps that utilize a solvent hydroxide as the base or an intramolecular hydrogen transfer of *o*-hydroxy hydrogen to C2 of 6OHFCoA. We found that the initial deprotonation step can be performed favorably by a solvent hydroxide (Supplementary Fig. 32). We also observe that the intramolecular deprotonation by the C2-C3 double bond is a possible pathway for His161-independent reaction (Supplementary Fig. 33).

Given the importance of acid-base chemistry in the proposed mechanism, we also calculated the reaction energetics for an alternative mechanism that avoids the simultaneous formation of the cationic histidine and an anionic hydroxide species. The alternative mechanism proceeds via a hydronium carrier and adds a step to the mechanism (Supplementary Fig. 34). However, we find that using hydronium as a carrier does not increase the favorability of the mechanism due to hydronium forming significantly weaker hydrogen bonds with Thr369 and Trp371 (Supplementary Fig. 35). Thus, the first proposed mechanism remains the most likely.

Evolutionary origins of COSY and its biochemical role in question

The unique biochemical activity of COSY as a BAHD acyltransferase-fold protein and its role in coumarin biosynthesis led us to probe its evolutionary origin in plants. To do this, we first performed a maximum-likelihood phylogenetic analysis of *At*COSY together with its closely related homologs from other representative plants covering major lineages of land plants. This analysis revealed an apparent COSY clade, sister to a clade that includes the characterized *A. thaliana* spermidine disinapoyl acyltransferase (*At*SDT; AT2G23510) (Supplementary Fig. 36). The COSY clade contains homologs from both monocot and eudicot species (Fig. 5a). This is intriguing because coumarins have not been reported as widespread metabolites in monocots, whereas they are commonly found in eudicots[23]. To resolve this quandary, we performed phylogenetic analysis of feruloyl-CoA 6′-hydroxylase 1 from *A. thaliana* (*At*F6′H1), which catalyzes the first committed step of coumarin biosynthesis, together with its closely related homologs from other plants (Supplementary Fig. 37). In contrast to the phylogenetic pattern of COSY homologs, F6′H seems to have emerged in early eudicots and is absent in monocots (Fig. 5a).

To examine whether COSY homologs within the COSY clade contain coumarin synthase activity, we produced and biochemically characterized two COSY homologs from the eudicot species *Glycine max* (Soybean) and *Solanum tuberosum* (Potato), and one from the monocot species *Sorghum bicolor* (Sorghum), respectively (Supplementary Fig. 8). *In vitro* enzyme assays using recombinant proteins showed that *Gm*COSY, *St*COSY and *Sb*COSY contain coumarin synthase activity comparable to that

of *At*COSY (Figs. 5b,c). Furthermore, we predicted the structures of *Gm*COSY, *St*COSY and *Sb*COSY using AlphaFold2.0[24], and compared their structural features to *At*COSY (Supplementary Fig. 38). We noticed a similar backbone fold and consistent positioning of most of the active-site-lining residues in these predicted structures compared to the solved apo-structure of *At*COSY. Altogether, these results suggest that COSY orthologs likely existed in the last common ancestor of monocots and eudicots, and predate the later emergence of coumarin biosynthetic pathway in eudicots, marked by the occurrence of F6′H. Such phylogenomic distribution pattern thus indicates that COSY may contain some additional ancestral biochemical activity prior to its recruitment to coumarin biosynthesis in eudicots. The biochemical and biological functions of COSY orthologs in monocots will be an interesting topic for future research.

Discussion

COSY is an unusual BAHD-family enzyme. Unlike conventional BAHD acyltransferases, which catalyze transfer of the acyl group from an acyl donor molecule to an acyl acceptor molecule, COSY catalyzes an intramolecular acyl transfer reaction involving a single *o*-hydroxylated *trans*-hydroxycinnamoyl-CoA substrate containing both the acyl donor and acceptor moieties. In line with this unconventional activity, COSY has evolved a unique active-site geometry where the conventional acyl acceptor binding pocket of those typical BAHD acyltransferases is occluded by a unique active-site loop. Moreover, coumarin formation requires *s-trans*-to-*s-cis* isomerization of the hydroxycinnamoyl-CoA substrate to occur prior to subsequent lactonization. Although

trans-cis isomerization of hydroxycinnamoyl-CoAs could occur non-enzymatically, we show that it indeed constitutes the most significant energy barrier in coumarin synthesis from the *o*-hydroxylated *trans*-hydroxycinnamoyl-CoA substrate, and that the COSY active site enables an exquisite proton exchange mechanism to lower the activation energy for this isomerization step. In contrast, deprotonation of the acyl-accepting -OH group, key to the catalytic mechanism of conventional BAHD acyltransferases, is no longer a major rate-limiting step in the full reaction cycle of coumarin formation. This notion is corroborated by the fact that the ubiquitously conserved catalytic histidine among BAHD acyltransferases can be mutated in COSY without a loss of its coumarin synthase activity. The updated catalytic mechanism of COSY described in this study (Fig. 4a) therefore improves upon Vanholme et al.'s proposal[8] and reveals previously unknown aspects of the coumarin formation reaction catalyzed by COSY.

Among the active-site-lining residues examined by site-directed mutagenesis, we identified Leu374, when mutated to alanine, completely abolishes the proton exchange mechanism. Interestingly, Leu374 is anchored on a dynamic loop that undergoes conformational change throughout the catalytic cycle. We postulate that the inward swing of Leu374 helps orient the reaction intermediate IM_1 in a catalytically favorable conformation for the subsequent protonation by water through van der Waals force (Fig. 4a). The important contribution of the dynamic Leu374 to the COSY catalytic mechanism is reminiscent of the previously described Arg356 residue in *At*HCT, where its positively charged side chain serves as a dynamic catalytic handle that engages an electrostatic interaction with the carboxyl group of the acyl acceptor substrate shikimate

to help orient it towards the catalytic center for regio-selective acylation[14]. In both cases, these unique but critical structural and dynamic features found in these enzymes are highly specialized for the context of the reactions catalyzed by COSY and HCT respectively, illustrating fine-grained mechanistic details underlying functional specialization within an enzyme family. Moreover, the observation that the L374A mutant still retains some coumarin synthase activity implicates additional contributions of the COSY active site to accelerate coumarin production besides the proton exchange mechanism. These contributions seem to be dispersed across numerous active-site-lining residues as revealed by our extensive site-directed mutagenesis analyses.

Beyond *trans*-to-*cis* isomerization of the hydroxycinnamoyl-CoA substrate in the context of COSY-catalyzed coumarin synthesis, *trans-cis* isomerization of olefin-containing compounds has been described in a wide range of biological processes. Examples include protein folding that involves the peptidyl-prolyl isomerase (PPIase)[25,26], bacterial adaptation to environmental stress through the *cis-trans* isomerase of unsaturated fatty acids[27], and regulation of cytokinin levels in plants that requires the activity of zeatin *cis-trans* isomerase[28,29]. Although *trans-cis* isomerization of olefins occurs spontaneously, sometimes at appreciable rates, precise regulation of these important biological processes relies on the flux balance of the involved isomeric compounds, which likely drove the evolution of discrete classes of isomerases that catalyze their production in a controlled manner. To the best of our knowledge, among the known catalytic mechanisms of various classes of olefin *trans-cis* isomerases, the

acid-base proton exchange-based isomerization mechanism discovered in the context of COSY is unique, which exemplifies how discrete enzyme families could explore alternative catalytic mechanisms to accelerate a common type of chemical reaction.

The unconventional catalytic activity of COSY as a BAHD-family enzyme is reminiscent of the recently reported isochorismoyl-glutamate pyruvoyl-glutamate lyase (IPGL) in the Arabidopsis salicylic acid (SA) biosynthetic pathway, encoded by the *ENHANCED PSEUDOMONAS SUSCEPTIBILITY 1* (*EPS1*) gene[30]. Despite being a BAHD-family enzyme, EPS1 catalyzes the production of plant defense hormone SA from isochorismoyl-glutamate, a reaction that otherwise occurs spontaneously at a slower rate[31]. Similar to the case of COSY, the active site of EPS1 has evolved unique geometry and dynamism to accommodate the highly specialized IPGL activity[31]. Unlike COSY, EPS1 further harbors substitutions to the highly conserved catalytic histidine and tryptophan residues, suggesting its derived catalytic function has entirely deviated from the ancestral acyltransferase activity[31]. Both EPS1 and COSY illustrate the mechanistic plasticity of the BAHD protein fold, which permits reconfiguration of its ancestral active site to access alternative catalytic functions through adaptive mutations. Reconstruction of such plausible mutational trajectories is an interesting topic for future research. It is worth noting that similar research has been previously pursued on the chalcone isomerase (CHI) fold proteins, which demonstrated the feasibility of arriving at the enantioselective CHI activity through laboratory evolution starting from a noncatalytic progenitor scaffold[32–34].

Several questions regarding the biochemical activities and biological functions of COSY have yet to be answered. We show that although the catalytic histidine and tryptophan remain unmutated among COSY orthologs, they are apparently nonessential to the coumarin synthase activity of COSY. Interestingly, our phylogenetic analyses suggest that the emergence of COSY predates the emergence of the coumarin biosynthetic pathway, which developed as an offshoot of the general phenylpropanoid metabolism only in eudicots. Could COSY still retain some ancestral activity outside the context of coumarin biosynthesis? What are the biological functions of COSY orthologs in monocots where coumarin biosynthesis is absent? Moreover, our study of COSY identified its enzymatic activity in converting *trans*-phenylpropanoids (in the form of CoA thioester) to their *cis* form. This observation raises the question regarding the biosynthetic origins of various *cis*-phenylpropanoids previously reported in plants[35,36]. For example, *cis*-cinnamic acid occurs in many plants and acts as an auxin transport inhibitor that promotes lateral root growth[37–39]. In addition, *cis*-*p*-coumaroylagmatine found in *Albizzia julibrissin* Durazz was reported to contain nyctinasty-inducing activities[35]. Although the occurrence of *cis*-phenylpropanoids has long been attributed to spontaneous *trans*-*cis* isomerization, perhaps enzymes like COSY are yet to be discovered that catalyze these physiologically important isomerization reactions.

Methods

Materials and general methods.

Scopoletin, umbelliferone, 2OHpCA were obtained from Sigma Aldrich. 6OHFA was synthesized by the saponification of scopoletin with 2 M NaOH solution following the previously characterized protocol[8]. All reagents used in biochemical assays were obtained from Sigma Aldrich. Deuterium oxide was obtained from Cambridge Isotopes Laboratories. All other reagents in this study were obtained from Thermo Fisher Scientific.

Protein expression and purification.

*At*COSY (AT1G28680) was amplified from root cDNA of *A. thaliana* and cloned into protein overexpression vector pHis8-4 containing an *N*-terminal 8xHis tag followed by a tobacco etch virus (TEV) cleavage site (Supplementary Table 4). The verified pHis8-4b::*AtCOSY* construct was transformed into *E. coli* strain BL21 (DE3) for recombinant protein expression. A 1L of terrific broth (TB) medium containing 50 µg/mL kanamycin was inoculated with 30 mL of an overnight starter culture. The resulting culture was grown with shaking at 200 rpm at 37 ºC to an OD_{600} of 0.6-0.8. Then protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by cold shock of the medium and subsequent growth with shaking at 200 rpm (18 °C for 18 h).

Cultures were harvested by centrifugation and the resulting cell paste (~10 g/L) was resuspended in lysis buffer (100 mM HEPES pH 8.0, 200 mM NaCl, 20 mM

imidazole, 10% (vol/vol) glycerol, 1 mM dithiothreitol) containing 1 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride. Cells were lysed via five passes through an M-110L microfluidizer (Microfluidics) and the resulting crude protein lysate was clarified by centrifugation at 19,000 x *g* for 1 hour. Then the clarified lysate was loaded onto QIAGEN nickel-nitrilotriacetic acid (Ni-NTA) gravity flow chromatographic column. After loading the clarified lysate, the Ni-NTA resin was washed with 10 column volumes of lysis buffer, followed by 10 column volumes of lysis buffer with 5 mM ATP. Then the purified 8xHis-tagged proteins are eluted with 1 column volume of elution buffer (100 mM HEPES pH 8.0, 200 mM NaCl, 250 mM imidazole, 10% (vol/vol) glycerol, 1 mM dithiothreitol). To cleave the 8xHis-tag from purified protein, 1 mg of His-tagged TEV protease was added, followed by dialysis at 4 °C for 16 h in dialysis buffer (25 mM HEPES pH 8.0, 200 mM NaCl, 5% (vol/vol) glycerol, 0.5 mM EDTA, 0.5 mM dithiothreitol). After dialysis, the protein solution was passed through Ni-NTA resin to remove uncleaved protein and His-tagged TEV. The recombinant proteins were further purified by gel filtration on an ÄKTA Pure fast protein liquid chromatography (FPLC) system (GE Healthcare Life Sciences). The principal peaks were collected, verified by SDS polyacrylamide gel electrophoresis and dialyzed into a storage buffer (20 mM HEPES pH 7.5 and 5%(vol/vol) glycerol). Finally, proteins were concentrated to >10 mg/mL using Amicon Ultra-15 Centrifugal Filters (Millipore). For some of the mutants assayed in this study, 8xHis-tags were not cleaved, as this led to higher yield of purified proteins. Similar process was performed for the purification of *At*4CL1 (AT1G51680), and *At*HCT (AT5G48930). For construction and purification of COSY

homologs *Gm*COSY, *St*COSY, and *Sb*COSY, the corresponding gene fragments were synthesized Integrated DNA Technologies (IDT) and followed a similar protocol (Supplementary Table 6).

Crystallization and X-ray structure determination.

Crystals for *At*COSY-apo, scopoletin-bound, umbelliferone-bound, and CoA-bound were grown at 25 ºC by hanging-drop vapor diffusion method with the drop containing 2 µL of protein sample and 2 µL of reservoir solution at a reservoir solution volume of 500 µL. The crystallization buffer for all of the *At*COSY structures were composed of 50 mM calcium acetate, 0.1 M sodium cacodylate pH 6.5, and 20% glycerol. The protein sample concentration was 5 mg/mL stored in 20 mM HEPES pH 7.5. The ligand bound *At*COSY drop also contained 500 µM ligands. No additional cryo-protectant was added. The *At*COSY-apo structure was determined first by molecular replacement using the native HCT structure from *C. canephora* (PDB:4G0B) as the search model via the EMBL-HH Auto-Rickshaw[40,41]. The resulting model was iteratively refined using Refmac 5.2[42] and then manually refined in Coot 0.7.1[43]. For the subsequent *At*COSY structures, Phaser-MR[44] was used using the *At*COSY-apo structure as the mode and refined using RefMac 5.2[42] and Coot 0.7.1[43]. All crystallographic diffraction data sets were collected at beamlines 24-ID-C and 24-ID-E of the Advanced Photon Source at the Argonne National Laboratory by single-wavelength anomalous diffraction methods.

Site-directed mutagenesis and *in vitro* biochemical assays.

*At*COSY mutants were generated according to the protocol described in the QuickChange II Site-Directed Mutagenesis Kit (Agilent Technologies) using plasmid pHis8-4b::*At*COSY as the template and the primer sequences in Supplementary Table 5. pHis8-4b::*At*COSY-H161A was used as the template for generating Y42F/H161A double mutant. For the construction of active-site-loop insertion domain swap mutant, we replaced the DNA sequences encoding for ^{392}LSNKLLGSMEPC^{403} in *At*COSY with that of the loop region in *At*HCT, ^{392}GGIPYEGLS^{400}. The resulting gene fragment was synthesized and purchased from Twist Biosciences and was cloned into protein overexpression vector pHis8-4 (Supplementary Table 6). The resulting mutant plasmid constructs were verified by sequencing. Recombinant mutant protein production and purification were carried out following the same procedure as described above.

Each *in vitro* biochemical assay for *At*COSY and its mutants with 6OHFA, D2-6OHFA, or 2OHpCA was carried out in 20 mM HEPES were performed with 1 mM hydroxycinnamic acid substrate, 1 mM ATP, 1 mM CoA, 0.5 mM $CaCl_2$ and 5 mM $MgCl_2$. 10 μM of *At*4CL1 was added to each assay 3 minutes before the addition of the respective enzyme at 2 μM. All assays were performed with a 15 minute reaction period and subsequent 20 mM Urea quench. Each assay panel had an *At*4CL1-only and no enzyme control, along with three replicates of each target sample. Similar buffer conditions were used for the *At*HCT assay except 1 mM *p*-coumaric acid and 1 mM shikimic acid were used as substrates instead of hydroxycinnamic acid.

For enzyme assay in D_2O, the WT *At*COSY reaction was carried out following the same procedure as described above, except the assay was conducted in 66% D_2O.

66% D_2O was used as the stock solutions of the buffer components were already prepared in H_2O and 100% D_2O was not achievable for the *in vitro* assay. For the purification of umbelliferone-*d1*, we scaled-up this reaction with 2OHpCA. The resulting assay solution was concentrated to 1 mL and purified by preparative reverse-phase HPLC (Shimadzu Preparative HPLC with LC-20AP pump and SPD-20A UV–VIS detector) using Kinetex 5µ C18 100 A, 150 × 21.2 mm column, solvent A-(H2O-0.1% TFA), solvent B-(CH3CN-0.1% TFA), 0–50 min 40–80% B, 10 mL/min) to provide umbelliferone-*d1* as a pale yellow oil. ^{1}H NMR spectra of umbelliferone and umbelliferone-*d1* were obtained using a Bruker Avance Neo 500 MHz spectrometer equipped with a 5 mm liquid nitrogen cooled Prodigy BBO cryoprobe.

Liquid-chromatography and mass-spectrometry

LC was conducted on a Vanquish Flex Binary UHPLC system (Thermo Fisher Scientific) using water with 0.1% formic acid as solvent A and acetonitrile with 0.1% formic acid as solvent B. Reverse phase separation of analytes was performed on a Kinetex C18 column, 150 × 3 mm^2, 2.6 µm particle size (Phenomenex). The column oven was held at 35 °C. Most injections were eluted with 5% B for 0.5 min, a gradient of 5–50% B for 6.5 min, 50-75% B for 0.5 min, 70% B for 0.5 min and 5% B for 2.0 min, with a flow rate of 0.5 mL/min. Most MS analyses were performed on a high-resolution Orbitrap Exploris 120 benchtop mass spectrometer (Thermo Fisher Scientific) operated in negative ionization mode with full scan range of 100–300 m/z and top four data-dependent MS/MS scans. The orbitrap resolution is 120,000 with RF lens of 70% and static spray

voltage of 3500 V. Initial enzyme activity assays were analyzed with a TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific) using multiple reaction monitoring at m/z transitions corresponding to $[M-H]^-$ of 6OHFA (165.1 *m/z* to 150.1 *m/z*) and $[M-H]^-$ of scopoletin (191.1 *m/z* to 176.1 *m/z)* with scan width of 0.5 and collision energy of 20 V. As 6OHFA and (D2)-6OHFA are readily fragmented under our MS collection methods, we monitored the abundance of their major fragment ions 165.05556 *m/z* and 167.06804 *m/z*, respectively, under negative ionization mode. For detection of D_2O labeling, single ion monitoring was used to scan the following *m/z* values: 161.05, 162.05, 165.03, 191.03, 192.04, 193.05, 209.05 with scan width of 0.5 and collision energy of 20 V. The MS analysis for quick detection of enzyme assay products was conducted on the TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific) operated in negative ionization mode with a full scan range of 100-300 *m/z*. Raw LC-MS data were analyzed using XCalibur (Thermo Fisher Scientific).

Classical Molecular Dynamics Simulations

The CoA-bound crystal structure (8DQR) was used for all MD simulations. The reactive state of the substrate was modeled using the conformation of scopoletin from the scopoletin-bound crystal structure (8DQP), which was then manually attached to CoA from 8DQR. Protonation states were added using the H++ webserver[45–47] with a pH of 7.0 and internal dielectric of 10 with all other defaults applied, while missing residues were added using Modeller[48]. The topology and coordinate files were prepared in

AMBER using tleap with the ff14SB force field[49]. The substrate was treated with general AMBER force field (GAFF) parameters and restrained electrostatic potential (RESP) charges[50] calculated with Gaussian16 at the HF/6-31G* level of theory[51] with the RED RESP server[52]. All systems were solvated with 15 Å periodic TIP3P water boxes and neutralized with Na^+ counterions[53]. The simulations were performed with the AMBER18 GPU-accelerated Particle Mesh Ewald molecular dynamics (PMEMD) code (pmemd.cuda)[54,55]. The Equilibration protocol is as follows: i) hydrogen atom minimization (1000 steps), sidechain minimization with a fixed backbone (2000 steps), and unrestrained minimization (2000 steps). ii) Controlled *NVT* heating was performed from 0 to 300 K for 10 ps using the Langevin thermostat and a collision frequency of 5.0 ps^{-1}, iii) and 1 ns of *NpT* simulation using the Berendsen barostat and a 2ps relaxation time. Following equilibration, 250 ns of production *NpT* dynamics were collected with SHAKE and 2 fs time steps.

QM Cluster Model Preparation

To investigate the energetic favorability of the proposed reaction mechanism, we performed geometry optimizations of large QM cluster models of the proposed intermediates[56]. While the crystal structures do not indicate the position of the reactant, we reconstructed 6OHFCoA by linking the substrates in the scopoletin-bound and the CoA-bound crystal structures after alignment with ChimeraX[57]. The scopoletin lactone ring was opened by changing the dihedral centered around the C2-C3 bond from *s-cis* to *s-trans*. The C1 carbonyl carbon was then connected to the thiol sulfur of CoA and

the thiol hydrogen was removed. The manually adjusted atoms were then optimized with the MMFF94 force field with all other atoms fixed to conserve the crystallographic positions in preparation from QM cluster optimization. No 6OHFCoA atoms involved in substrate-protein contacts were modified during this process.

We next constructed a large QM cluster model that included all side-chains and backbone atoms in contact with the reactant substrate. The final cluster model included twenty amino acids, which corresponded to three residue chains: 40-42, 159-169, and 369-374. Full residues, including sidechain and backbone atoms, were Phe40, Tyr42, Ile159, His161, Cys164, Gly166, Leu167, Gly168, Ala169, Thr369, Asp370, Trp371, Arg372, Tyr373, and Leu374. Residues whose side chains did not interact with the substrate but whose backbone did, were substituted with glycine to minimize the size of the cluster model. Since backbone atoms were fixed to maintain the conformation of the active site, alanine and glycine substitutions are effectively the same. The substituted residues include Arg41, His160, Ala162, Ile163, and Asp165. A key water molecule proposed to be important in acid-base chemistry with the substrate, was modeled between Thr369 and Trp371. In total, the resulting cluster model contained 395 atoms. The substrate and protein environment both had charges of zero and maintained net charge neutrality. The starting QM cluster models are provided in the Source data. The backbone atoms for the terminal residues of each chain were capped with hydrogen atoms.

QM Cluster Model Calculations

QM cluster geometry optimizations and constrained geometry scans were performed with developer version 1.9 of the GPU-accelerated quantum chemistry package TeraChem[58–60]. All calculations were carried out at the B3LYP/6-31G* level of theory. Additional calculations were carried out at the ωPBEh/6-31G* level of theory to verify that the results were insensitive to the functional (Supplementary Table 3). Given that the results were comparable for the range-separated hybrid functional, we will discuss the B3LYP results in the text. A moderate level of theory was selected to balance cost accuracy trade-offs and allow for a large QM cluster model that could account for all interactions between coumarin, CoA, and the protein environment. Given the large size of the cluster model, zero-point vibrational energy corrections were not computed and electronic energies are instead reported in the text. It is expected that this has a limited effect on conclusions from computational modeling because the computed differences between intermediates were large. To account for long-range interactions within the QM cluster model, semi-empirical DFT-D3 with default Becke-Johnson damping was also applied[60]. The calculations were carried out with implicit conductor-like polarizable continuum model (C-PCM) with epsilon set to 10 to approximate the electronic environment of the binding pocket[61]. To conserve the overall structure and crystallographic conformation of the active site during the optimization, backbone carbons were frozen. All other atoms were allowed to move freely. Putative transition states were identified using constrained geometry scans as implemented in TeraChem. During the scans, a selected reaction coordinate was fixed while optimizing all other degrees of freedom. In this implementation of the constrained geometry scan, each

subsequent step uses the geometry of the previous step as input and the previous wavefunction as an initial guess to improve continuity. The strength of the hydrogen bonds present in each intermediate were estimated based on bond critical points (BCP)[62] from the quantum theory of atoms in molecules (QTAIM)[21] and the potential energy density calculated using the software package Multiwfn[63]. To better understand the double bond character of the C2-C3 bond involved in the trans to cis isomerization, we performed Mayer bond order analysis using Multiwfn[22].

Chemical synthesis and spectral verification of

(*E*)-3-(2,4-Dihydroxy-5-methoxyphenyl-3-*d*)acrylic-2-*d* Acid (D2-6OHFA)

All reagents and solvents were used as supplied without further purification. Column chromatography was conducted using a Buchi Pure C-815 Flash Automated Chromatography System equipped with Buchi FlashPure EcoFlex SiO_2 prepacked columns (40–63 μm). Analytical TLC was conducted using Millipore $SiO_2$60 F_{254} TLC (0.250 mm) plates. ^{1}H and ^{13}C NMR spectra were obtained using a Bruker Avance Neo 500 MHz spectrometer equipped with a 5 mm liquid nitrogen cooled Prodigy BBO cryoprobe. IR spectra were obtained using a Bruker Alpha 2 with a Platinum ATR accessory. Mass spectrometric analysis was performed on a ThermoFisher Scientific Orbitrap Exploris 120.

A vigorously stirred 2.0 M solution of NaOH in D_2O (750 mL) at 23 °C was treated with scopoletin (25 mg, 0.13 mmol) [8] (Supplementary Fig. 12a). The reaction mixture was then warmed to 80 °C and stirred for 1 h. After 1 h, the reaction mixture was cooled

to 23 °C, acidified to pH 5 by dropwise addition of 1.0 M HCl(aq), then extracted with EtOAc (3 × 1 mL). The combined organic extracts were dried over Na_2SO_4 and concentrated on a rotary evaporator. Flash chromatography (SiO_2, 20–100% EtOAc/hexanes + 0.1% AcOH) provided **3** as a pale yellow oil (7.8 mg, 28%): ^{1}H NMR (acetone-d6, 500 MHz) δ 7.98 (s, 1H), 7.19 (s, 1H), 3.85 (s, 3H); ^{13}C NMR (acetone-d6, 125 MHz) δ 172.2, 168.9, 152.9, 151.1, 142.6, 141.0, 114.9, 114.8, 114.6, 114.4, 113.3, 111.8, 104.1, 104.0, 103.8, 103.6, 56.9; IR (film) ν_{max} 3267, 1661, 1582, 1495, 1455, 1433, 1268, 1199, 1121, 1061, 1038, 961, 666, 539, 515 cm^{-1}; HRMS (H-ESI-Orbitrap Exploris 120) *m/z* 211.05806 ($C_{10}H_8D_2O_5$ - H^+ requires 211.05755).

***In planta* deuterium exchange**

A. thaliana wild type (Col-0) and *cosy* mutant (*AT1G28680*) seeds were obtained from Arabidopsis Biological Resource Center. All seeds were sterilized by washing with 70% (v/v) ethanol and 0.1% Tween 20, then with 100% ethanol. Seeds were subsequently washed with water to remove residual ethanol before vernalization at 4 ºC for 2 days. The vernalized seeds were planted on Murashige and Skoog/agar media (half strength, 0.9% agar) prepared in 0% D_2O or 50% D_2O. 50% D_2O was used because *A. thaliana* plants did not grow on media prepared in 100% D_2O. Plants were grown under a 16hr-light/8-hr-dark cycle with fluorescent light of 80 to 120 µmol photons m^{-2} s^{-1} at 22°C and 60% RH. The roots of Col-0 and *cosy* plants were harvested after 20 days and flash frozen under liquid nitrogen. The frozen samples were homogenized on a TissueLyser II (QIAGEN) at 25 s^{-1} for 2 min. Metabolites were extracted using 100

volumes (dry w/v) of 50% MeOH at room temperature for 15 min. Extracts were centrifuged once (13,000 x *g*, 10 min) to pellet plant tissue, and the supernatants were filtered using 0.2 µm filter vials (Thomson Instrument Company) for LC-HRAM-MS analysis.

Phylogenetic analysis of COSY and F6′H.

The amino acid sequences for phylogenetic analysis were obtained from curated plant proteomes in the Phytozome v12.1.6 database. pBLAST searches with *At*COSY or *At*F6′H1 as query sequences were conducted and top 5 full length protein sequences were collected from each of the curated plant proteomes covering the angiosperm phylogeny and selected pre-angiosperm species. Multiple sequence alignments were built using the MUSCLE[64] algorithm in MEGA X[65]. Evolutionary histories were inferred using the maximum-likelihood method on the basis of the JTT matrix-based model in MEGA X[65]. Bootstrap statistics were calculated using 200 replicates. Similar procedure was followed for multiple sequence alignment of COSY and HCT homologous sequences, which was visualized using ESPript 3[66]. All alignment files can be found under Source Data.

Acknowledgments

We thank the staff at the CCP4/APS School for Macromolecular Crystallography for assistance with x-ray data collection and analysis. This work was supported by the Keck Foundation (J.K.W.), the Family Larsson-Rosenquist Foundation (J.K.W.), the Beckman

Young Investigator Program (J.K.W.), a Burroughs Wellcome Fund Career Award at the Scientific Interface (H.J.K., D.W.K), and a National Science Foundation Graduate Research Fellowship under Grant #1745302 (D.W.K.).

Accession codes

Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 8DQO (COSY Apo), 8DQP (COSY + scopoletin), 8DQQ (COSY + umbelliferone), and 8DQR (COSY + CoA).

Author contributions

C.Y.K., A.J.M., and J.K.W. designed the research. C.Y.K., A.J.M., M.G., and J.K.W. generated and processed the crystallographic data. C.Y.K., A.J.M., C.E.A. and M.G. performed the cloning, protein purification, and biochemical assays. C.Y.K. and C.E.A. conducted the *in planta* deuterium feeding study and phylogenetic analyses. C.M.G. performed chemical synthesis and characterization. D.W.K. and H.J.K. conducted and analyzed the QM cluster models and MD simulations. C.Y.K., D.W.K., H.J.K., and J.K.W. interpreted the results and wrote the manuscript. All authors reviewed the manuscript.

Competing financial interests

J.K.W. is a member of the Scientific Advisory Board and a shareholder of DoubleRainbow Biosciences, Galixir and Inari Agriculture, which develop

biotechnologies related to natural products, drug discovery and agriculture. All other authors have no competing interests.

Figures

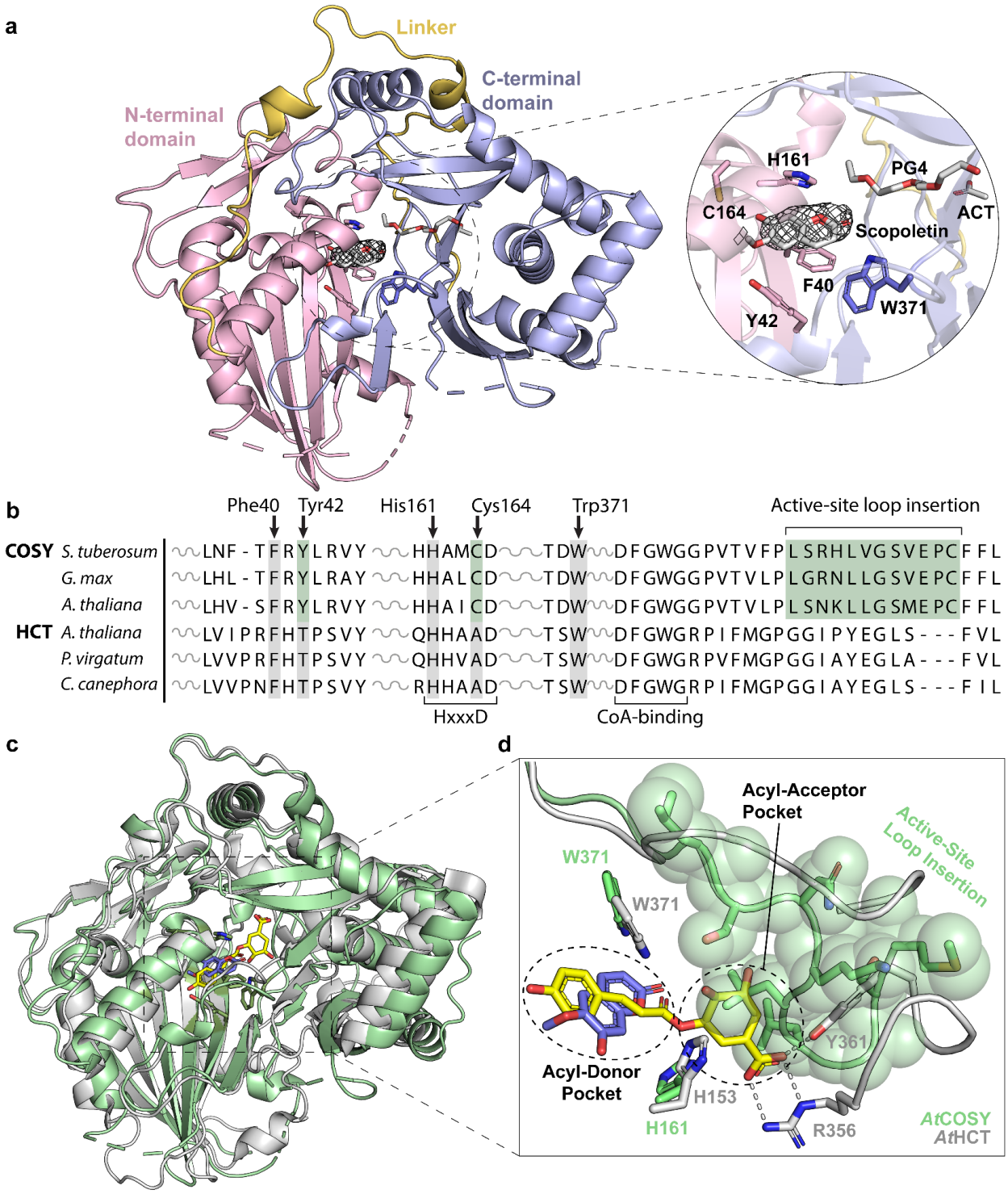


Fig. 1 | Structural and sequence characterization of COSY in comparison with HCT. (**a**) Scopoletin-bound *At*COSY at 2.5 Å resolution. The overall fold showcases a pseudo-symmetric N-terminal (residues 1-181; light pink) and C-terminal (residues 230-451; light blue) domains, connected by a linker loop (residues 182-229; yellow).

The conserved BAHD catalytic residues (His161 and Trp371), along with product stabilizing residues (Phe40, Tyr42, and Cys164) are showcased within the active site pocket. Tetraethylene glycol (PG4) and acetate ion (ACT) are displayed. The mFo – DFc simulated annealing omit map with electron density map around SCO ligand is contoured at 1.0 σ. (**b**) Multiple sequence alignment of COSY orthologs from *Solanum tuberosum*, *Glycine max*, and *A. thaliana* with HCT orthologs from *A. thaliana*, *Panicum virgatum* and *Coffea canephora*. The full sequence alignment is shown in Supplementary Fig. 4. (**c**) Structural alignment of *At*COSY (green) and *At*HCT (PDB: 5KJU; gray). The product of *At*HCT, *p*-coumaroyl shikimate, is shown in yellow. Scopoletin is shown in purple. (**d**) Structural comparison between *At*COSY and *At*HCT reveals an active site loop insertion (residues 390-406) in *At*COSY that occludes the corresponding acyl acceptor pocket in *At*HCT.

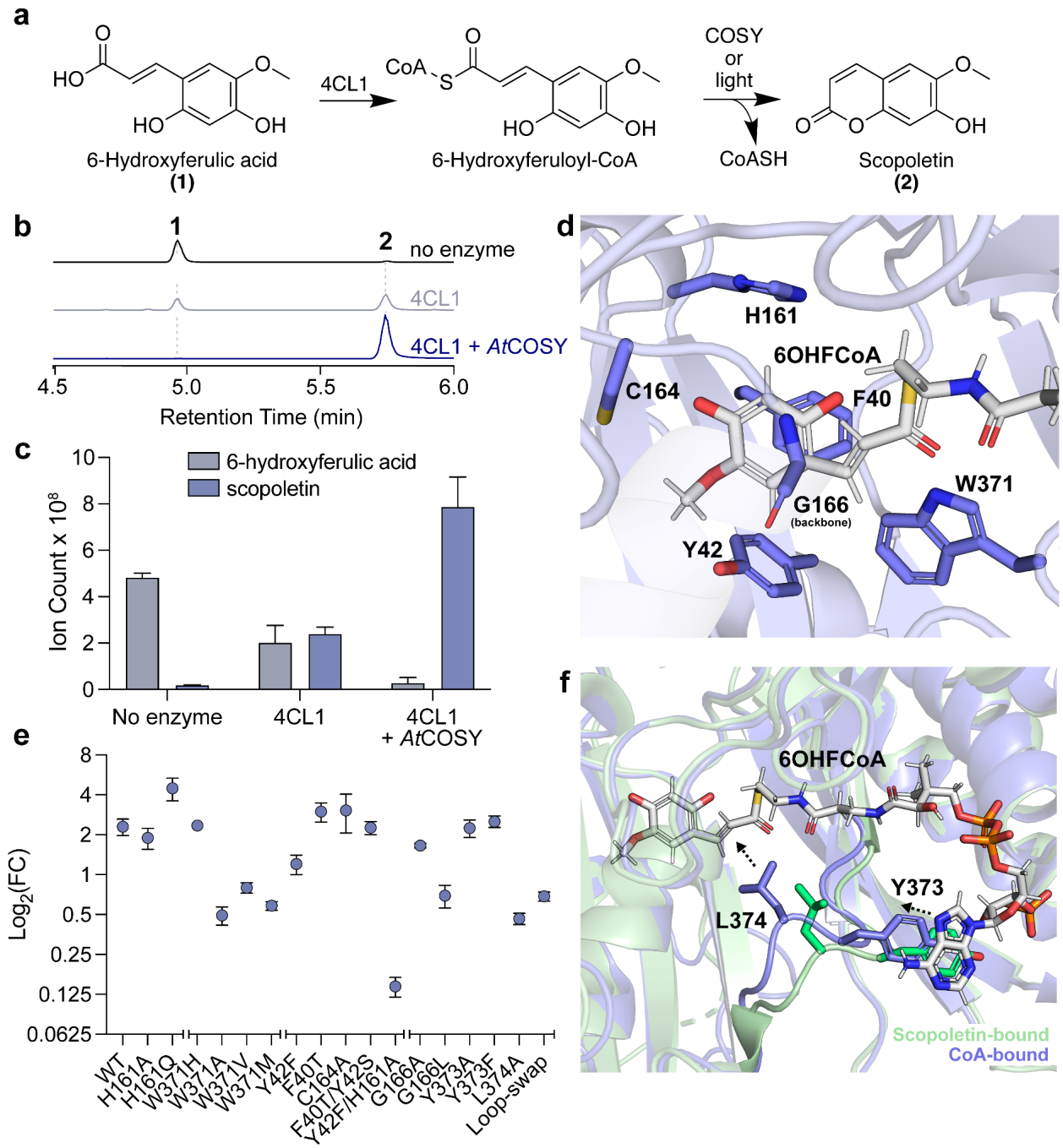


Fig. 2 | *In vitro* coupled biochemical assay that examines the coumarin synthase activity of *At*COSY. (**a**) Reaction schematic of the conversion of 6-hydroxyferulic acid (6OHFA) to scopoletin using 4-coumarate-CoA ligase 1 from *A. thaliana* (*At*4CL1) and coumarin synthase (*At*COSY). (**b**) Combined LC-MS extracted ion chromatograms (XICs) of 165.05556 m/z and 209.04549 m/z; **1**, and 191.03443 m/z; **2**. As scopoletin production is spontaneous, assay carried out in absence of *At*COSY serves as the baseline control for the coumarin synthase activity measurement. 4CL1+COSY sample shows complete conversion of 6OHFA into scopoletin. (**c**) Total ion count of 6OHFA and scopoletin from LC-MS for the enzymatic assays in triplicates. (**d**) Reconstructed pose

of *trans*-6OHFCoA in the active site of the CoA-bound *At*COSY structure. Phe40, Tyr42, His161, Cys 164, Trp371, and backbone atoms of Gly166 are highlighted as they make close contact with the substrate. (**e**) Relative coumarin synthase activity of wild-type and mutant COSY. Log_2 fold-change between the LC-HRAM-MS peak areas of coumarin ($[M-H]^-$ = 191.03443 m/z) in 4CL1+*At*COSY (WT or mutant) samples and that of 4CL1-only samples are shown. We performed the mutant assays for the following active site mutants (H161A, H161Q, W371H, W371A, W371V, W371M, Y42F, F40T, C164A, F40T/Y42S, and Y42F/H161A). All assays were performed in triplicates and the error bars represent standard error of the mean. (**f**) Structural alignment of CoA-bound (blue) and scopoletin-bound (green) *At*COSY structures highlighting the dynamic loop containing Tyr373 and Leu374. Arrows indicate change in orientation of Tyr373 and Leu374 in CoA-bound structure compared to scopoletin-bound structure.

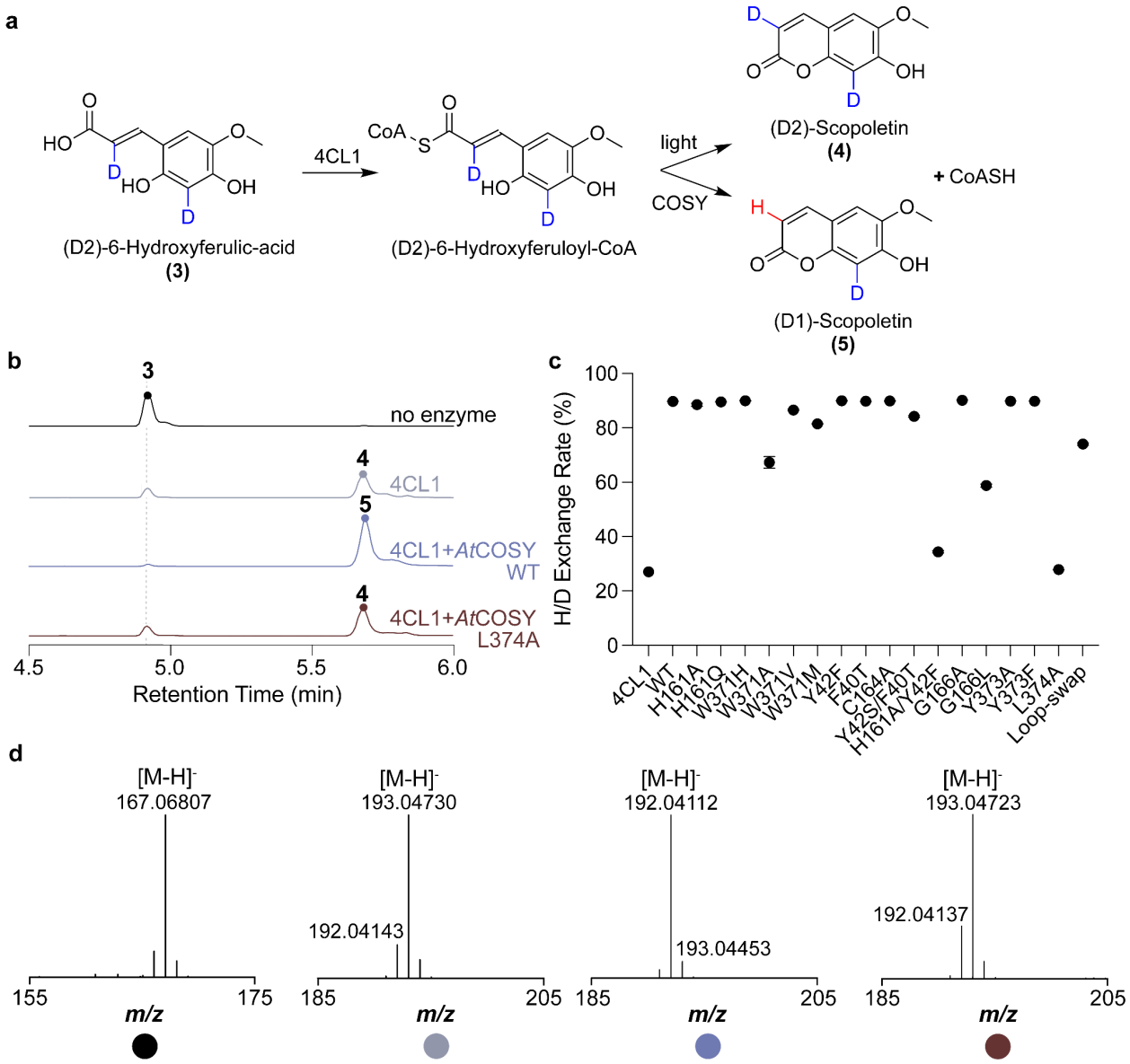


Fig. 3 | *In vitro* coupled enzyme assay that examines the H/D exchange at C2 of the (D2)-6OHFA substrate. (**a**) Reaction schematic for the conversion of **3**, (D2)-6-Hydroxyferulic-acid into **4**, (D2)-Scopoletin by light and **5**, (D1)-Scopoletin by COSY. (**b**) Combined LC-MS extracted ion chromatograms (XICs) of 167.06804 m/z; **3**, 193.04747 m/z; **4** and 192.04114 m/z; **5**. The colored circles represent the relevant substrate or product peaks in each XIC trace corresponding to the compound number in **a**. (**c**) H/D exchange rate of assays conducted with no COSY, wild-type COSY, and various COSY mutants. H/D exchange rate is defined as the percentage of **5** out of the total ion intensity for *m/z* values corresponding to **4** and **5**. All assays were performed with 4CL1 in duplicates with the error bars representing the standard error of the mean. (**d**) MS^1 spectra of the relevant peaks from assays shown in **b** corresponding to compounds highlighted in **a**.

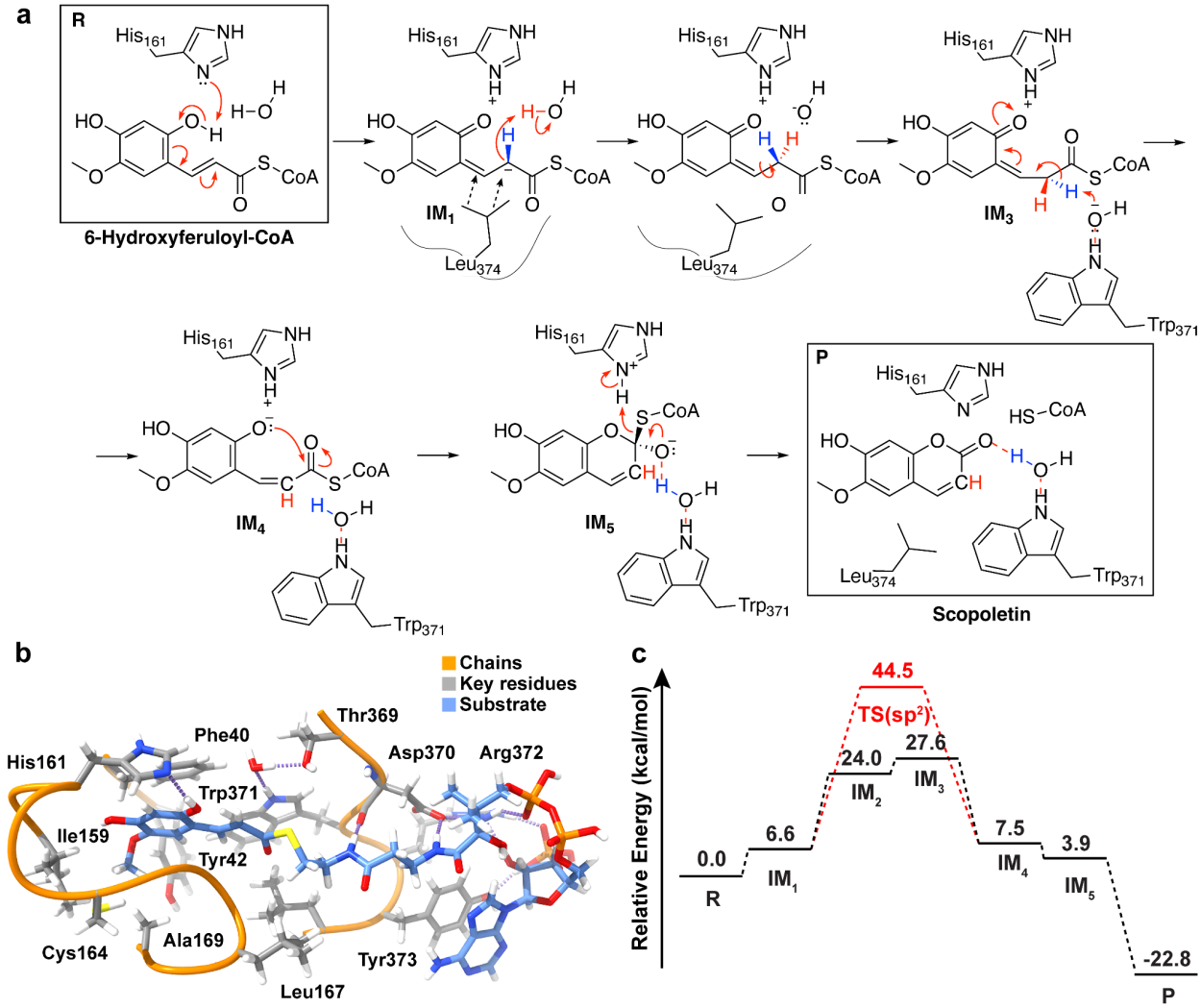


Fig. 4 | Computational insights into the proposed intermediates and reaction energetics of COSY-catalyzed coumarin synthesis. (**a**) A schematic of the proposed stepwise mechanism of COSY-catalyzed coumarin synthesis from 6OHFCoA. The reaction begins with deprotonation of the 6-OH of 6OHFCoA by His161, followed by delocalization of electrons through the conjugated double bond system involving C2 of 6OHFCoA. Then, Leu374 helps orient IM_1 towards the catalytic water molecule through van der Waals force, leading to the addition of a proton at C2. The C2-C3 bond then undergoes isomerization from the *s-trans* to the *s-cis* configuration. Following isomerization, lactonization and release of CoA lead to the generation of scopoletin. The oxyanion of IM_5 is depicted to be stabilized by a water molecule coordinated by Trp371, although the indole nitrogen of Trp371 may stabilize the oxyanion directly. Hydrogens that are relevant to stereoselective proton exchange are colored in red and blue. (**b**) The QM optimized large cluster model in the reactant state. Continuous chains of amino acids are shown as orange ribbons and included side-chains are shown with gray

carbons. The substrate is depicted with blue carbons. Hydrogen bonds are depicted as dotted purple lines. (**c**) The overall energetics for all intermediates of the proposed mechanism. Reactants are labeled as R and products are labeled as P. A putative transition state for the reaction mechanism in the absence of C2-protonation is illustrated in red.

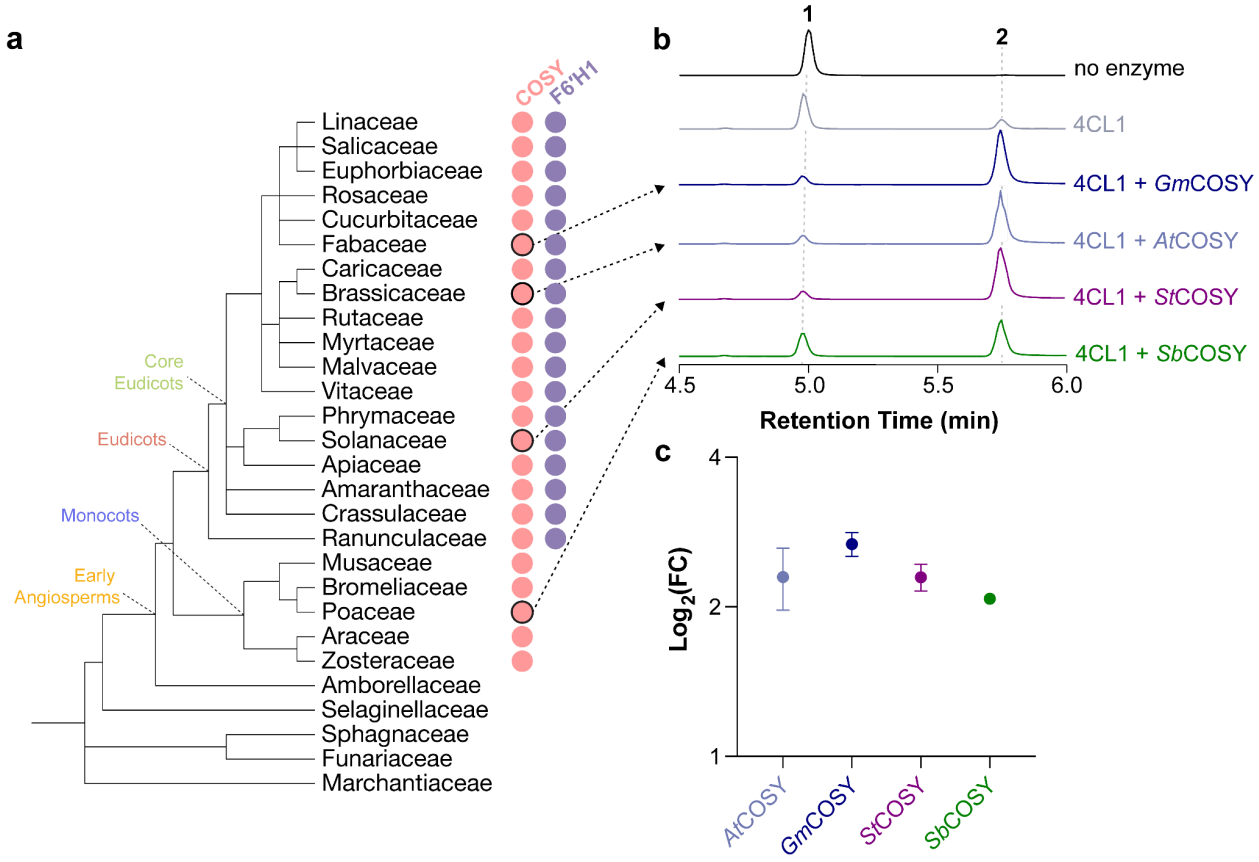


Fig. 5 | Evolutionary origins of COSY in land plants. (**a**) Presence of COSY (pink circles) and F6′H1 (purple circles) orthologs among major lineages of land plants. This information is inferred from maximum-likelihood phylogenetic analyses of COSY and F6′H1 together with their homologs from representative plants with sequenced genomes (Supplementary Figs. 28 and 29). (**b**) Combined LC-MS XICs of 165.05556 m/z; **1**, 209.04549 m/z; **1** and 191.03443 m/z; **2** showcasing the *in vitro* coumarin synthase activity of COSY orthologs from *G. max* (Soybean), *A. thaliana* (Arabidopsis), *S. tuberosum* (Potato), and *S. bicolor* (Sorghum). (**c**) Relative coumarin synthase activity of the four COSY orthologs quantified by the LC-HRAM-MS peak areas derived from the *in vitro* enzyme assay as shown in **b**. All assays were performed in triplicates and the error bars represent standard error of the mean.

References

1. Musa, M. A., Cooperwood, J. S. & Khan, M. O. F. A review of coumarin derivatives in pharmacotherapy of breast cancer. *Curr. Med. Chem.* **15**, 2664–2679 (2008).
2. Robe, K., Izquierdo, E., Vignols, F., Rouached, H. & Dubos, C. The Coumarins: Secondary Metabolites Playing a Primary Role in Plant Nutrition and Health. *Trends Plant Sci.* **26**, 248–259 (2021).
3. Stassen, M. J. J., Hsu, S.-H., Pieterse, C. M. J. & Stringlis, I. A. Coumarin Communication Along the Microbiome–Root–Shoot Axis. *Trends in Plant Science* vol. 26 169–183 Preprint at https://doi.org/10.1016/j.tplants.2020.09.008 (2021).
4. Mishra, S., Pandey, A. & Manvati, S. Coumarin: An emerging antiviral agent. *Heliyon* **6**, e03217 (2020).
5. Sahoo, C. R. *et al.* Coumarin derivatives as promising antibacterial agent(s). *Arabian Journal of Chemistry* vol. 14 102922 Preprint at https://doi.org/10.1016/j.arabjc.2020.102922 (2021).
6. Lowenthal, J. & Birnbaum, H. Vitamin K and coumarin anticoagulants: dependence of anticoagulant effect on inhibition of vitamin K transport. *Science* **164**, 181–183 (1969).
7. Kai, K. *et al.* Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in Arabidopsis thaliana. *Plant J.* **55**, 989–999 (2008).
8. Vanholme, R. *et al.* COSY catalyses trans–cis isomerization and lactonization in the biosynthesis of coumarins. *Nature Plants* vol. 5 1066–1075 Preprint at

https://doi.org/10.1038/s41477-019-0510-0 (2019).

9. Dudareva, N., D'Auria, J. C., Nam, K. H., Raguso, R. A. & Pichersky, E. Acetyl-CoA:benzylalcohol acetyltransferase--an enzyme involved in floral scent production in Clarkia breweri. *Plant J.* **14**, 297–304 (1998).
10. Fujiwara, H. *et al.* Anthocyanin 5-aromatic acyltransferase from Gentiana triflora. Purification, characterization and its role in anthocyanin biosynthesis. *Eur. J. Biochem.* **249**, 45–51 (1997).
11. Yang, Q., Reinhard, K., Schiltz, E. & Matern, U. Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, Dianthus caryophyllus L. *Plant Mol. Biol.* **35**, 777–789 (1997).
12. Power, R., Kurz, W. G. & De Luca, V. Purification and characterization of acetylcoenzyme A: deacetylvindoline 4-O-acetyltransferase from Catharanthus roseus. *Arch. Biochem. Biophys.* **279**, 370–376 (1990).
13. D'Auria, J. C. Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* vol. 9 331–340 Preprint at https://doi.org/10.1016/j.pbi.2006.03.016 (2006).
14. Levsh, O. *et al.* Dynamic Conformational States Dictate Selectivity toward the Native Substrate in a Substrate-Permissive Acyltransferase. *Biochemistry* **55**, 6314–6326 (2016).
15. Ma, X., Koepke, J., Panjikar, S., Fritzsch, G. & Stöckigt, J. Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J. Biol. Chem.*

280, 13576–13583 (2005).

16. Yamane, M., Takenoya, M., Yajima, S. & Sue, M. Molecular and structural characterization of agmatine coumaroyltransferase in Triticeae, the key regulator of hydroxycinnamic acid amide accumulation. *Phytochemistry* **189**, 112825 (2021).
17. Wang, L., Chen, K., Zhang, M., Ye, M. & Qiao, X. Catalytic function, mechanism, and application of plant acyltransferases. *Crit. Rev. Biotechnol.* 1–20 (2021).
18. Wu, D. *et al.* Structure of human carnitine acetyltransferase. Molecular basis for fatty acyl transfer. *J. Biol. Chem.* **278**, 13159–13165 (2003).
19. Greene, N. P., Crow, A., Hughes, C. & Koronakis, V. Structure of a bacterial toxin-activating acyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3058–66 (2015).
20. Bayoumi, S. A. L., Rowan, M. G., Blagbrough, I. S. & Beeching, J. R. Biosynthesis of scopoletin and scopolin in cassava roots during post-harvest physiological deterioration: The E-Z-isomerisation stage. *Phytochemistry* vol. 69 2928–2936 Preprint at https://doi.org/10.1016/j.phytochem.2008.09.023 (2008).
21. Bader, R. F. W. Atoms in Molecules. *Encyclopedia of Computational Chemistry* Preprint at https://doi.org/10.1002/0470845015.caa012 (2002).
22. Mayer, I. Bond order and valence indices: a personal account. *J. Comput. Chem.* **28**, 204–221 (2007).
23. Stringlis, I. A., de Jonge, R. & Pieterse, C. M. J. The Age of Coumarins in Plant-Microbe Interactions. *Plant Cell Physiol.* **60**, 1405–1419 (2019).
24. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*

596, 583–589 (2021).

25. Shaw, P. E. Peptidyl-prolyl isomerases: a new twist to transcription. *EMBO Rep.* **3**, 521–526 (2002).
26. Schmidpeter, P. A. M. & Schmid, F. X. Prolyl Isomerization and Its Catalysis in Protein Folding and Protein Function. *Journal of Molecular Biology* vol. 427 1609–1631 Preprint at https://doi.org/10.1016/j.jmb.2015.01.023 (2015).
27. Heipieper, H. J., Meinhardt, F. & Segura, A. The cis-trans isomerase of unsaturated fatty acids in Pseudomonas and Vibrio: biochemistry, molecular biology and physiological function of a unique stress adaptive mechanism. *FEMS Microbiol. Lett.* **229**, 1–7 (2003).
28. Hluska, T., Šebela, M., Lenobel, R., Frébort, I. & Galuszka, P. Purification of Maize Nucleotide Pyrophosphatase/Phosphodiesterase Casts Doubt on the Existence of Zeatin Cis–Trans Isomerase in Plants. *Frontiers in Plant Science* vol. 8 Preprint at https://doi.org/10.3389/fpls.2017.01473 (2017).
29. Bassil, N. V., Mok, D. & Mok, M. C. Partial Purification of a cis-trans-Isomerase of Zeatin from Immature Seed of Phaseolus vulgaris L. *Plant Physiol.* **102**, 867–872 (1993).
30. Torrens-Spence, M. P. *et al.* PBS3 and EPS1 Complete Salicylic Acid Biosynthesis from Isochorismate in Arabidopsis. *Mol. Plant* **12**, 1577–1586 (2019).
31. Torrens-Spence, M. P. *et al.* Mechanistic basis for the emergence of EPS1 as a catalyst in plant salicylic acid biosynthesis. Preprint at https://doi.org/10.1101/2021.08.21.457228.

32. Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat. Struct. Biol.* **7**, 786–791 (2000).
33. Waki, T. *et al.* A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat. Commun.* **11**, 870 (2020).
34. Kaltenbach, M. *et al.* Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14**, 548–555 (2018).
35. Ueda, M., Tashiro, C. & Yamamura, S. cis-p-Coumaroylagmatine, the genuine leaf-opening substance of a nyctinastic plant, Albizzia julibrissin Durazz. *Tetrahedron Letters* vol. 38 3253–3256 Preprint at https://doi.org/10.1016/s0040-4039(97)00578-9 (1997).
36. Vanholme, R., De Meester, B., Ralph, J. & Boerjan, W. Lignin biosynthesis and its integration into metabolism. *Current Opinion in Biotechnology* vol. 56 230–239 Preprint at https://doi.org/10.1016/j.copbio.2019.02.018 (2019).
37. Steenackers, W. *et al.* cis-Cinnamic acid is a natural plant growth-promoting compound. *J. Exp. Bot.* **70**, 6293–6304 (2019).
38. El Houari, I. *et al.* Seedling developmental defects upon blocking CINNAMATE-4-HYDROXYLASE are caused by perturbations in auxin transport. *New Phytol.* **230**, 2275–2291 (2021).
39. Vanholme, B., El Houari, I. & Boerjan, W. Bioactivity: phenylpropanoids' best kept secret. *Current Opinion in Biotechnology* vol. 56 156–162 Preprint at https://doi.org/10.1016/j.copbio.2018.11.012 (2019).

40. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 449–457 (2005).
41. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1089–1097 (2009).
42. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
44. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
45. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, W537–41 (2012).
46. Gordon, J. C. *et al.* H : a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research* vol. 33 W368–W371 Preprint at https://doi.org/10.1093/nar/gki464 (2005).
47. Myers, J., Grothaus, G., Narayanan, S. & Onufriev, A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins*

63, 928–938 (2006).

48. Webb, B. & Sali, A. Protein Structure Modeling with MODELLER. *Methods in Molecular Biology* 1–15 Preprint at https://doi.org/10.1007/978-1-4939-0366-5_1 (2014).
49. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* vol. 11 3696–3713 Preprint at https://doi.org/10.1021/acs.jctc.5b00255 (2015).
50. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* vol. 97 10269–10280 Preprint at https://doi.org/10.1021/j100142a004 (1993).
51. Hariharan, P. C. & Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* vol. 28 213–222 Preprint at https://doi.org/10.1007/bf00533485 (1973).
52. Vanquelef, E. *et al.* R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research* vol. 39 W511–W517 Preprint at https://doi.org/10.1093/nar/gkr288 (2011).
53. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* vol. 79 926–935 Preprint at https://doi.org/10.1063/1.445869 (1983).

54. Götz, A. W. *et al.* Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).

55. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).

56. Himo, F. & de Visser, S. P. Status report on the quantum chemical cluster approach for modeling enzyme reactions. *Communications Chemistry* vol. 5 Preprint at https://doi.org/10.1038/s42004-022-00642-2 (2022).

57. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

58. Seritan, S. *et al.* TeraChem : A graphical processing unit -accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Computational Molecular Science* vol. 11 Preprint at https://doi.org/10.1002/wcms.1494 (2021).

59. Kulik, H. J., Luehr, N., Ufimtsev, I. S. & Martinez, T. J. Ab Initio Quantum Chemistry for Protein Structures. *The Journal of Physical Chemistry B* vol. 116 12501–12509 Preprint at https://doi.org/10.1021/jp307741u (2012).

60. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).

61. Lange, A. W. & Herbert, J. M. A smooth, nonsingular, and faithful discretization

scheme for polarizable continuum models: the switching/Gaussian approach. *J. Chem. Phys.* **133**, 244111 (2010).

62. Bader, R. F. W. ChemInform Abstract: A Quantum Theory of Molecular Structure and Its Applications. *ChemInform* vol. 23 no–no Preprint at https://doi.org/10.1002/chin.199203324 (2010).
63. Lu, T. & Chen, F. Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* **33**, 580–592 (2012).
64. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
65. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
66. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–4 (2014).

Supplementary information for

Emergence of a proton exchange-based isomerization and lactonization mechanism in the plant coumarin synthase COSY

Kim et al.

*Corresponding author email: wengj@wi.mit.edu

Supplementary Table 1 | Data collection and structure refinement statistics. Data were collected from one crystal per condition and the statistics for the highest resolution shell are shown in parentheses.

Structures	***At*COSY (Apo) PDB ID: 8DQO**	***At*COSY (Scopoletin) PDB ID: 8DQP**	***At*COSY (Umbelliferone) PDB ID: 8DQQ**	***At*COSY (CoA) PDB ID: 8DQR**
		Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell a, b, c (Å)	57.411, 57.903, 270.395	58.459, 58.349, 273.281	57.907, 58.771, 273.611	58.514, 88.886, 96.901
Ɑ, B, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
R_{merge}	1.361e-17 (1.178e-17)	0.04009 (0.7324)	0.0714 (0.7117)	0.03387 (0.7622)
R_{pim}	1.361e-17 (1.178e-17)	0.04009 (0.7324)	0.0714 (0.7117)	0.03387 (0.7622)
$CC_{1/2}$	1 (1)	0.999 (0.474)	0.991 (0.441)	0.999 (0.425)
I/σ((I)	13.42 (3.53)	10.64 (0.89)	9.40 (1.00)	12.69 (0.89)
Completeness (%)	98.65 (89.64)	97.46 (96.88)	97.43 (92.20)	97.02 (90.69)
		Refinement		
Resolution range (Å)	48.72 - 1.9 (1.968 - 1.9)	57.17 - 2.48 (2.569 - 2.48)	136.8 - 2.33 (2.413 - 2.33)	50.09 - 2.26 (2.341 - 2.26)
Reflections used in refinement	71349 (6349)	33351 (3261)	40000 (3722)	23599 (2172)
R_{work} / R_{free} (%)	19.6 / 23.0 (23.8 / 25.9)	17.3 / 28.1 (29.2 / 37.7)	25.0 / 30.9 (30.9 / 35.3)	18.8 / 28.4 (38.4 / 39.3)
		Number of atoms		
Number of non-hydrogen atoms	7232	6718	6862	3586
macromolecules	6835	6555	6734	3388
ligands	46	41	18	55
solvent	351	122	110	143
Protein residues	873	835	862	433
Ramachandran allowed (%)	1.97	4.27	5.19	2.36
Ramachandran outliers (%)	0.12	1.22	0.59	0.71
		R.M.S. deviations		
Bond lengths (Å)	0.012	0.024	0.003	0.013
Bond angles (°)	1.48	2.35	0.69	1.68
		$B_{factors}$ (Å^2)		
Average B-factor	34.27	79.56	58.89	54.53
macromolecules	34.16	79.71	58.86	54.36
ligands	41.20	73.51	56.61	57.41
solvent	35.64	73.19	60.93	57.53

Supplementary Table 2 | Statistical analysis of *At*COSY mutant assays. Unpaired, two-tailed t-test was performed using Prism[1] for each mutant vs. WT.

Mutants	p-value	t	df	Significantly different (P < 0.05)?
H161A	0.4433	0.8498	4	no
H161Q	0.0798	2.335	4	no
W371H	0.8868	0.1516	4	no
W371A	0.0058	5.363	4	yes
W371V	0.0110	4.473	4	yes
W371M	0.0066	5.188	4	yes
Y42F	0.0466	2.846	4	yes
F40T	0.3080	1.167	4	no
C164A	0.5104	0.7217	4	no
F40T/Y42S	0.9275	0.09688	4	no
Y42F/H161A	0.0028	6.556	4	yes
G166A	0.1281	1.914	4	no
G166L	0.0107	4.517	4	yes
Y373A	0.9228	0.1052	3	no
Y373F	0.6200	0.5367	4	no
L374A	0.0052	5.540	4	yes
Loop-swap	0.0083	4.854	4	yes

Supplementary Table 3 | Tabulated energetics for the intermediates of the proposed mechanism. Energies are reported in kcal/mol. Reactants are labeled as R and products are labeled as P. Separate geometry optimizations were performed with each functional.

System	B3LYP/6-31G* ΔE (kcal/mol)	wPBEh/6-31G* ΔE (kcal/mol)
R	0.00	0.00
IM_1	6.56	4.88
IM_2	24.00	23.97
IM_3	27.55	30.50
IM_4	7.54	8.94
IM_5	3.91	1.74
P	-22.82	-20.08

Supplementary Table 4 | Strains and plasmids used in this study.

Strain	Description	Source
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS*	New England BioLabs

Plasmid	Description	Source
pHis8-4b-His_8-*At*COSY	His_8-*At*COSY (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY H161A	His_8-*At*COSY H161A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY H161Q	His_8-*At*COSY H161Q (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY W371H	His_8-*At*COSY W371H (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY W371A	His_8-*At*COSY W371A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY W371V	His_8-*At*COSY W371V (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY W371M	His_8-*At*COSY W371M (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY Y42F	His_8-*At*COSY Y42F (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY F40T	His_8-*At*COSY F40T (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY C164A	His_8-*At*COSY C164A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY F40T/Y42S	His_8-*At*COSY F40T/Y42S (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY Y42F/H161A	His_8-*At*COSY Y42F/H161A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY C308S	His_8-*At*COSY C308S (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY C308A	His_8-*At*COSY C308A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY G166A	His_8-*At*COSY G166A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY G166L	His_8-*At*COSY G166L (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY Y373A	His_8-*At*COSY Y373A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY Y373F	His_8-*At*COSY Y373F (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY L374A	His_8-*At*COSY L374A (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*COSY-HCT-loop-swap	His_8-*At*COSY-HCT-loop-swap (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*At*4CL1	His_8-*At*4CL1 (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*Gm*COSY	His_8-*Gm*COSY (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*St*COSY	His_8-*St*COSY (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study
pHis8-4b-His_8-*Sb*COSY	His_8-*Sb*COSY (T7), *lacI*, *lacO*, Kan^R, F1-ori	This study

Supplementary Table 5 | Oligonucleotide sequences reported in this study.

Name	Sequence (5′ - 3′)
His_8-*At*COSY-F	GAAAACTTGTACTTCCAGGCCCATGGCATGGCGACACTTGAAATTACCGATATAGCC
His_8-*At*COSY-R	CGGGCTTTGTTAGCAGCCGGATCGCCATGGTCAAGACAGGGCAAATTCACCTTTGTGG
His_8-*At*COSY H161A-F (QuikChange)	TCGCAGATCGCAGCATGAATCGATGCACCTAGAACCC
His_8-*At*COSY H161A-R (QuikChange)	GGGTTCTAGGTGCATCGATTCATGCTGCGATCTGCGA
His_8-*At*COSY H161Q-F (QuikChange)	ATCGCAGATCGCCTGATGAATCGATGCACCTAGAAC
His_8-*At*COSY H161Q-R (QuikChange)	GTTCTAGGTGCATCGATTCATCAGGCGATCTGCGAT
His_8-*At*COSY W371H-F (QuikChange)	GGAATGCCCCAAGTATCGATGGTCCGTGAATCCGGTCAC
His_8-*At*COSY W371H-R (QuikChange)	GTGACCGGATTCACGGACCATCGATACTTGGGGCATTCC
His_8-*At*COSY W371A-F (QuikChange)	GCCCCAAGTATCGCGCGTCCGTGAATCCGG
His_8-*At*COSY W371A-R (QuikChange)	CCGGATTCACGGACGCGCGATACTTGGGGC
His_8-*At*COSY W371V-F (QuikChange)	ATGCCCCAAGTATCGCACGTCCGTGAATCCGGTC
His_8-*At*COSY W371V-R (QuikChange)	GACCGGATTCACGGACGTGCGATACTTGGGGCAT
His_8-*At*COSY W371M-F (QuikChange)	ATGCCCCAAGTATCGCATGTCCGTGAATCCGGTC
His_8-*At*COSY W371M-R (QuikChange)	GACCGGATTCACGGACATGCGATACTTGGGGCAT
His_8-*At*COSY Y42F-F (QuikChange)	GACGCGGAGGAAACGGAAGCTGACGTG
His_8-*At*COSY Y42F-R (QuikChange)	CACGTCAGCTTCCGTTTCCTCCGCGTC
His_8-*At*COSY F40T-F (QuikChange)	CGCGGAGGTAACGGGTGCTGACGTGGAGGT
His_8-*At*COSY F40T-R (QuikChange)	ACCTCCACGTCAGCACCCGTTACCTCCGCG
His_8-*At*COSY C164A-F (QuikChange)	TCGCACCTAACCCATCGGCGATCGCATGATGAATCG
His_8-*At*COSY C164A-R (QuikChange)	CGATTCATCATGCGATCGCCGATGGGTTAGGTGCGA
His_8-*At*COSY F40T/Y42S-F (QuikChange)	GGAGTAGACGCGGAGGCTACGGGTGCTGACGTGGAGGTTG
His_8-*At*COSY F40T/Y42S-R (QuikChange)	CAACCTCCACGTCAGCACCCGTAGCCTCCGCGTCTACTCC
His_8-*At*COSY Y42F/H161A-F (QuikChange)	AGCTTCCGTTTTCTCCGCGT
His_8-*At*COSY Y42F/H161A-R (QuikChange)	ACGCGGAGAAAACGGAAGCT
His_8-*At*COSY C308S-F (QuikChange)	GCTACTGGGGAAATGGATCAGTGCCAATGTATGCTCAGAT
His_8-*At*COSY C308S-R (QuikChange)	ATCTGAGCATACATTGGCACTGATCCATTTCCCCAGTAGC
His_8-*At*COSY C308A-F (QuikChange)	GCTACTGGGGAAATGGAGCAGTGCCAATGTATGCTCAGAT
His_8-*At*COSY C308A-R (QuikChange)	ATCTGAGCATACATTGGCACTGCTCCATTTCCCCAGTAGC
His_8-*At*COSY G166A-F (QuikChange)	GACTCGCACCTAACGCATCGCAGATCGCG
His_8-*At*COSY G166A-R (QuikChange)	CGCGATCTGCGATGCGTTAGGTGCGAGTC
His_8-*At*COSY G166L-F (QuikChange)	GACTCGCACCTAATAGATCGCAGATCGCG
His_8-*At*COSY G166L-R (QuikChange)	CGCGATCTGCGATCTATTAGGTGCGAGTC
His_8-*At*COSY Y373A-F (QuikChange)	CGTGGAATGCCCCAAGGCTCGCCAGTCCGTGAAT
His_8-*At*COSY Y373A-R (QuikChange)	ATTCACGGACTGGCGAGCCTTGGGGCATTCCACG
His_8-*At*COSY Y373F-F (QuikChange)	TGGAATGCCCCAAGAATCGCCAGTCCGTG
His_8-*At*COSY Y373F-R (QuikChange)	CACGGACTGGCGATTCTTGGGGCATTCCA
His_8-*At*COSY L374A-F (QuikChange)	CGTGGAATGCCCCGCGTATCGCCAGTCCGTG
His_8-*At*COSY L374A-R (QuikChange)	CACGGACTGGCGATACGCGGGGCATTCCACG
His_8-*At*4CL1-F	ATGGCGCCACAAGAACAAGCAGT
His_8-*At*4CL1-R	TCACAATCCATTTGCTAGTTTTGCCC
His_8-*Gm*COSY-F	CCAGGCCCATGGCATGCAAAGGATCAAAACCTCAGAAC
His_8-*Gm*COSY-R	AGCCGGATCGCCATGGCTATATGTGGCTCAGCCACTGC
His_8-*St*COSY-F	CCAGGCCCATGGCATGGAAGTAACTATCGGCGAAAC
His_8-*St*COSY-R	CGGATCGCCATGGTCAAGCAAGTCCATGCTCCAACTTTTCC

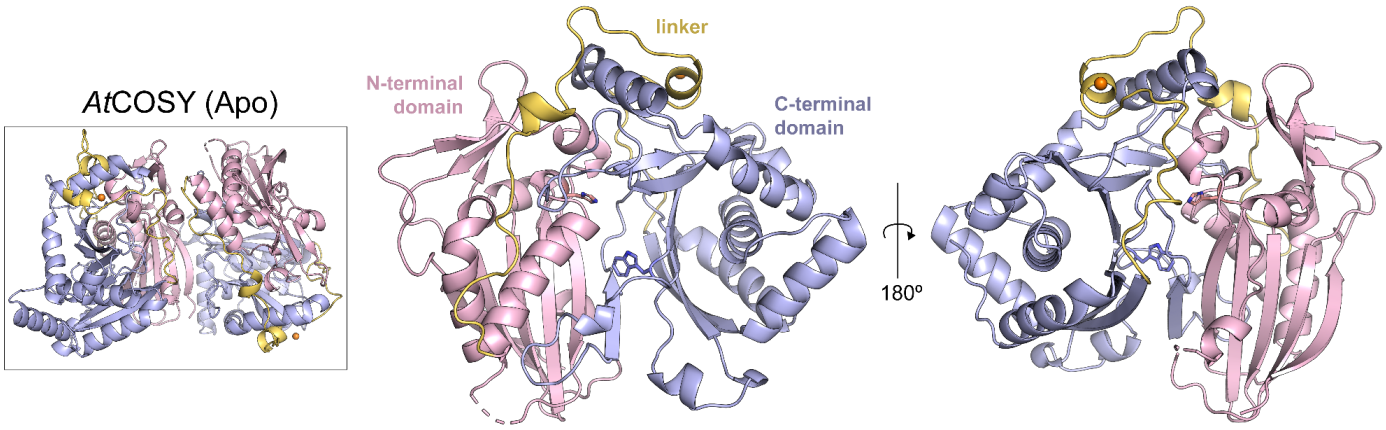
Name	**Sequence (5′ - 3′)**
His_8-*Sb*COSY-F	CCAGGCCCATGGCATGGAGCCACAGCAGCTGCGACTACG
His_8-*Sb*COSY-R	GCAGCCGGATCGCCATGGTCAAAGTTTTCCCGCTGAACACTGC
His_8-*At*COSY-HCT-loop-swap-F	CCAGGCCCATGGCATGGCGACACTTGAAATTACCG
His_8-*At*COSY-HCT-loop-swap-R	GCAGCCGGATCGCCATGGTCAAGACAGGGCAAATTCACC

Supplementary Table 6 | Synthetic gene sequences reported in this study.

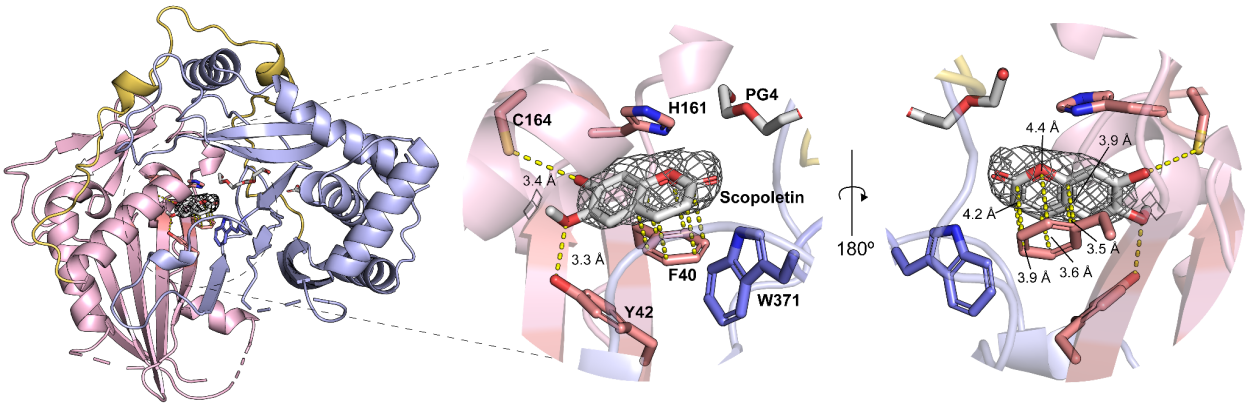
Name	Sequence (5′ - 3′)
*Gm*COSY	ATGCAAAGGATCAAAACCTCAGAACGCACTTTGATCTTCCCTTCCCACCCTCCTTTTCTCCAAGACCACCCC TTCCCTCTCTCCCACCTCGACACTGATCCCAACCTCCACCTTACCTTCCGCTACCTCCGCGCGTACACCTC AACAACAACAACAACCTCCCTCGACCCCTTCCACGTCATCTCCTCCTCCCTCTCCCACGCCCTCCCCCACT TCTACCCCCTCACCGCCACCCTCCGCCGCCAACAAACCTCCCCCCACCGCCTCCAACTCTGGTGCGTCG CCGGCCAGGGCATCCCCCTCATCCGCGCCACCGCGGACTTCACCCTCGAGTCCGTGAACTTCCTCGACA ACCCGGCCTCGAGCTTCTTGGAGCAGTTAGTGCCCGACCCGGGACCCGAGGAGGGGATGGAGCACCCG TGCATGCTCCAGGTGACGGTGTTCGCGTGCGGGGGATTCACCCTCGGCGCGGCGATGCACCACGCGCTC TGCGACGGCATGGGCGGGACGCTGTTCTTCAATGCGGTGGCGGAGCTGGCGCGTGGGGCGACCCGGAT AACGTTGGACCCGGTTTGGGACCGTGCGAGGTTGCTGGGTCCCAGGGACCCGCCCCTGGTGGATTCGC CGTTGATTGGGGAGTTTCTGCGTTTGGAGAAGGGAGTTTTGCCGTACCAACAGAGTGTTGGTGGGGTCGC GAGAGAGTGCTTTCACGTGAAGGATGAGTGCTTGGACAATTTCAAGAGGACCTTGTTGGAGCAATCTGGGT TGAACTTCACCGTTTTTGAGGCTCTTGGTGCCTACATCTGGAGGGCTAAGGTCAGGGCCTCGGGAATCCAG GCTGATGAAAAGGTGAAGTTTGCATACTCAATTAACATACGGAGACTGGTAAAGCCACCACTGCCTGGTGGC TATTGGGGTAATGGTTGTGTGCCAATGTACGTACAACTTAGTGCCAAAGATTTGATAGAGAAACCCGTTTGCG AAACCGCAGAGCTAATAAAAAAGAGTAAAAGCAATGTCACTGATGAGTATGTTAAGTCCTACATCGATTATCAG GAGCTGCATTTTGCTGATGGAATCACTGCGGGAAAAGAGGTTAGTGGGTTCACGGATTGGAGGCACTTGGG CCATTCAACTGTGGACTTTGGGTGGGGTGGCCCAGTTACTGTTTTGCCCCTTGGAAGGAACTTACTTGGGA GTGTTGAGCCTTGCTTTTTTTTGCCTTATTCAACAGCCACTTCAGAGAAGAAAGAGGGGTTCAAGGTTTTGG TGACTTTGAGAGAGGCTGCGTTGCCTGCTTTCAGAGAAGACATGAAAGTGTTTTGCAATAGCCAAGAGCAG TGGCTGAGCCACATATAG
*St*COSY	ATGGAAGTAACTATCGGCGAAACTAACACAATCTACCCTTCTAAGCTACCTTTCACTGAAAACCATGTCCTTC CTCTCTCCCATATTGACACCGATCGGAACCTCAATTTCACTTTCCGATACCTCCGTGTCTACGTCAACGACG ACACGACACAACAAAAACCAGACCCATATGAAGCTCTCACCTCCTCCCTCTCCGCCGCTCTCGTCCACTAC TACCAATTCGCCGGCTCTCTCCGCCGTCGTCCGTCGGACAACCGTCTAGAGCTTCACTGTCAAGTAGGGAA TGGCGTTCCCGTAATTCTGTCCACGGTGGATTGTACCTTGGCTTCTATTAACTACCTGGATGATCCAGATTATA ACTTCGCTGAGAAGCTGGTACCCGACCCGAGAGATGAGGAGGTGTTGACCCGACCCTTGATTTTGCAACT GAACCGTTTTAAGTGTGGTGGGTGGGTTTTCGGAACGGCGGTTCATCACGCGATGTGTGATGGAATGGGAT CGACGCTGTTTTTTCATGCTATGGCGGAGATTGCTCGTGGGGAGAACGGGATGAAGATTGAGCCGGTGTG GGACCGGTCGAATTTGCTTGGACCGAGGAATCCACCGCGAGTTGAATTTCCGGTTCATGAGTTTCTCAGTT TGGATAGGGATTCGTCTCCATACTTGGAATCAGATAAACCGGCGGTTCGAGAGTGCTTTGAGGTGAAGGATG AGTGGTTGGATAGGCTAAAGGGGTTTCTTCATGAGCAATCCGGTTCAAATTATACAACTTTTGAAGCTTTGGG AGCTTTCATATGGCGAGCAAAGGCAAAGGCTTCTAAAATCTCAGATGATGAGACAGTGAAGTATGCCTATCTA ACCAATATCAGAAGAAGAGTGAAGCCACCATTGCCAGCAGGTTACTGGGGCAATGGGTGCGTGCCAATATA CGTTCAGCTCCTCGCAAAGGACCTCATCAATGAACCTATCTGGAAAGCAGCAGATGCGATAAAGAAGAGCA AGGACATCATCACAGATGAGTATGTTCATTCGTTCATTGATTTCCAGGAGCTGCATTATGATGAAGGGATCACA TCAGGGAATAGAGTGAGTGCATTTACAGATTGGCGGCATGTAGGCCACGAGACGGTTGATTTTGGATGGGG TGGCCCTGTGACTGTCTTTCCTCTGTCTAGACACTTGGTTGGGAGTGTTGAACCTTGTTTTTTCTTGCCTTATT CTTCCGCCACTCAAGGTAAGAAAGATGGTTTCAAAGTTTTGGTCTGTCTGCAACAAGAAGCCATGCCTGTTT TCATGGAAGAGATGGAAAAGTTGGAGCATGGACTTGCTTGA
*Sb*COSY	ATGGAGCCACAGCAGCTGCGACTACGAATTCTGGACACCGTGCGCCTGTCGCCGCCACCCGCCCCGGC GCAGCACGCGGCGGCCCTCCCGCTCTCCGGCCTCGACGCCGACCGCAACGCGCTCCACGTCACCTTCC GCACGCTCCGCTTCTTCCCACCTCCCCCGCCGTCCATCGACCCCCACGCGGTCCTCCCGCCCGCGTTC GAGGCCGCGCTCGGTCTCTTCCCGGCGCTCGCGGGCCGCCTCCGCGTCCGCGACGGCCACGTTGTGG TCGGCGCTGCCGCCGTGCCCGTCGTCCTCGCGGAGTCCGGTCTGTCCGCGGCCGACGTCGACACCGAC TGCCCCGGCTCGGCGCTGCTCGACCACCTCGCGCCGCCGGGGGACGGCGACGCGGATACCCCGGTGC TCGCGCTCCAGGCCACGCGGTTCGCGTGCGGCGGCGTCGCGCTCGGGATGCGGGTCGCGCACGCCCT CTGCGACGGCGCGGGCGCCACCAAGTTCCTCGCCGCCGCGGCGCGGTTCGCGCGGGGGCAGGGGAT GCCGGAGGTGGCGCCTGTGTGGGACCGGCGGGAGCTGCTGGGCCCGAGGCGCCCGCCGCGCGTCGC GACGCTGGTGTTCGACCGCGTCCTCGCGCTCGACGGCGGCGTCGCCCGGTGTGGGCCGTACGGAGCC GTGCGGGAATGGCGCGAACAGCAGCGGCAGCTGACGAGGGAATGCTTCCACGTGAGCAACGCGCGCGT GGAGGCGCTGAGAGCGCGGATCGCCGACGAGGCTGGCGTCAAGCTCACGACCTTCGAGGTTGTCGCAG CGTTCATCTGGCGCGCCAGGGTCAAGGCAAACGGGACCAGCTCCGGCGAGGTGGTGAAGATGGTGTACT CCATGAACATCAGCAAGCTCGTCGACCCGCCGCTCCCCGACGGGTACTGGGGCAACGTGTGCGTCCCTG TCTACGTCACCCTGGCCGCCTGCGACCTCGTCGCCCAGCCACTGGCGGCAACCGCCGCCCTGATAAAGA AGAGCAAGCAGGCCGTGGACGACGAGTACGTGCGATCCTACGTCGACTTCCAGGAGCTGCACCGCCACG ATGGGGTCACGGCGGGAGCCGTGAGTGCGTTCACGGACTGGCGGCGGCTCGGCCATGGGGAGGTGGA CTTCGGGTGGGGCGGCCCCGACTCCGTGCTGCCGCTCTCGTGGAGGATCCTCGGGAGCACGGAGCCCT GCTTCCTGCTGCCCTACGGCGCGGGGGATGAGCGGCGGCGGGGAGGGTTCAAGGTGTTCATCGCGCTG CAGCGCACGGCGCTGGCTGGTTTCAGAGAGGAAATGCAGGAGCTGCTGTTGCAACCGAAGCAGTGTTCA GCGGGAAAACTTTGA

Name	Sequence (5′ - 3′)
*At*COSY-HCT-loop-swap	ATGGCGACACTTGAAATTACCGATATAGCCCTGGTTCAACCTTCTCACCAACCACTCTCCAACGACCAAACT CTCTCTCTTTCCCATCTCGACAATGATAACAACCTCCACGTCAGCTTCCGTTACCTCCGCGTCTACTCCTCCT CCTCTTCCACCGTCGCCGGAGAAAGCCCCTCTGCCGTCGTATCCGCCTCTCTTGCCACCGCTCTCGTTCA CTACTACCCTCTTGCTGGCTCTCTCCGTCGCTCTGCCTCCGATAACCGATTTGAACTACTCTGCTCCGCTGG TCAAAGCGTGCCTTTAGTCAACGCTACAGTGAACTGTACGCTTGAGTCAGTCGGGTATTTGGATGGACCCGA TCCAGGTTTCGTCGAGAGATTGGTACCGGATCCGACCCGGGAGGAAGGAATGGTCAATCCTTGTATCCTCC AGGTCACTATGTTTCAGTGTGGTGGTTGGGTTCTAGGTGCATCGATTCATCATGCGATCTGCGATGGGTTAGG TGCGAGTCTGTTCTTCAACGCTATGGCGGAATTAGCTCGCGGAGCGACAAAGATTTCGATCGAACCGGTTT GGGACAGAGAACGTTTACTTGGTCCAAGGGAGAAGCCTTGGGTTGGAGCTCCAGTTCGTGACTTCTTGAG CCTGGATAAGGACTTTGATCCATATGGACAAGCCATTGGAGACGTCAAAAGAGACTGCTTCTTTGTGACCGA CGATTCTTTGGATCAATTGAAGGCTCAATTACTCGAGAAATCGGGTCTGAATTTCACCACATTCGAAGCTCTT GGTGCTTACATTTGGCGTGCAAAGGTAAGAGCTGCAAAGACTGAGGAAAAGGAGAATGTGAAATTTGTGTAT TCGATAAATATAAGGAGATTGATGAATCCACCTTTGCCTAAAGGCTACTGGGGAAATGGATGTGTGCCAATGT ATGCTCAGATCAAAGCTGGAGAACTCATTGAGCAACCAATCTGGAAAACTGCAGAGCTCATAAAACAGAGCA AATCCAATACGAGTGATGAATATGTACGCTCCTTTATCGACTTCCAAGAGCTGCATCACAAAGATGGAATCAAT GCCGGTACAGGAGTGACCGGATTCACGGACTGGCGATACTTGGGGCATTCCACGATTGATTTTGGATGGG GAGGACCTGTGACGGTTTTGCCAGGTGGAATTCCATACGAGGGTTTGTCTTTTTTCTTGCCATATTCTACTGA TGCTGCAGCTGGAAGCAAGAAAGACAGTGGGTTTAAGGTTTTGGTAAATCTGCGCGAATCTGCAATGCCTG AGTTTAAAGAGGCCATGGATAAGTTCCACAAAGGTGAATTTGCCCTGTCTTGA

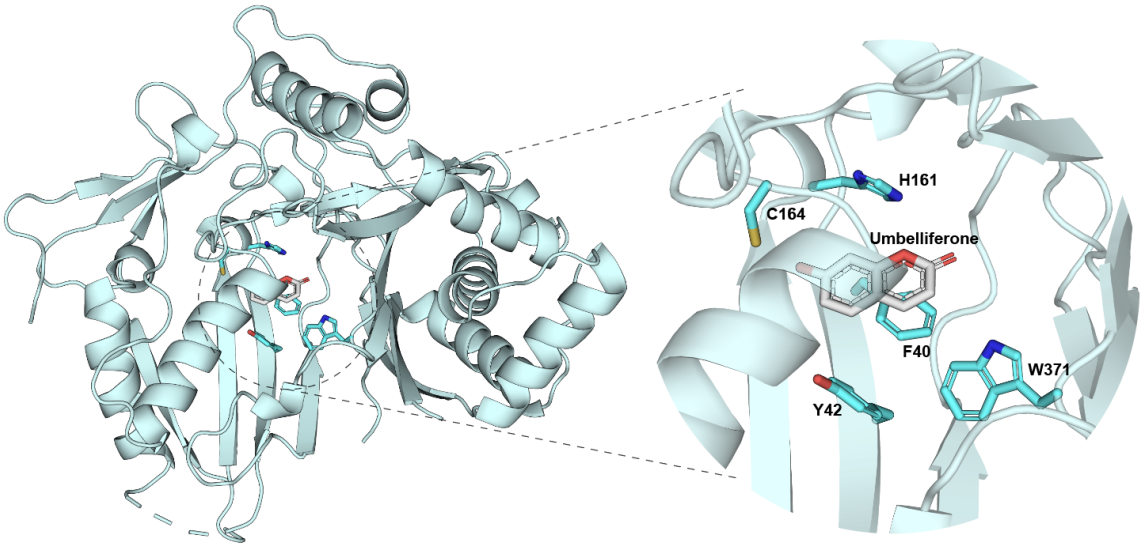
Supplementary Figures



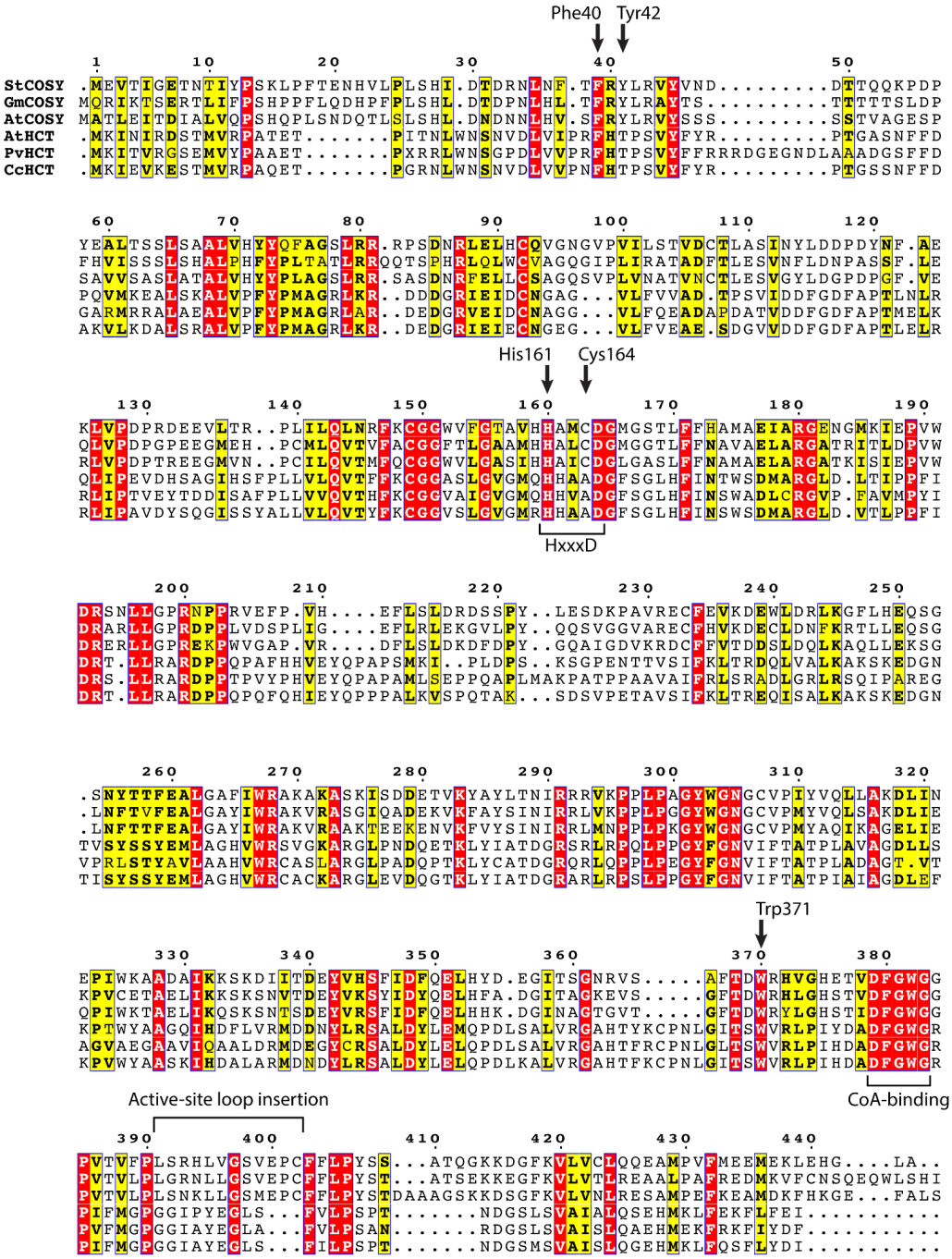
Supplementary Fig. 1 | Apo structure of *At*COSY. Apo structure of *At*COSY at 1.9 Å resolution. The overall fold showcases a pseudo-symmetric N-terminal (residues 1-181; light pink) and C-terminal (residues 230-451; light blue) domains, connected by a linker loop (residues 182-229; yellow). Ca^{2+} ion is displayed as an orange sphere near the linker. The conserved BAHD catalytic residues (His161 and Trp371) are showcased within the active site pocket.



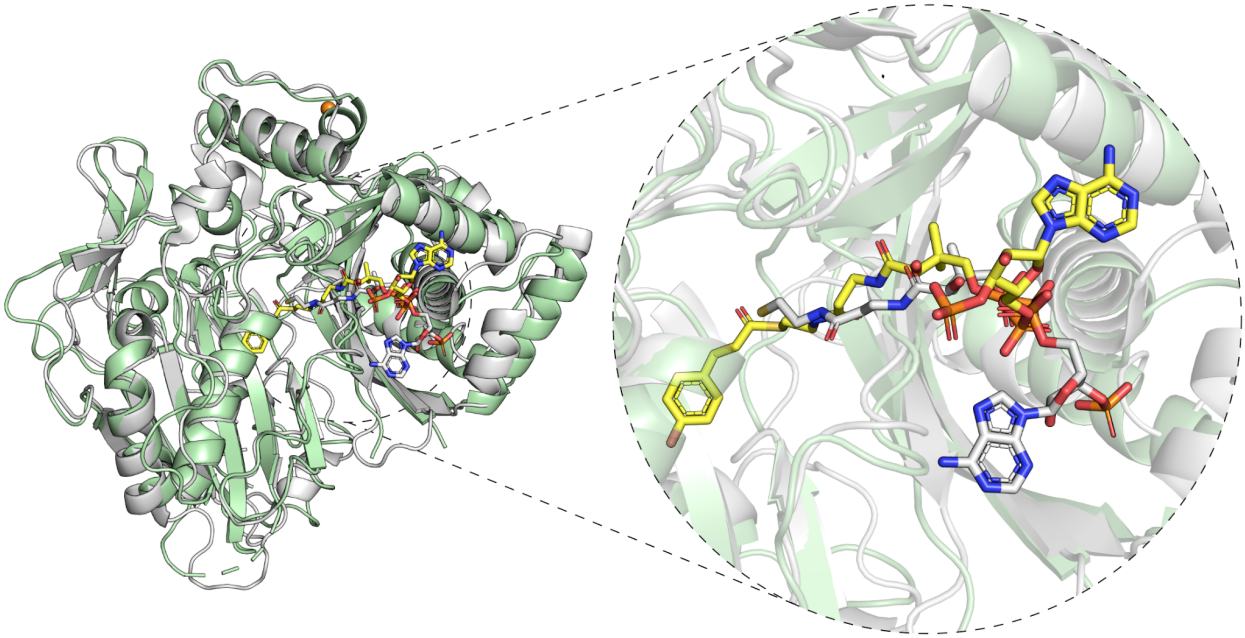
Supplementary Fig. 2 | Active-site residues involved in scopoletin stabilization and catalysis. Scopoletin-bound structure of *At*COSY is displayed in N-terminal (residues 1-181; light pink) and C-terminal (residues 230-451; light blue) domains, connected by a linker loop (residues 182-229; yellow). The conserved BAHD catalytic residues (His161 and Trp371), along with product stabilizing residues (Phe40, Tyr42, and Cys164) are showcased within the active site pocket. The mFo – DFc simulated annealing omit map with electron density map around SCO ligand is contoured at 1.0 σ.



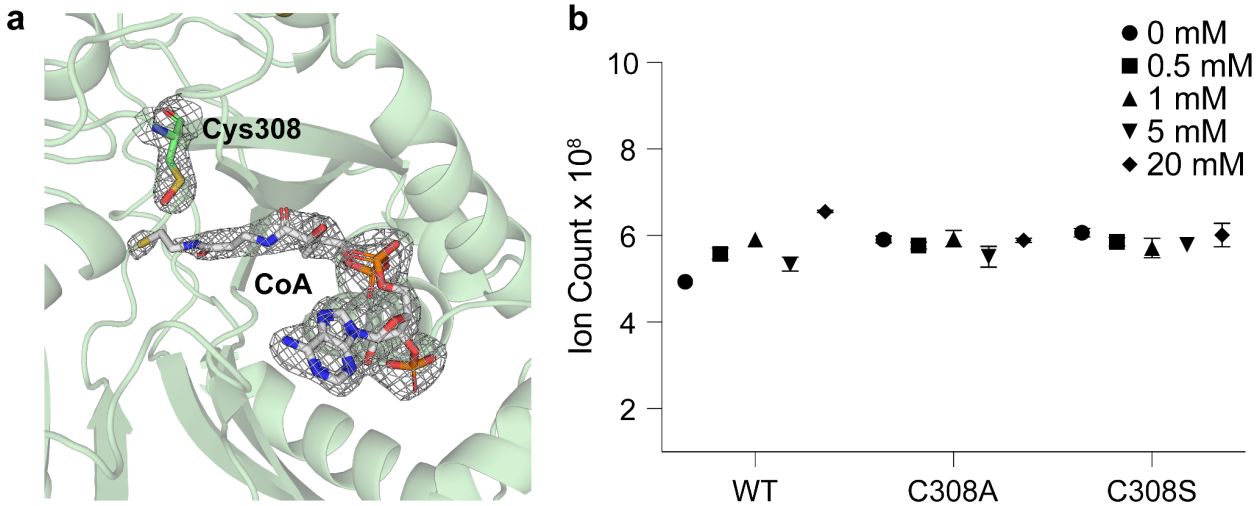
Supplementary Fig. 3 | Structure of *At*COSY co-crystallized with umbelliferone. The conserved BAHD catalytic residues (His161 and Trp371), along with product stabilizing residues (Phe40, Tyr42, and Cys164) are showcased within the active site pocket.



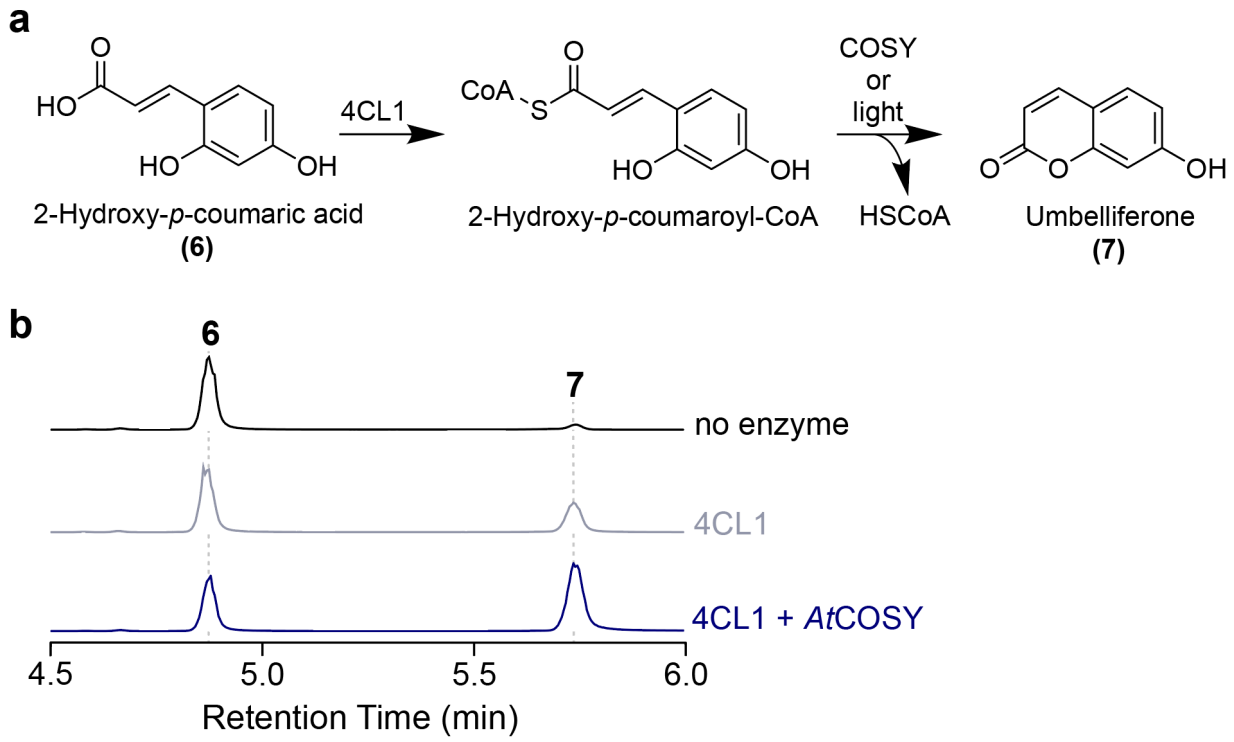
Supplementary Fig. 4 | Sequence alignment of COSY and HCT sequences. COSY homologous sequences were selected from *A. thaliana* (NCBI Accession # NP_174189.1), *S. tuberosum* (NCBI Accession # KAH0704856.1), and *G. max* (NCBI Accession # XP_003541712.1). HCT homologous sequences were selected from *A. thaliana* (NCBI Accession # NP_199704.1), *P. virgatum* (NCBI Accession # AGM90558.1) and *C. canephora* (NCBI Accession # ABO47805.1). Black arrows indicate residues involved in stabilization of scopoletin and canonical residues of BAHD acyltransferases. Conserved sequence motifs in BAHD acyltransferases: HxxxD (catalytic domain) and DFGWG (CoA binding domain) are highlighted, along with the observed active-site loop insertion in COSY. The multiple sequence alignment was built using MUSCLE[2] and the figure was generated using ESPript3.0[3].



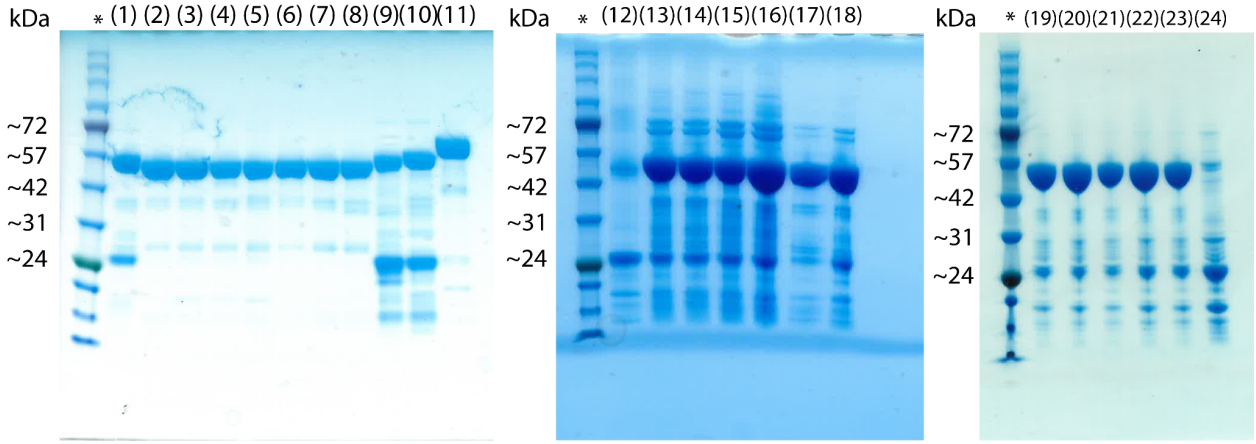
Supplementary Fig. 5 | Structural alignment of *At*COSY-CoA and *At*HCT. Free-CoA bound *At*COSY structure highlighted in green is aligned to *At*HCT highlighted in gray (PDB: 5KJT). CoA from *At*COSY structure is colored in white, whereas p-coumaroyl-CoA from *At*HCT is colored in yellow.



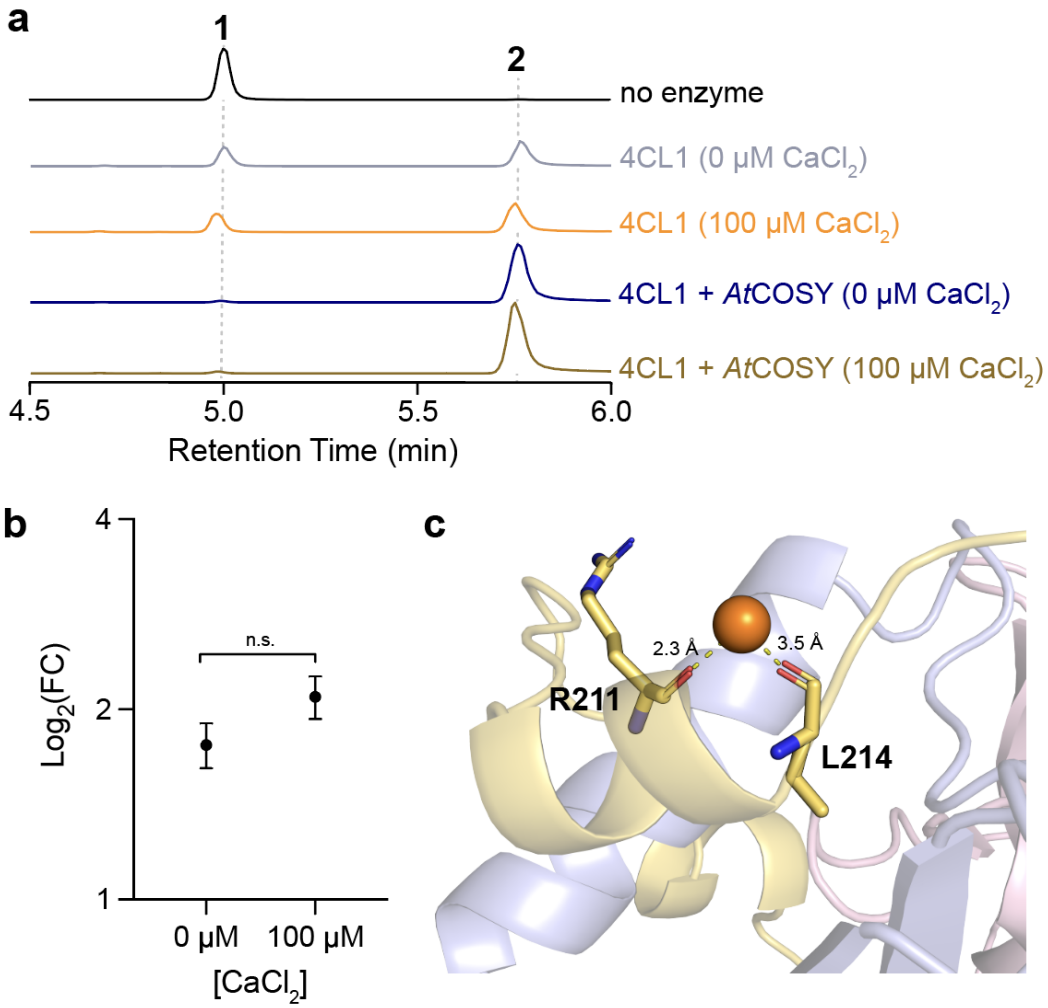
Supplementary Fig. 6 | Observation of Cys380 oxidation in *At*COSY structure and assay for redox-dependency. (**a**) Free-CoA bound *At*COSY structure highlighted in green portrays a single oxidation of Cys308 to sulfenic acid oxidation. |2mFo – DFc| electron density map for Cys308 and CoA (white) is contoured at 1.0 σ. (**b**) Total ion count of scopoletin from LC-MS for WT, C308A, and C308S enzymatic assays in duplicates across varying [H_2O_2]. The error bars represent standard error of mean. Unpaired, two-tailed t-test for statistical comparison of the coumarin synthase activity between C308 mutants to WT indicated no significant difference.



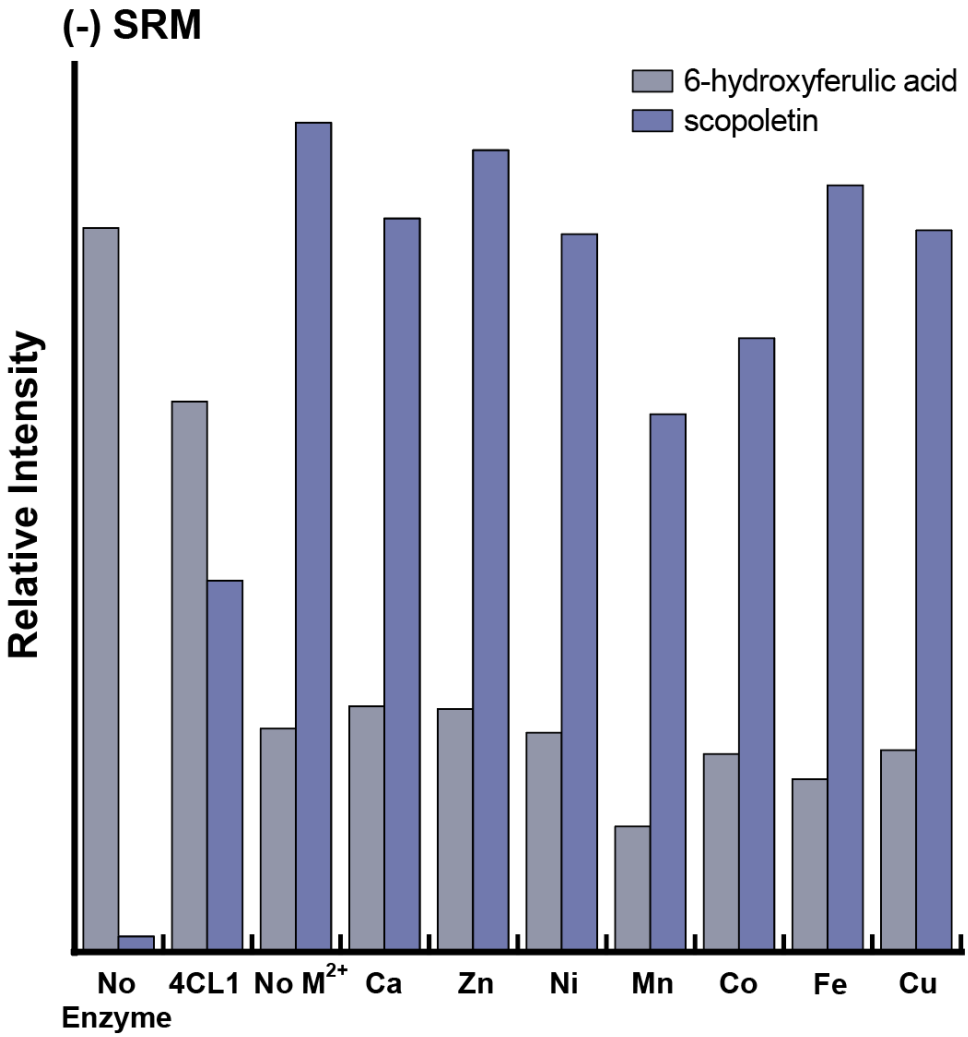
Supplementary Fig. 7 | *In vitro* biochemical assay of *At*COSY for *trans-cis* isomerization and lactonization activity on 2OHpCA. (**a**) Reaction schematic of the conversion of 2-hydroxy-*p*-coumaric acid (2OHpCA) to umbelliferone using 4-coumarate-CoA ligase 1 from *A. thaliana* (*At*4CL1) and coumarin synthase (*At*COSY). (**b**) Combined LC-MS extracted ion chromatograms (XICs) of 135.04512, 179.03494 *m/z*; **6** and 161.02428 *m/z*; **7**. As umbelliferone production is spontaneously catalyzed by light, 4CL1 without *At*COSY sample serves as an experimental control for the coumarin synthase activity of *At*COSY. 4CL1+COSY sample shows complete conversion of 2OHpCA into umbelliferone.



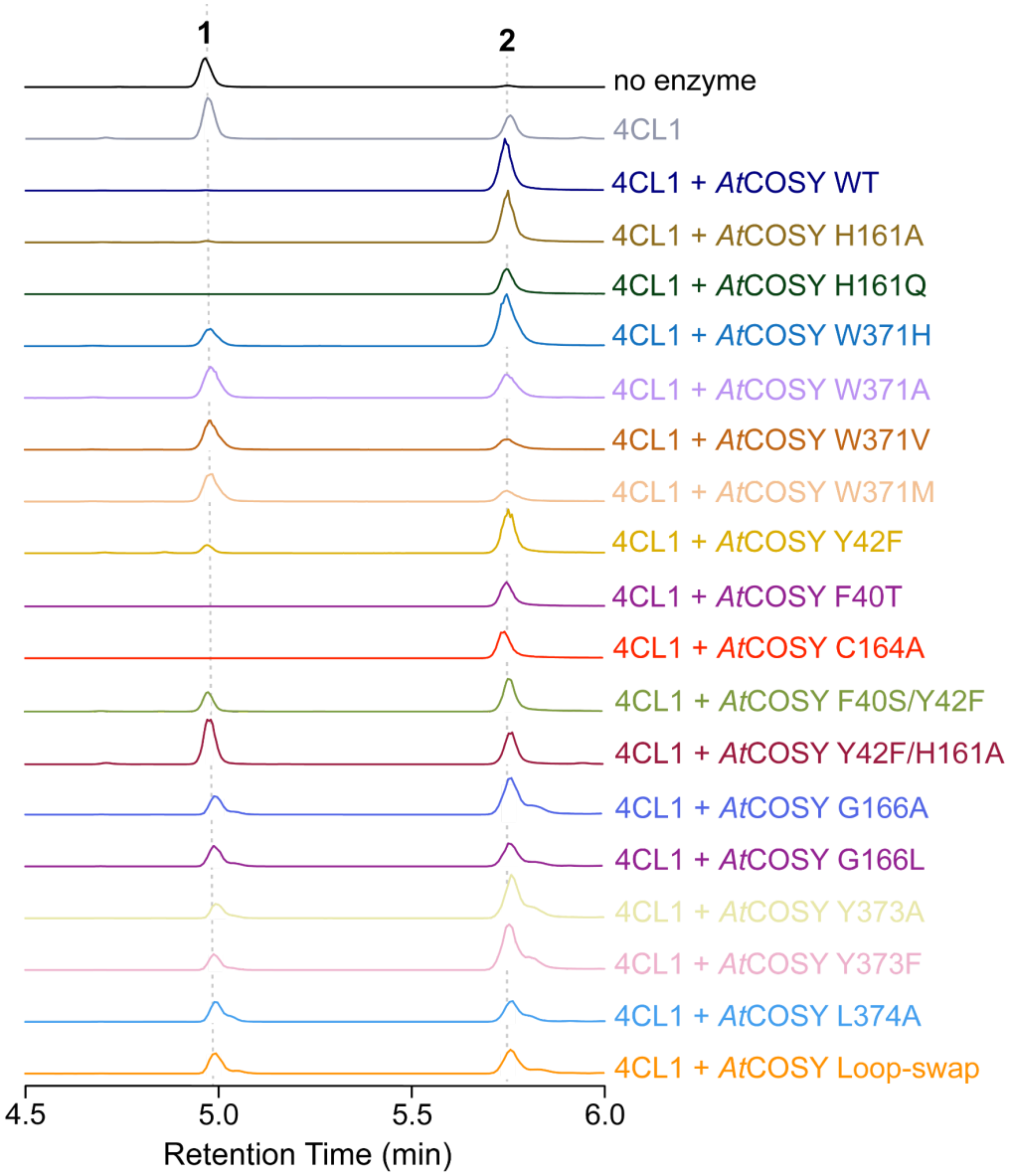
Supplementary Fig. 8 | Purified recombinant proteins assayed in this study. The lanes in the gel correspond to the following: (1) *At*COSY WT, (2) *At*COSY H161A, (3) *At*COSY Y42F, (4) *At*COSY Y42F/H161A, (5) *At*COSY H161Q, (6) *At*COSY F40T, (7) *At*COSY C164A, (8) *At*COSY F40T/Y42S, (9) *Gm*COSY WT, (10) *St*COSY WT, (11) *At*4CL1 WT, (12) *Sb*COSY WT, (13) *At*COSY W371H, (14) *At*COSY W371A, (15) *At*COSY W371V, (16) *At*COSY W371M, (17) *At*COSY C308S, (18) *At*COSY C308A, (19) *At*COSY L374A, (20) *At*COSY Y373A, (21) *At*COSY Y373F, (22) *At*COSY G166A, (23) *At*COSY G166L, and (24) *At*COSY$^{\text{HCT-Loop-swap}}$. (1) and (19)-(24) have N-terminal 8xHis-tags. The asterisk indicates the lane for BlueStain™ protein ladder (GoldBio).



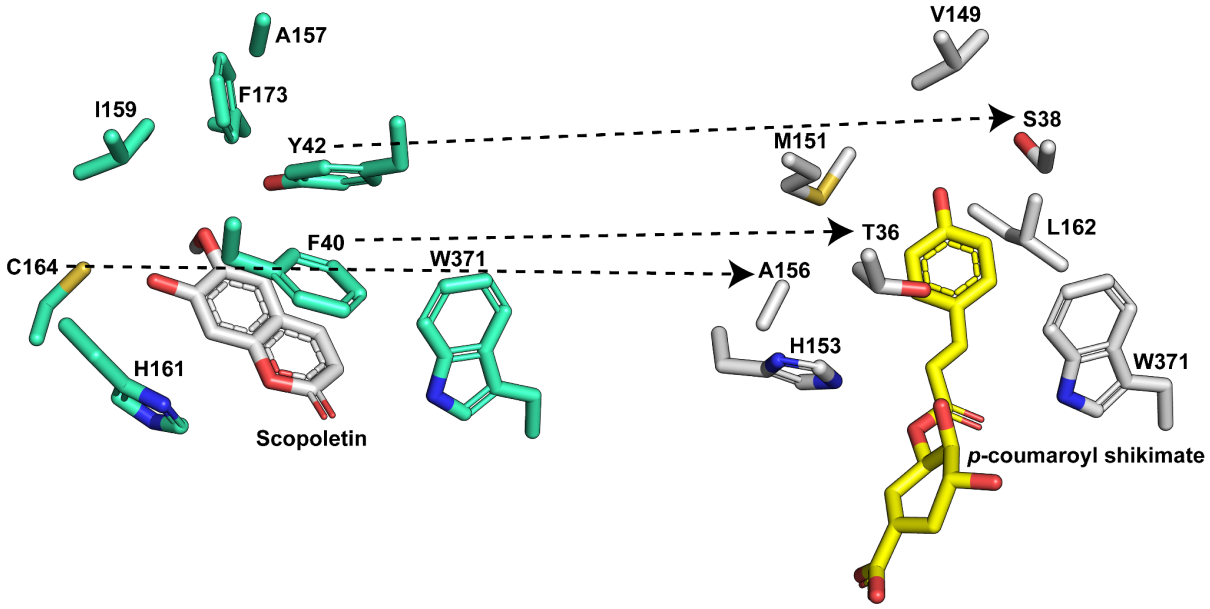
Supplementary Fig. 9 | Observation of Ca^{2+} ion in *At*COSY structures and assay for Ca^{2+} dependency. (**a**) Combined LC-MS extracted ion chromatograms (XICs) of 135.04512, 179.03494 *m/z*; **6** and 161.02428 *m/z*; **7**. As umbelliferone production is spontaneously catalyzed by light, 4CL1 without *At*COSY sample serves as an experimental control for the coumarin synthase activity of *At*COSY. 4CL1+COSY sample shows complete conversion of 2OHpCA into umbelliferone.Assay with Ca-dependency (**b**) Unpaired, two-tailed t-test for statistical comparison of the coumarin synthase activity compared to the respective *At*4CL1 only control. 3 replicates of 0 μM and 100 μM [$CaCl_2$] were added for the enzyme assays. (**c**) Ca^{2+} ion present in the scopoletin-bound structure. Backbone carbonyls of R211 and L214 coordinate the Ca^{2+} ion.



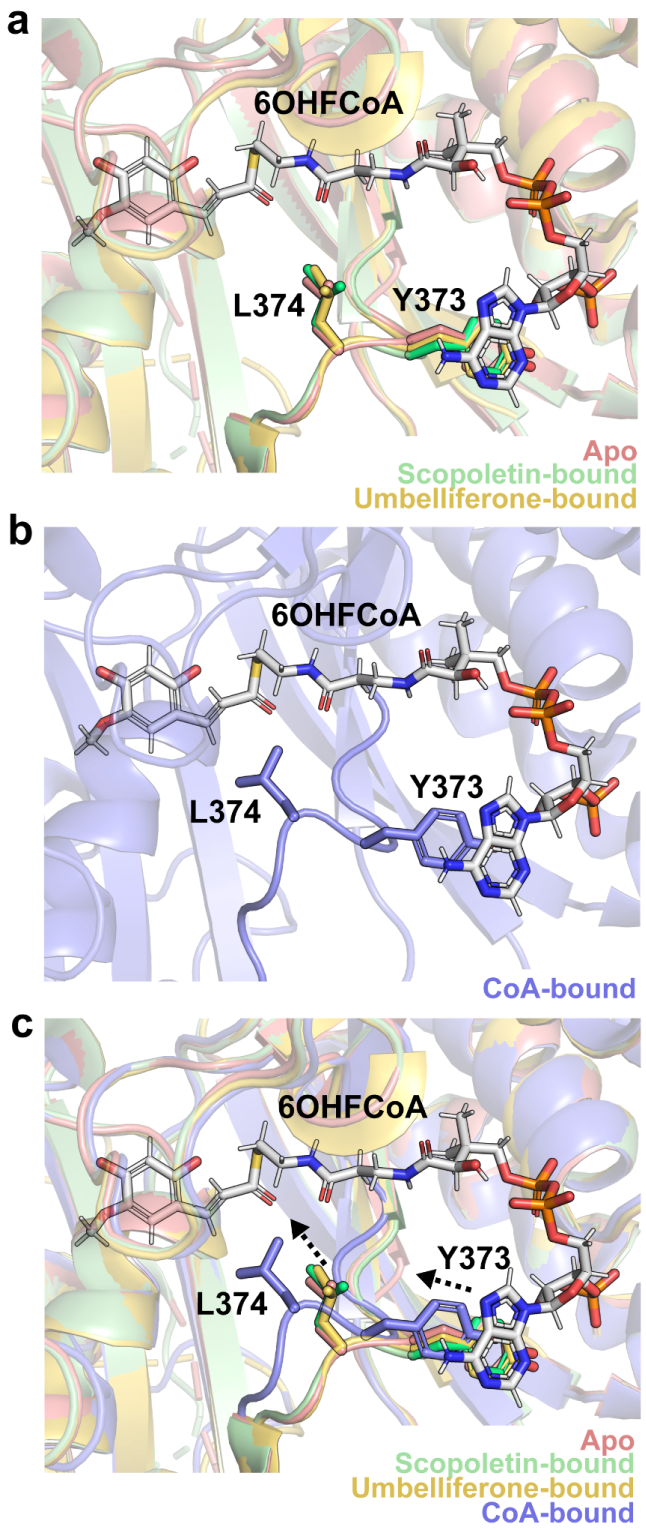
Supplementary Fig. 10 | Assay of divalent cation-dependent coumarin synthase activity in *At*COSY. The enzyme assay was carried out by adding each of the following sources of divalent cations $CaCl_2$, $ZnCl_2$, $NiCl_2$, $MnCl_2$, $CoCl_2$, $FeSO_4$ and $CuCl_2$ at 100 µM each. Single reaction monitoring (SRM) operated in negative ionization mode was used to detect the $[M-H]^-$ *m/z* values of the expected compounds: 165.1 *m/z* to 150.1 *m/z*; **1** and 191.1 *m/z* to 176.1 *m/z;* **2** both with mass windows of 0.5 *m/z*. The peak area for each sample is reported as relative intensity. This mass spectrometric analysis was performed on a TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific).



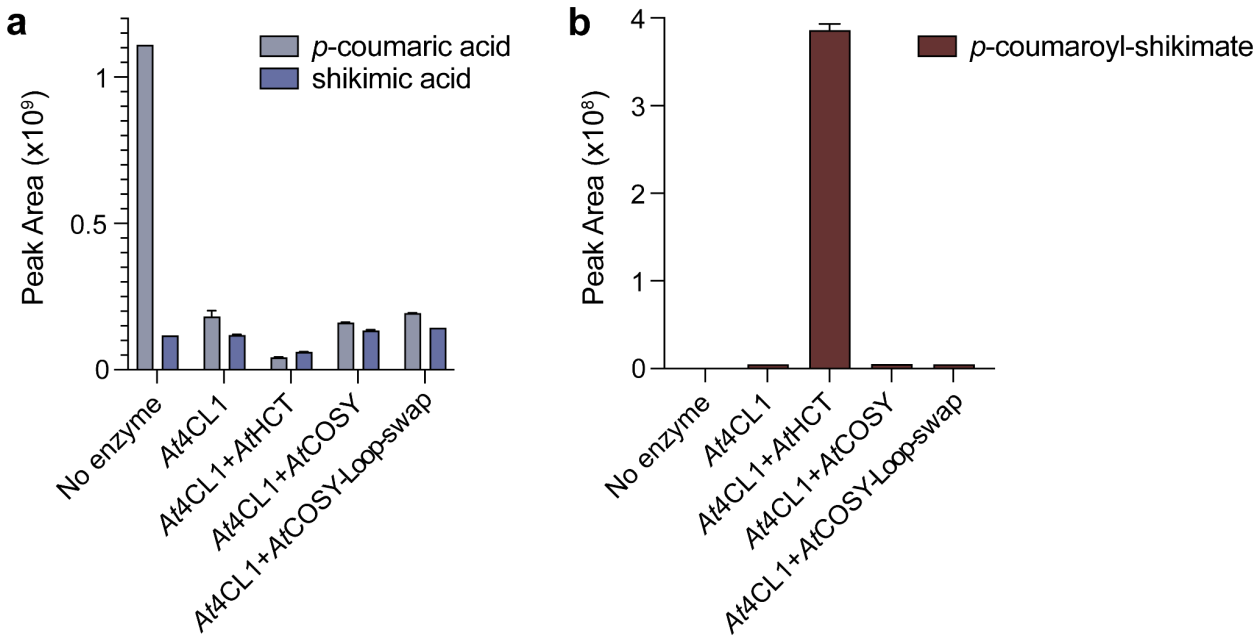
Supplementary Fig. 11 | Qualitative LC-MS analysis of *At*COSY mutant assays. XICs showing the highest conversion of 6OHFA; **1** to scopoletin; **2** for each of the mutant assays are selected and highlighted in various colors corresponding to each mutant (n=3). W371H, W371A, W371V, W371M, G166A, G166L, Y373A, Y373F, L374A, and Loop-swap mutant assays were performed separately from other mutants with proper WT control and scale-adjusted based on the WT assays. Mass windows used for displaying the XICs: 165.05556 m/z and 209.04549 m/z; **1**, and 191.03443 m/z; **2**. There are no side products observed in these assays. (**a**)



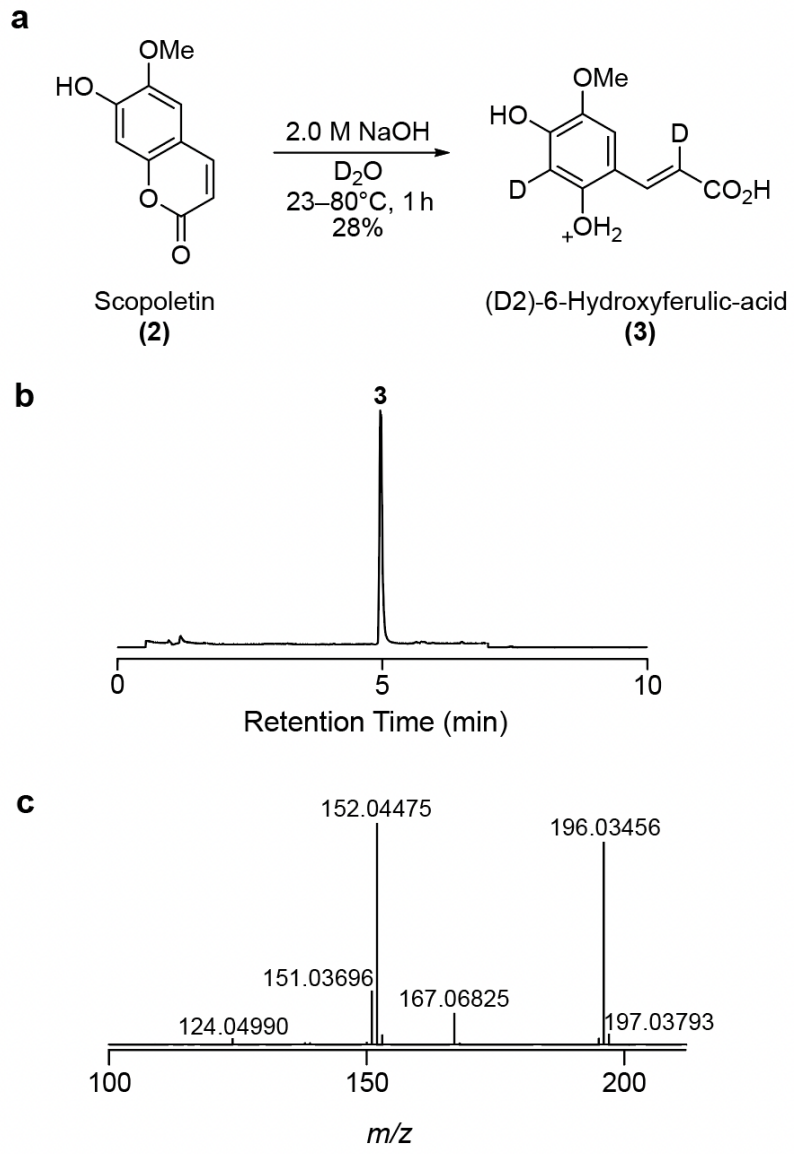
Supplementary Fig. 12 | Structural comparison between *At*COSY and *At*HCT. The active site comparison between *At*COSY and *At*HCT revealed several substitutions at corresponding scopoletin stabilizing residues (F40->T36; Y42->S38; C164->A156).



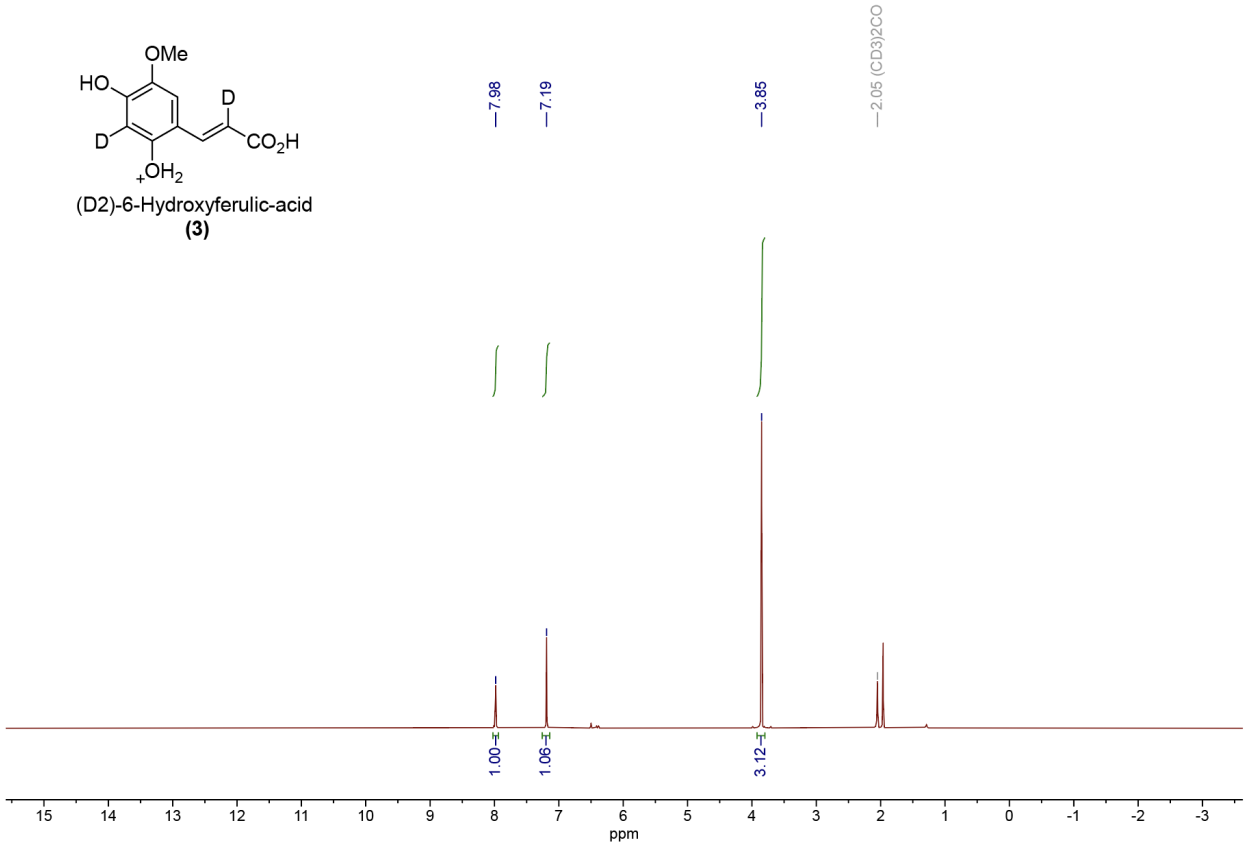
Supplementary Fig. 13 | Structural alignment of active-site loop in obtained *At*COSY structures. The dynamic loop containing Tyr373 and Leu374 is shown relative to *trans*-6OHFCoA in structural alignment of (**a**) apo, scopoletin-bound, and umbelliferone-bound *At*COSY structures; (**b**) CoA-bound *At*COSY structure (**c**) all *At*COSY structures reported in this study. Arrows indicate change in orientation of Tyr373 and Leu374 in CoA-bound *At*COSY structure compared to others in structural alignment.



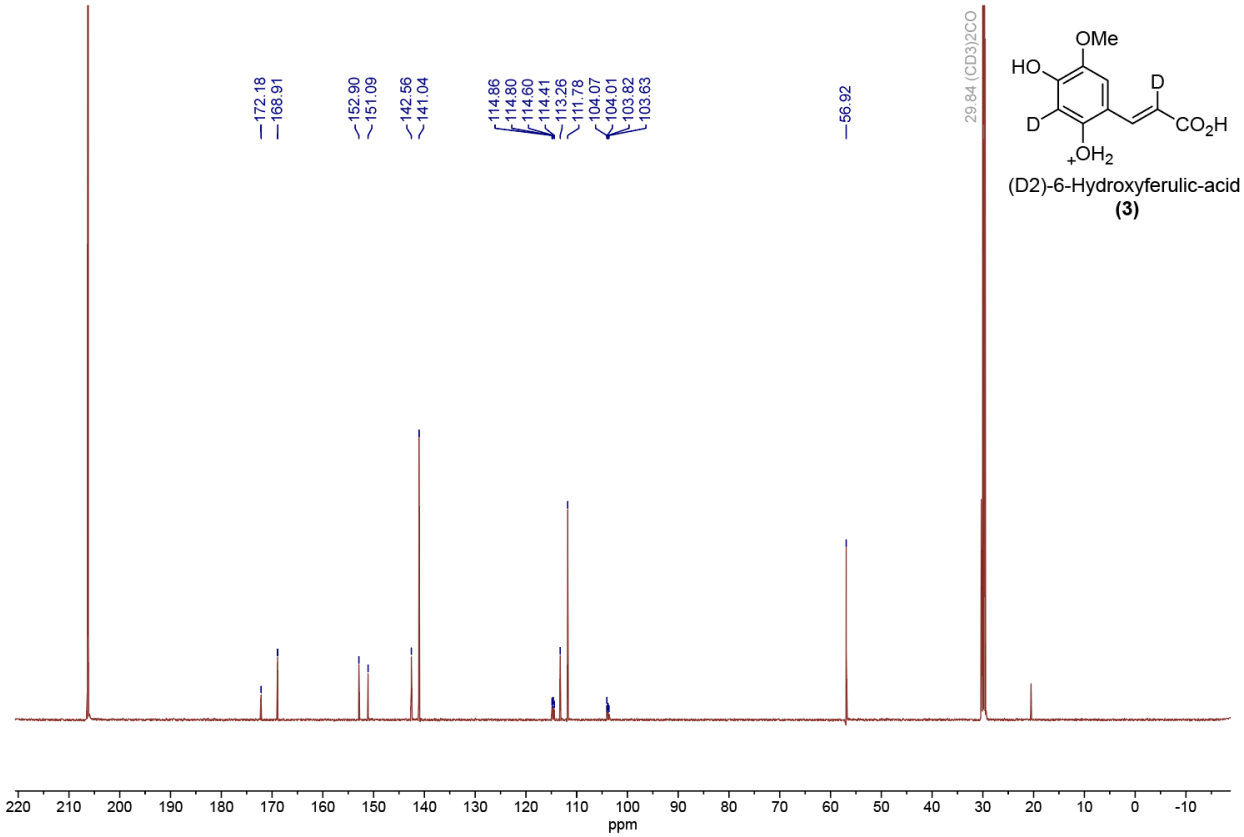
Supplementary Fig. 14 | Assay of *p*-coumaroyl-shikimate producing activity in *At*COSY active-site-loop domain swap assay. (**a**) Total ion count of the LC-HRAM-MS peak area of *p*-coumaric acid ($[M-H]^-$ = 163.0400 *m/z*) and shikimic acid ($[M-H]^-$ = 173.0457 *m/z*) for no enzyme, *At*4CL1-only, *At*4CL1+*At*HCT, *At*4CL1+*At*COSY, and *At*4CL1+*At*COSY-Loop-swap mutant. The error bars represent standard error of mean (n=3). (**b**) Total ion count of the LC-HRAM-MS peak area of *p*-coumaroyl-shikimate ($[M-H]^-$ = 319.0823 *m/z*). The error bars represent standard error of mean (n=3). Unpaired, two-tailed t-test for statistical comparison of the *p*-coumaroyl-shikimate producing activity between *At*4CL1+*At*COSY-Loop-swap mutant and *At*4CL1-only sample indicated no significant difference.



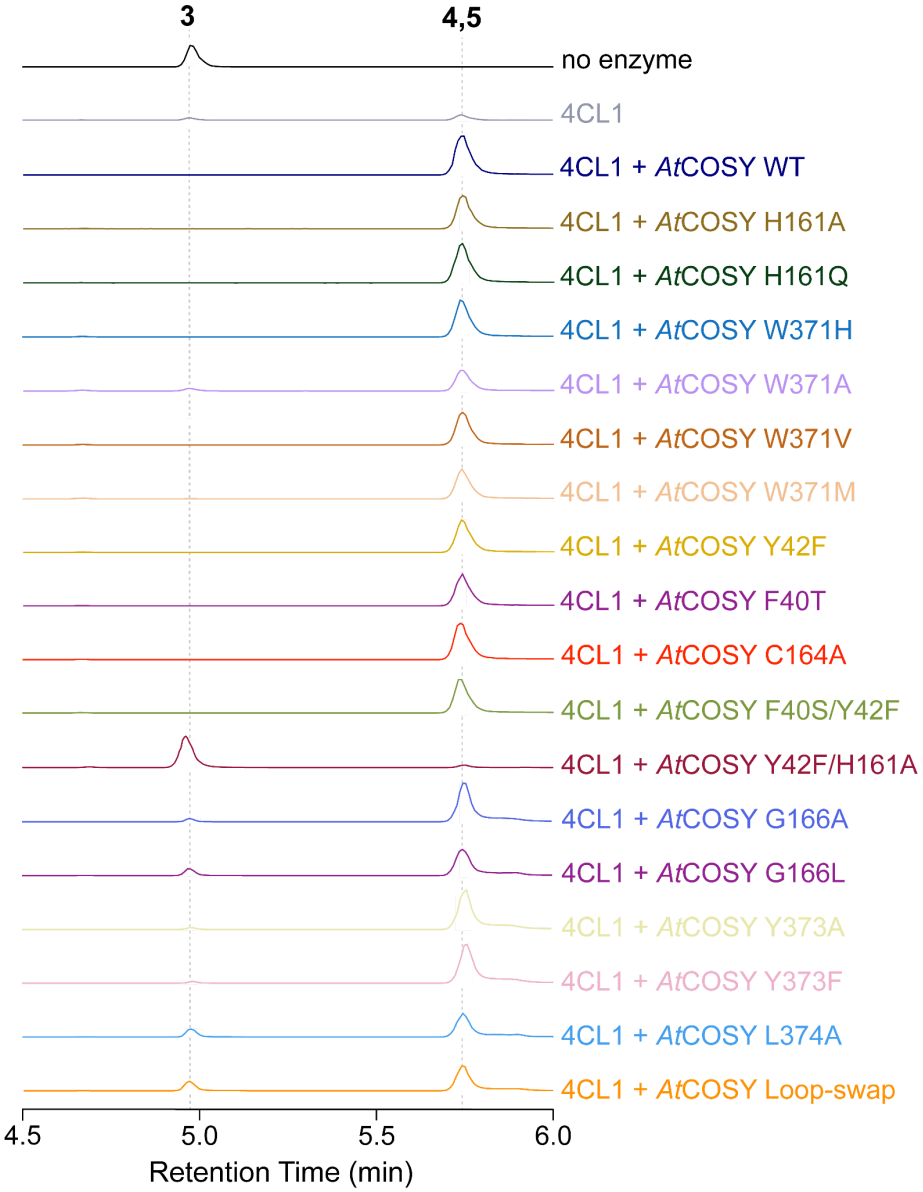
Supplementary Fig. 15 | Chemical synthesis of D2-6OHFA. (**a**) Reaction diagram of saponification reaction in D_2O of scopoletin (**2**) to (D2)-6-hydroxyferulic-acid (**3**). (**b**) TIC of the purified **3** shows no significant side-products from the deuterated-saponification reaction and purification process. (**c**) MS^2 fragmentation spectra of 211.0581 *m/z*; **3**, with the major peaks 196.03456 *m/z* and 152.04475 *m/z*.



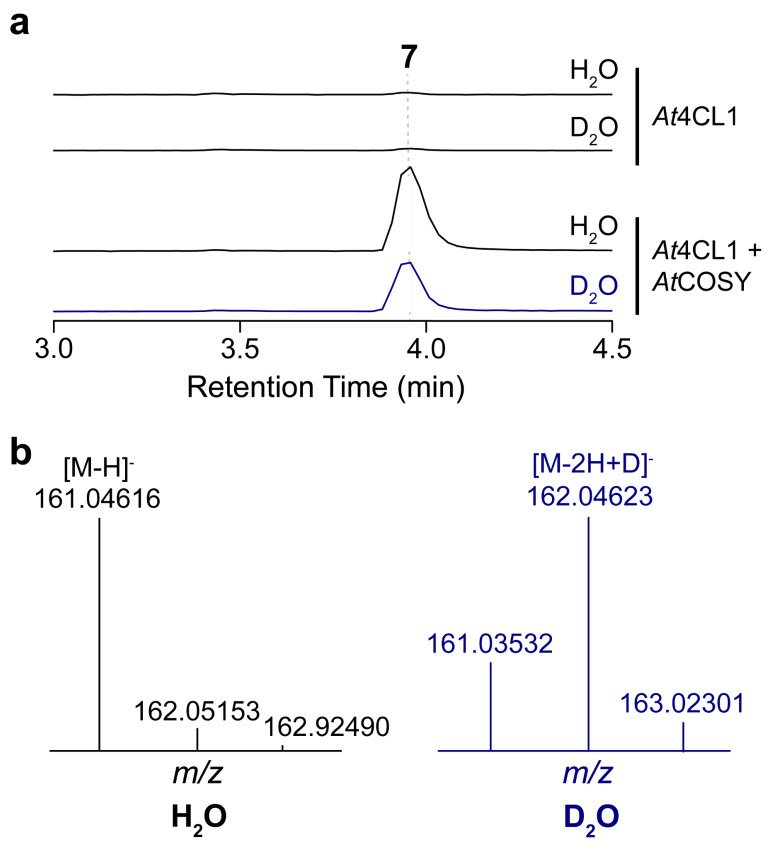
Supplementary Fig. 16 | ^{1}H-NMR spectra of D2-6OHFA. ^{1}H NMR (acetone-d6, 500 MHz) δ 7.98 (s, 1H), 7.19 (s, 1H), 3.85 (s, 3H).



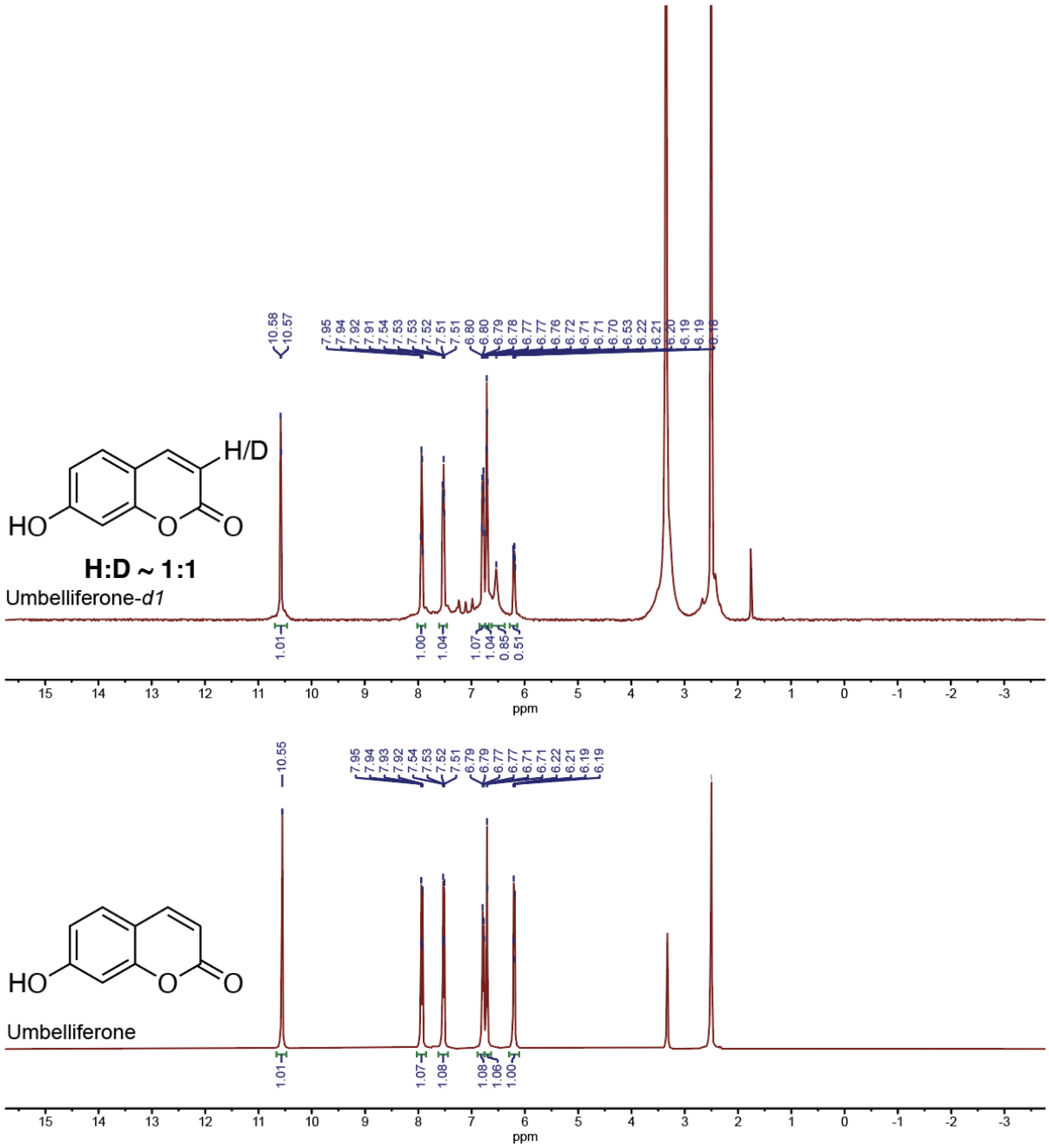
Supplementary Fig. 17 | ^{13}C-NMR spectra of D2-6OHFA. ^{13}C NMR (acetone-d6, 125 MHz) δ 172.2, 168.9, 152.9, 151.1, 142.6, 141.0, 114.9, 114.8, 114.6, 114.4, 113.3, 111.8, 104.1, 104.0, 103.8, 103.6, 56.9.



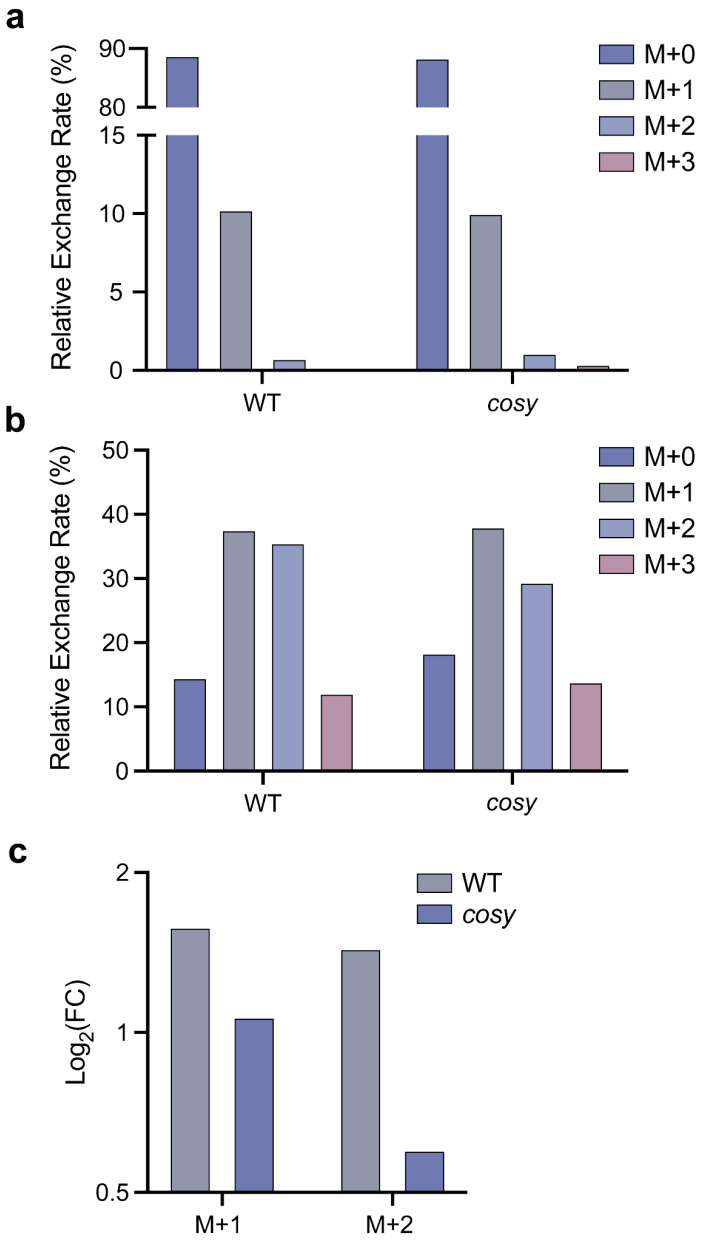
Supplementary Fig. 18 | Qualitative LC-MS analysis of *At*COSY mutant assays against D2-6OHFA. XICs showing the highest conversion of D2-6OHFA (**3**) to the product (**4, 5**) for each of the mutant assays are selected and highlighted in various colors corresponding to each mutant (n=2). G166A, G166L, Y373A, Y373F, L374A, and Loop-swap mutant assays were performed separately from other mutants with proper WT control and scale-adjusted based on the WT assay. Mass windows used for displaying the XICs: **3**, 167.06804 m/z; **4**, 193.04747 m/z; and **5**, 192.04114 m/z. There are no side products observed in these assays.



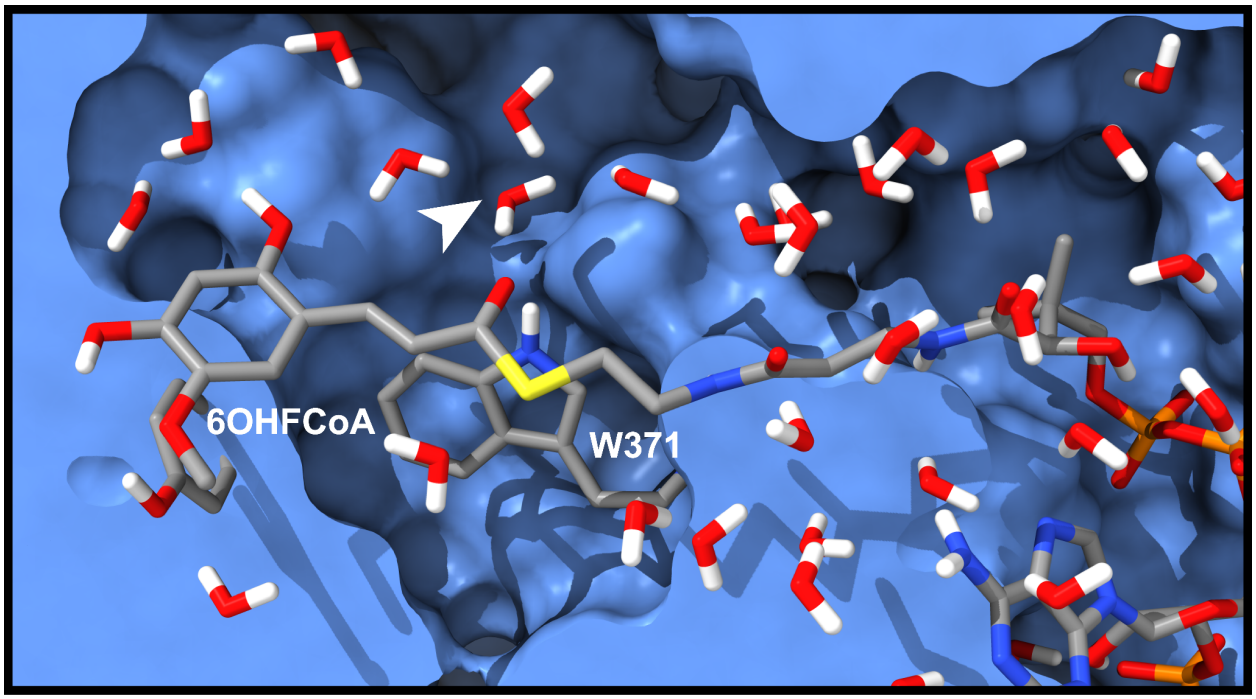
Supplementary Fig. 19 | Enzyme activity of WT *At*COSY in D_2O against 2OHpCACoA. (**a**) Single ion monitoring (SIM) of umbelliferone in enzymatic assay with *At*4CL1 (top two traces) and *At*4CL1+WT *At*COSY (bottom two traces) in 100% H_2O or 66% D_2O. LC-MS trace in blue showcases the umbelliferone peak with H/D-exchange. (**b**) MS^1 spectra of umbelliferone products obtained under SIM mode. WT *At*COSY reaction in 100% H_2O yielded umbelliferone product with 161.04616 *m/z* (black) $[M-H]^-$, whereas in 66% D_2O yielded product with 162.04623 *m/z* (blue) corresponding to the $[M-2H+D]^-$ value of umbelliferone.



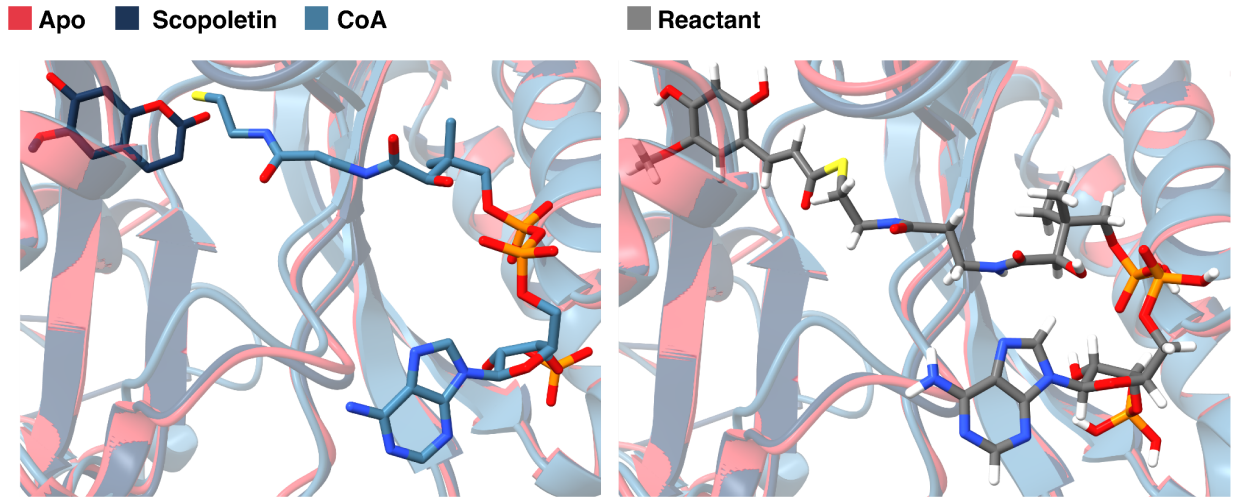
Supplementary Fig. 20 | NMR spectroscopy structural analysis of (7).



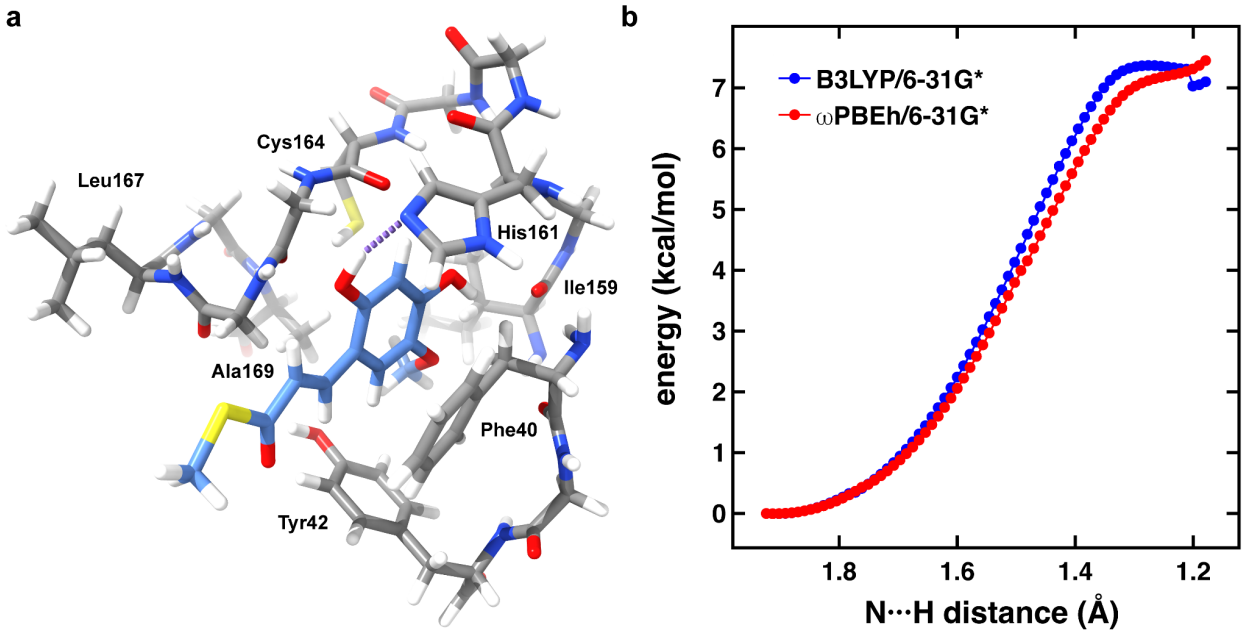
Supplementary Fig. 21 | *In planta* feeding of D_2O for WT and *cosy A. thaliana*. Relative H/D exchange rate for scopoletin (M = 191.03508 m/z) and its corresponding isotopes (M+1 = 192.04123 m/z; M+2 = 193.04744 m/z; M+3 = 194.05373 m/z) observed in (**a**) WT and *cosy* roots grown in H_2O, and (**b**) WT and *cosy* roots grown in 100% D_2O. (**c**) Log_2 fold-change of M+1 and M+2 peaks compared to the corresponding parent scopoletin peak in WT and *cosy* roots grown in 100% D_2O. All assays were performed in 3 technical replicates and small error bars are not shown at this scale.



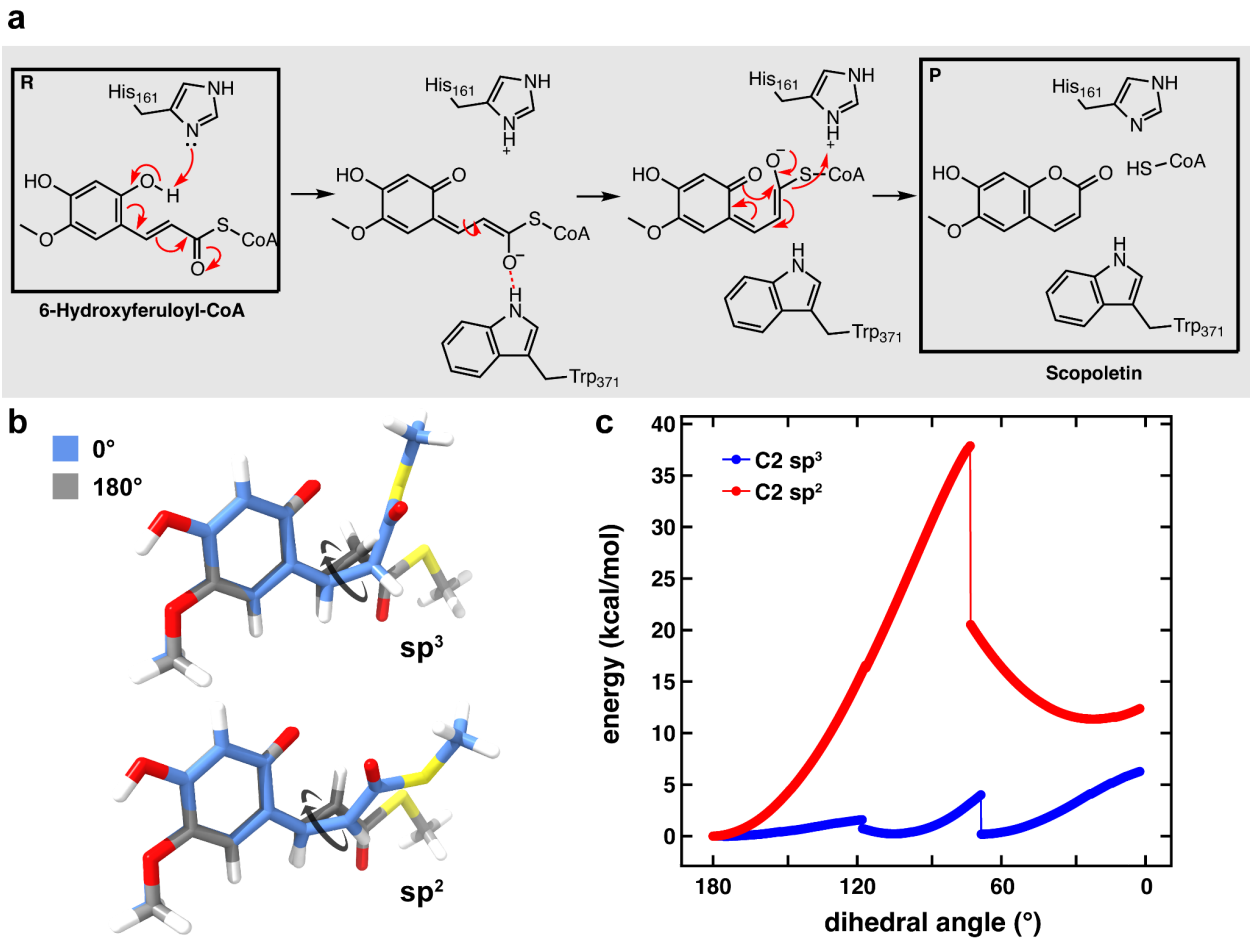
Supplementary Fig. 22 | Water configuration from MD simulations. The water configuration of the active site from the clustered MD simulation. A white arrow indicates the location of the proposed catalytic hydroxide. The substrate 6OHFCoA and Trp371 are labeled. The protein environment is highlighted as light blue. Atoms are colored as follows: carbons in gray, sulfur in yellow, nitrogen in blue, oxygen in red, and hydrogen in white.



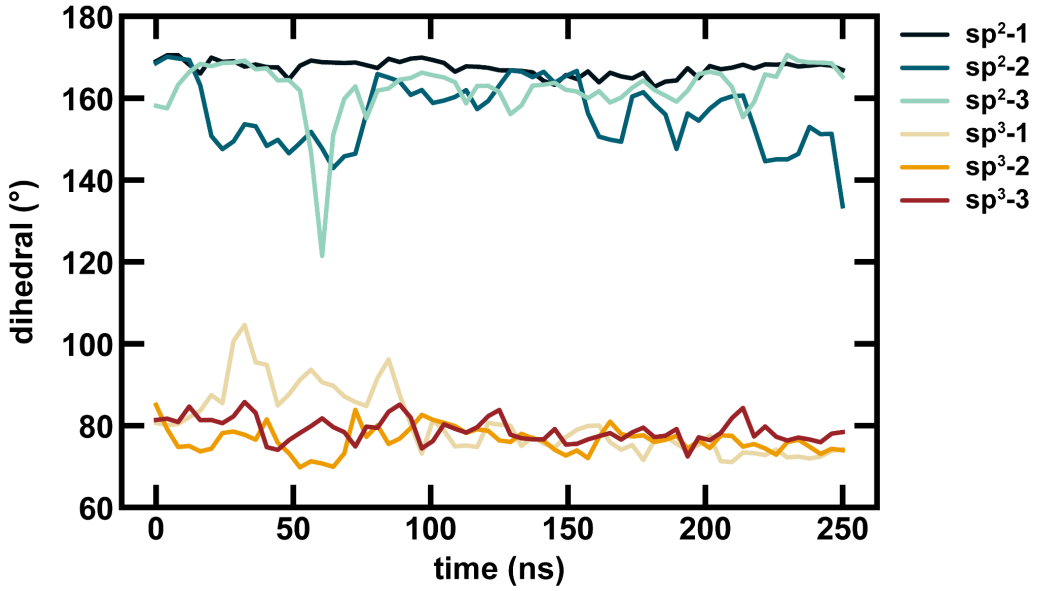
Supplementary Fig. 23 | Modeling the reactant 6-Hydroxy-feruloyl-CoA. (**left**) A depiction of the scopoletin-bound, CoA-bound, and apo crystal structures superimposed. Scopoletin is depicted with black carbons, and CoA is depicted with blue carbons. (**right**) The final 6-Hydroxy-feruloyl-CoA structure modeled by adding hydrogens and connecting scopoletin and CoA and performing a geometry optimization at the B3LYP/6-31G* level of theory.



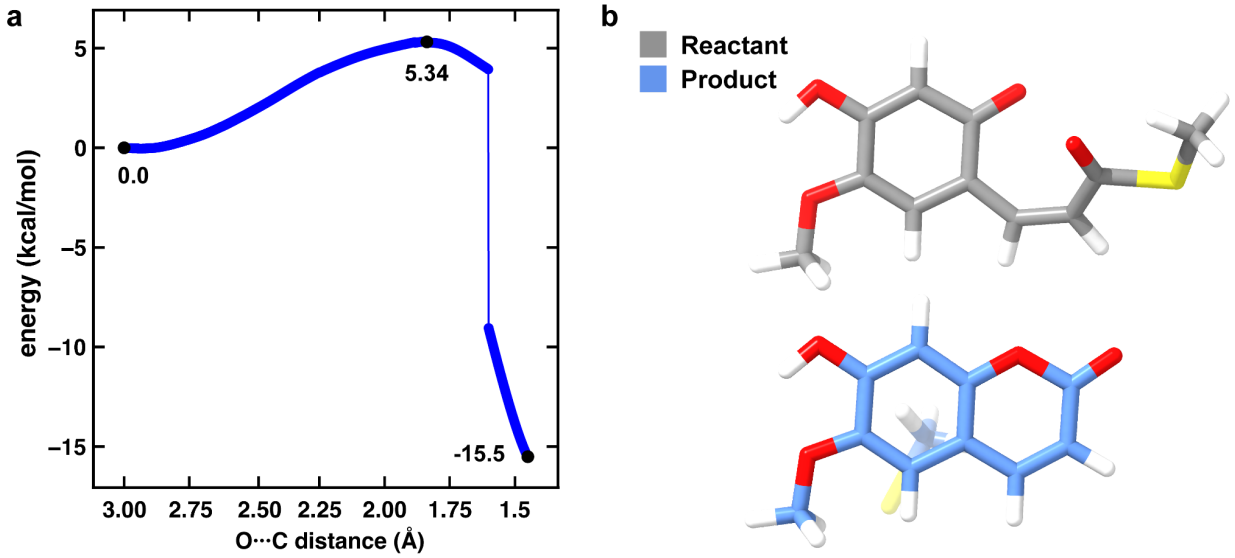
Supplementary Fig. 24 | Dissociation of a hydrogen from 6OHFCoA to His161. (**a**) The minimal cluster model used for the constrained geometry scans. Substrate carbons are colored light blue and the hydrogen bond between the substrate and His161 is shown as a dotted purple line. (**b**) Constrained geometry energy scans following the N···H coordinate between the substrate and His161. The energies were calculated with both B3LYP (blue) and wPBEh (red).



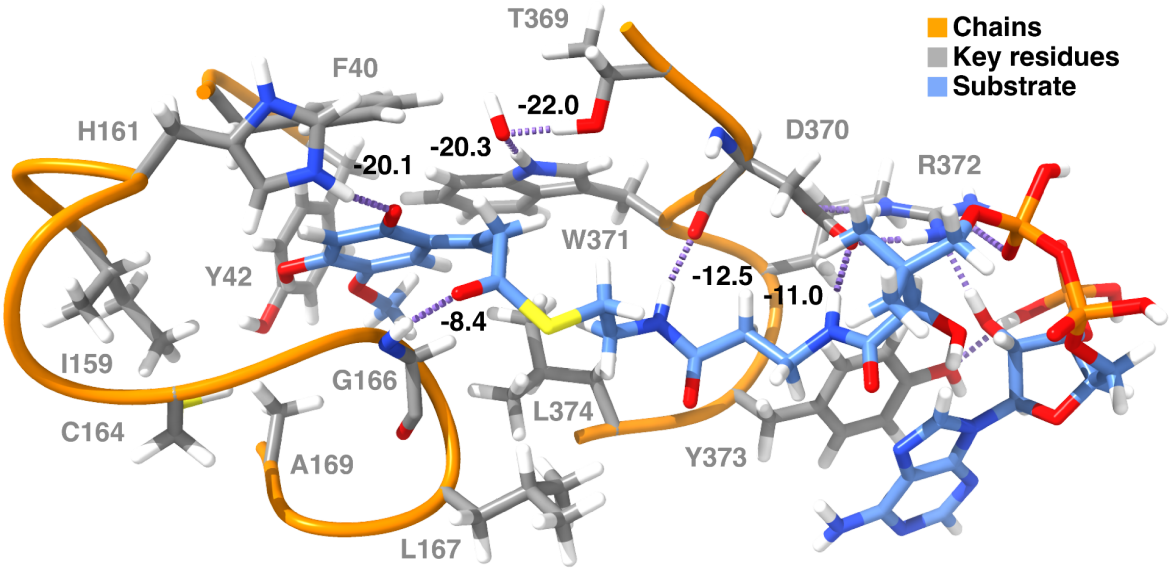
Supplementary Fig. 25 | Energetic evaluation of the *s-trans*-to-*s-cis* isomerization. (**a**) Proposed mechanism from Vanholme et al.[4] for the *trans*-to-*cis* isomerization of 6OHFCoA via intramolecular electron delocalization prior to lactonization. This mechanism does not entail proton exchange at C2. (**b**) (top) The structure of the C2-protonated substrate with the C1-C2(sp^3)-C3-C4 dihedral at 0° (gray) and 180° (blue). (bottom) The structure of the conjugated substrate with the C1-C2(sp^2)-C3-C4 dihedral at 0° (blue) and 180° (gray). (**c**) The energy profile calculated from the constrained geometry scan for the C2-protonated substrate (blue) and the conjugated substrate (red). The constrained geometry scan of the conjugated species reveals two peaks at 4.0 kcal/mol and 6.3 kcal/mol. The smaller peak corresponds with a brief steric clash between one of the C2 hydrogens and the ring carbonyl. The putative transition state identified for the isomerization of the conjugated species is 37.8 kcal/mol. The drop in energy for the C1-C2(sp^2)-C3-C4 dihedral scan occurs when the hydrogen on C2 is repositioned once the double bond is finally broken.



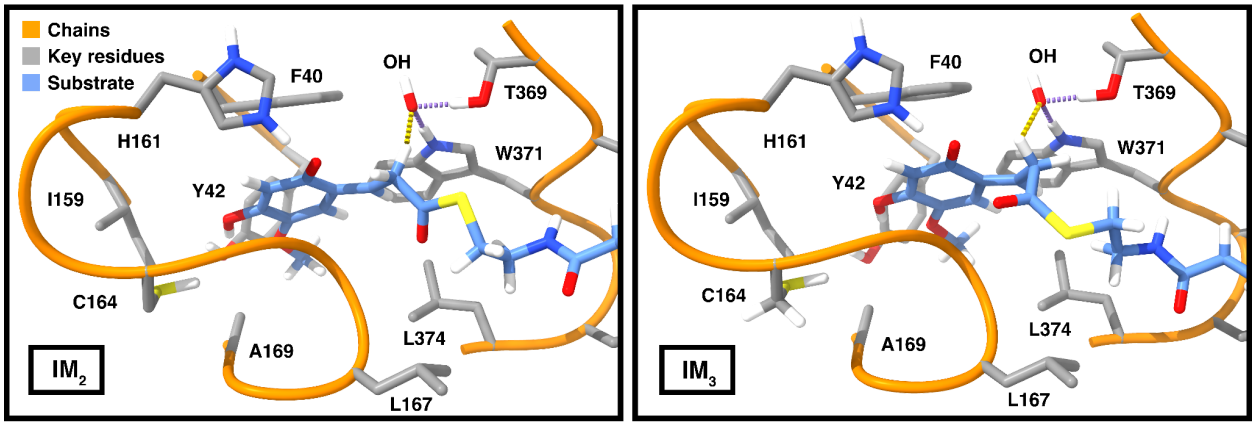
Supplementary Fig. 26 | Isomerization with either an sp^2 or sp^3 carbon at C2. 250 ns MD simulations were performed for *At*COSY with either an sp^2 (dark blue, blue, and light blue) or sp^3 carbon (red, orange, yellow) at C2 of the substrate. For each case, production dynamics were collected for three replicates at 2 ps intervals and averaged across 4 ns intervals to minimize noise. For all simulations, the substrate started with a dihedral angle of 180° for the C1-C2-C3-C4 dihedral. All replicates with a sp^3 carbon at C2 readily isomerized during the equilibration steps to a dihedral of approximately 80°, which remained relatively constant throughout the simulations.



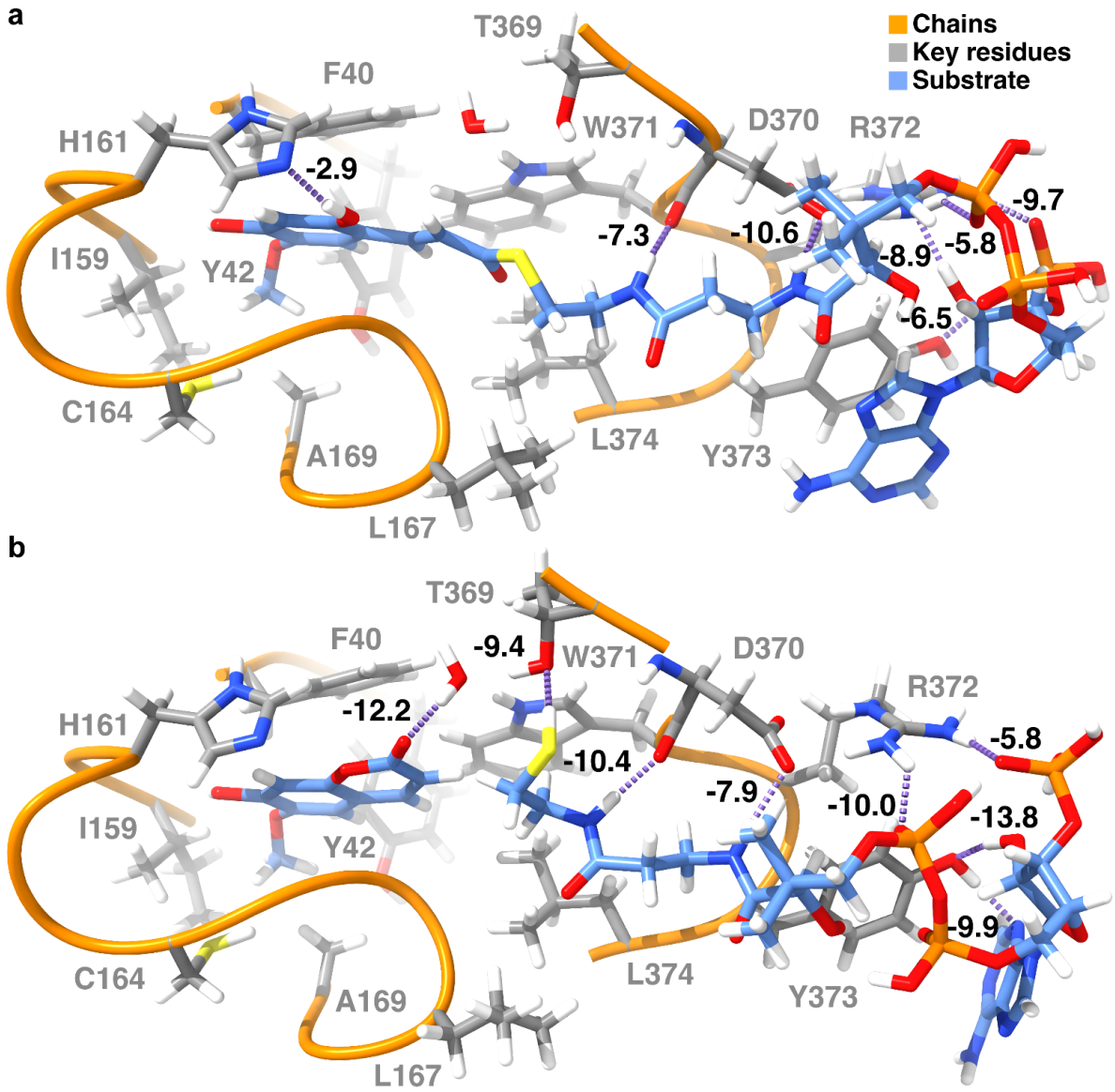
Supplementary Fig. 27 | Energetic evaluation of the lactonization step in implicit solvent with dielectric constant of water. (**a**) The energy profile calculated from the constrained geometry scan for lactonization following the trans to cis isomerization. The large drop in energy after the transition state is due to the breaking of C-S bond representative of CoA release. (**b**) (top) The structure of the starting cis conformation (gray). (bottom) The structure of the product following lactonization (blue).



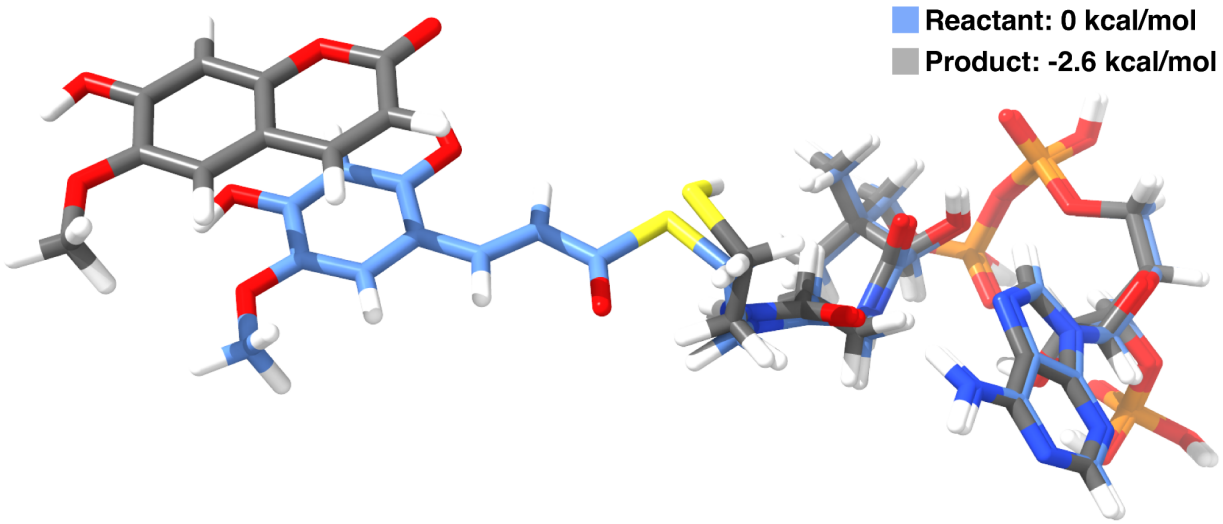
Supplementary Fig. 28 | Quantum mechanical hydrogen bonding analysis of the high energy IM3 intermediate. A hydrogen bond with Gly166, unique to the *s-cis* isomer, stabilizes the high energy conformation by -8.4 kcal/mol. Trp371 and Thr369 stabilize the hydroxide species by -20.3 and -22.0 kcal/mol respectively. Continuous chains of amino acids are shown as orange ribbons and included side chains are shown with gray carbons. The substrate is depicted with blue carbons. Hydrogen bonds are depicted as dotted purple lines. Hydrogen bond strengths are labeled in black, and residues are labeled in gray.



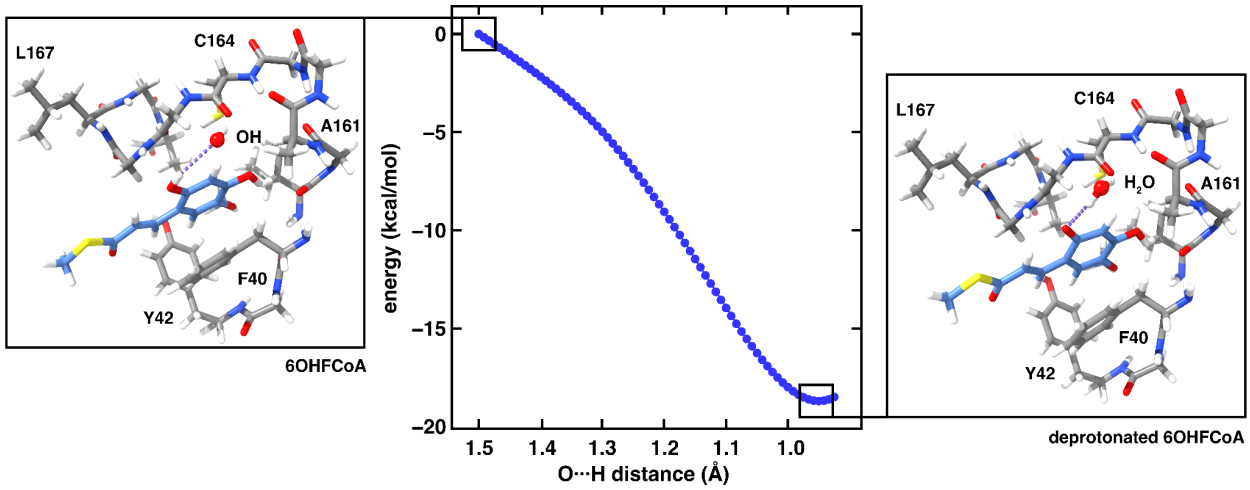
Supplementary Fig. 29 | Positioning of the catalytic hydroxide in intermediates IM_2 and IM_3. Hydrogen bonds between the proposed catalytic hydroxide and Thr369 and Trp371 are shown as purple dashed lines and the distance between the hydroxide and the target hydrogen is shown as a yellow dashed line. Continuous chains of amino acids are shown as orange ribbons and included side chains are shown with gray carbons. The substrate is depicted with blue carbons. Hydrogen bond strengths are labeled in black, and residues are labeled in gray.



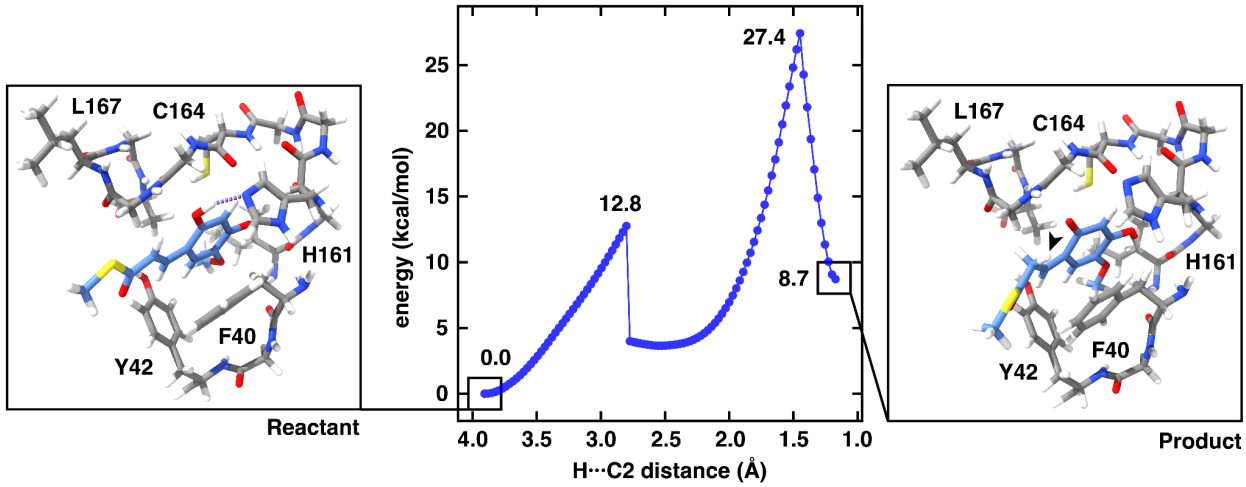
Supplementary Fig. 30 | Quantum mechanical hydrogen bonding analysis of the reactant and product states. A depiction of the hydrogen bonds between the substrate and the protein environment in the (**a**) reactant state and (**b**) product states. Hydrogen bond energies are reported in kcal/mol. Continuous chains of amino acids are shown as orange ribbons and included side chains are shown with gray carbons. The substrate is depicted with blue carbons. Hydrogen bonds are depicted as dotted purple lines. Hydrogen bond strengths are labeled in black, and residues are labeled in gray.



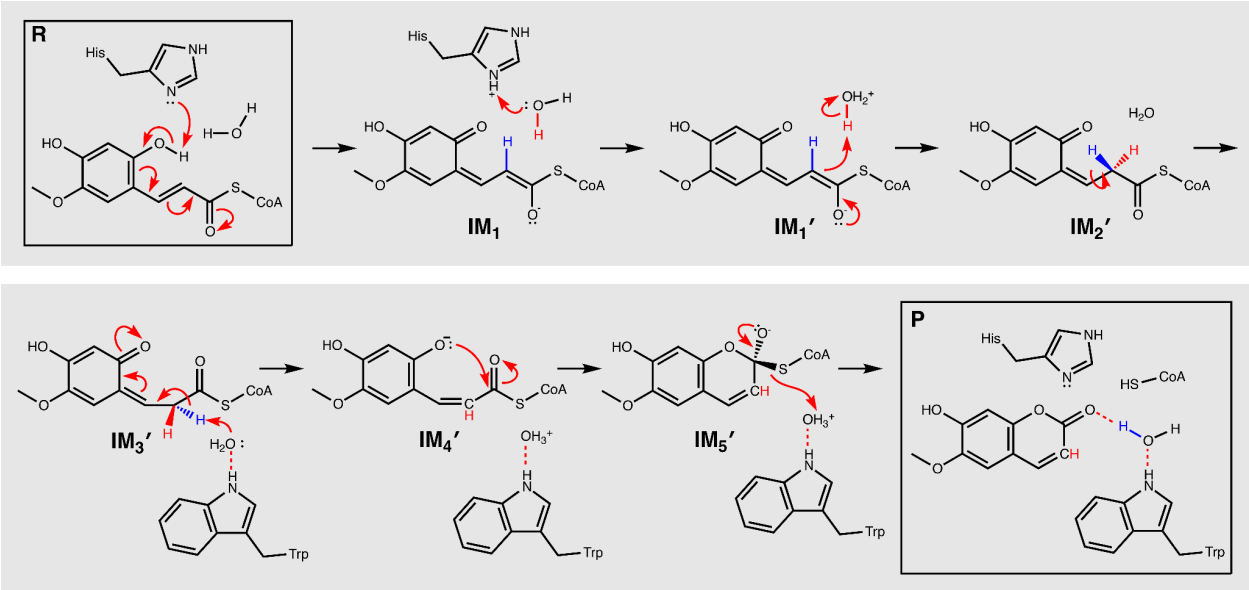
Supplementary Fig. 31 | Energetics of CoA release without the protein environment. The geometry optimized structure of the reactant (blue) and the products (grey) without the protein environment.



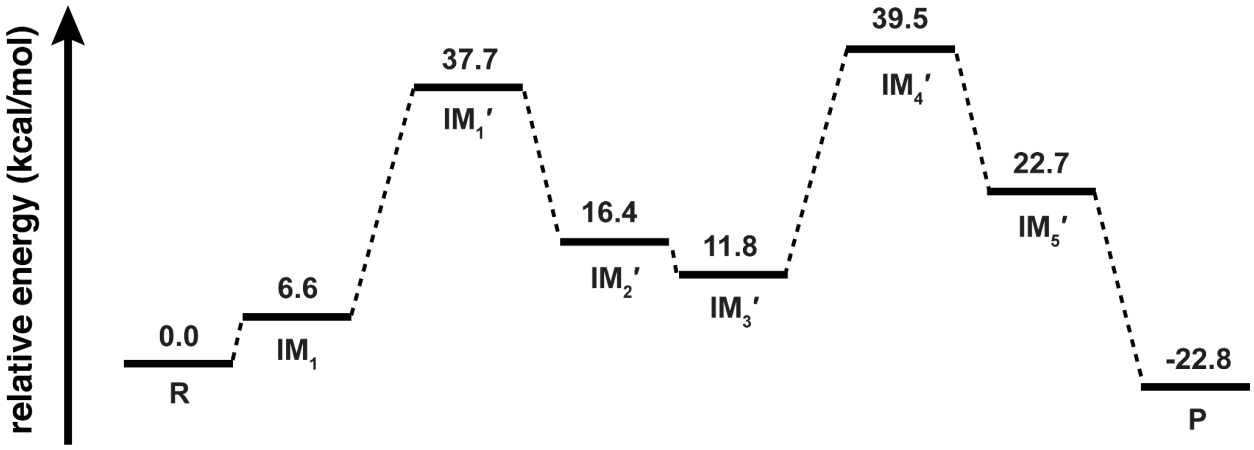
Supplementary Fig. 32 | Dissociation of a proton from 6OHFCoA to hydroxide in His161Ala mutant. A constrained geometry scan was performed using a minimal cluster model where the distance between the target hydrogen of 6OHFCoA and the oxygen of a free hydroxide was used as a reaction coordinate and is indicated with a purple dashed line. 6OHFCoA is highlighted with light blue carbons whereas the protein backbone carbons are colored in gray. Remaining atoms are colored as follows: sulfur in yellow, nitrogen in blue, oxygen in red, and hydrogen in white.



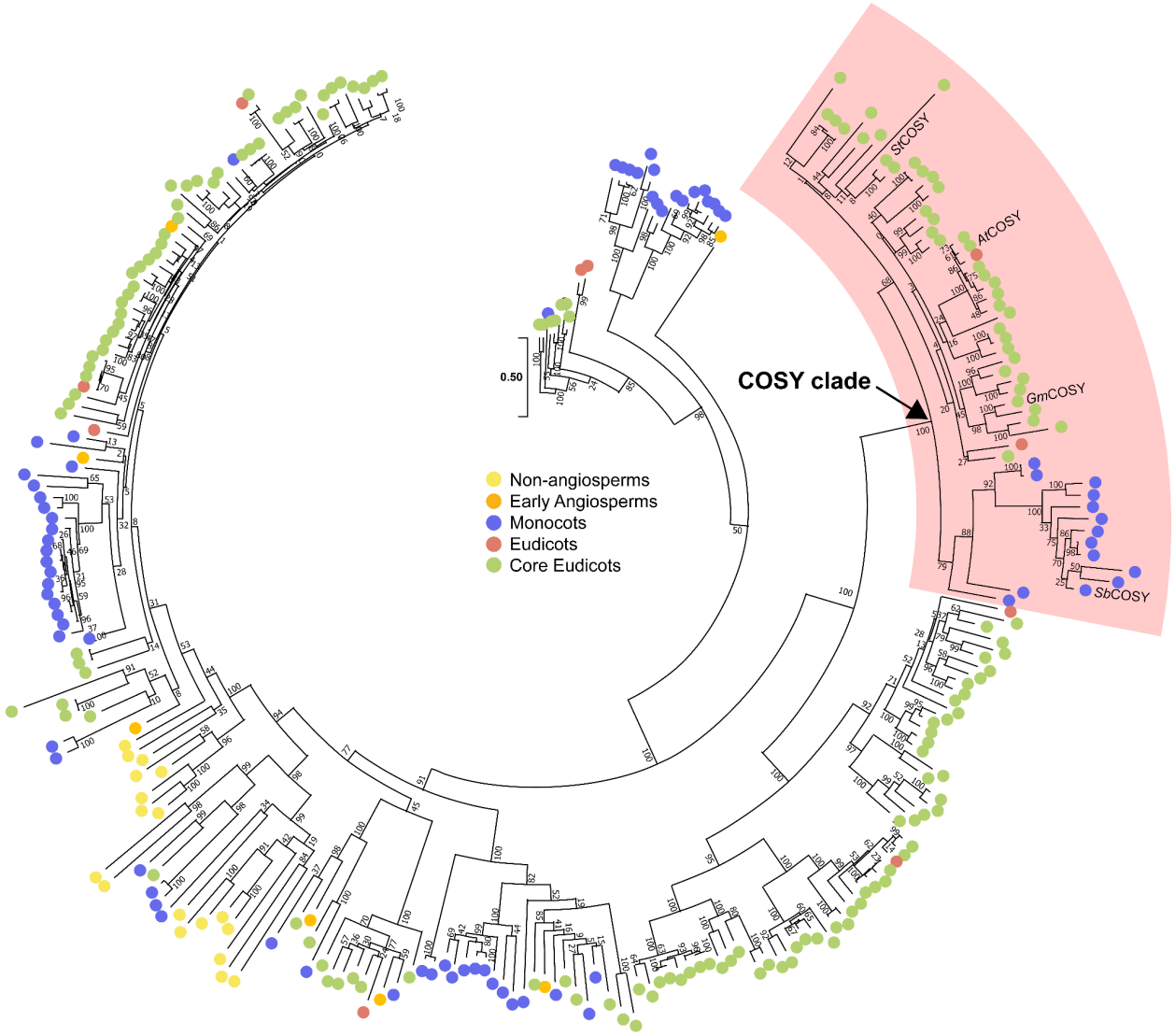
Supplementary Fig. 33 | Energetics of the intramolecular transfer of *o*-hydroxy hydrogen to C2 of 6OHFCoA. A constrained geometry scan of the alternate pathway for the protonation of C2 via the intramolecular transfer of a hydrogen atom was performed. The scan revealed two energetic barriers. The first barrier of 12.8 kcal/mol corresponds to the breaking of the strong hydrogen bond between the substrate and His161 and is shown as a purple dashed line. The second barrier of 27.4 kcal/mol corresponds to the protonation of C2. The position of the transferred hydrogen atom in the product is indicated with a black arrow. 6OHFCoA is highlighted with light blue carbons whereas the protein backbone carbons are colored in gray. Remaining atoms are colored as follows: sulfur in yellow, nitrogen in blue, oxygen in red, and hydrogen in white.



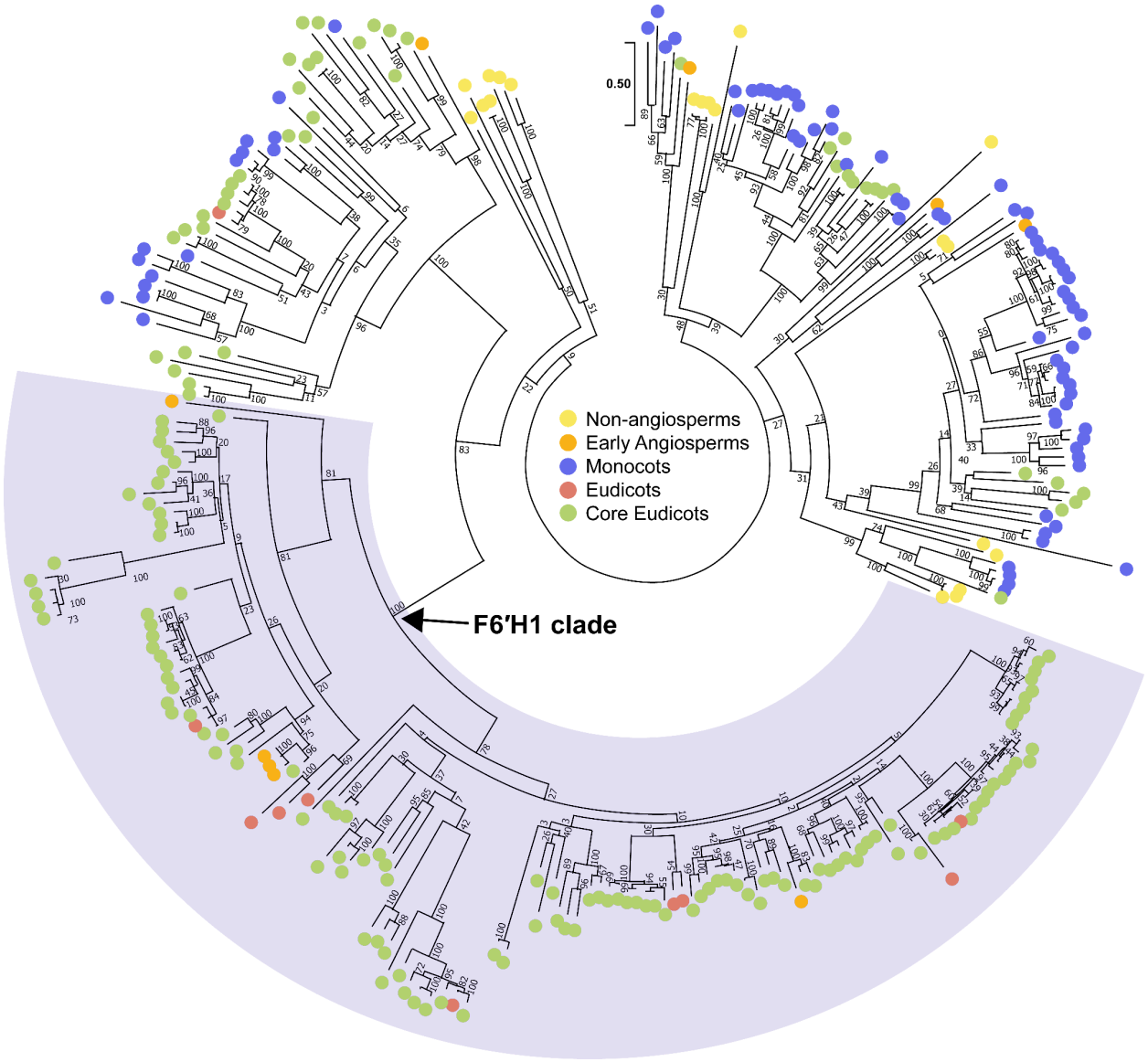
Supplementary Fig. 34 | Proposed alternative acid-base pathway. Schematic depicting an alternate acid-base mechanism that proceeds via a hydronium carrier. Stereospecific hydrogens are labeled either red or blue.



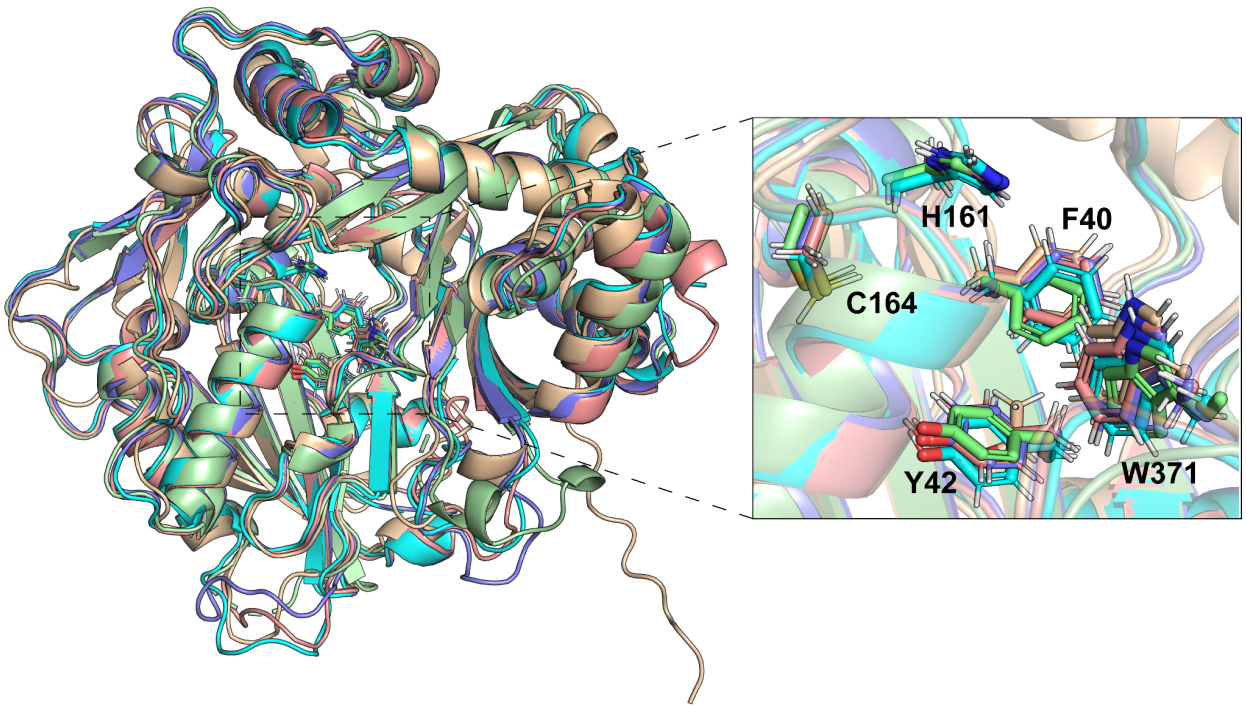
Supplementary Fig. 35 | Energetics of the alternative acid-base pathway. Energies are reported in kcal/mol. IM1′ and IM4′ are the highest energy intermediate. The highest energy intermediates contain hydronium.



Supplementary Fig. 36 | Maximum likelihood phylogenetic tree of *At*COSY with its homologous sequences. The phylogenetic tree of COSY is built with top 5, complete sequences from pBLAST search against select species in Phytozome v.12.1.6 proteome database. Bootstrap statistics (200 replicates) are indicated at the tree nodes. The scale measures evolutionary distance in substitutions per amino acid. Sequences from species that belong to non-angiosperms, early-angiosperms, monocots, eudicots, and core eudicots are colored as yellow, orange, purple, red, and green, respectively. The arrow points to the COSY-specific clade highlighted in pink, which includes homologous sequences from monocots, eudicots, and core eudicots. *St*COSY, *At*COSY, *Gm*COSY, and *Sb*COSY are labeled at the branch tips in the COSY clade.



Supplementary Fig. 37 | Maximum likelihood phylogenetic tree of *At*F6′H1 with its homologous sequences. The phylogenetic tree of F6′H1 is built with top 5, complete sequences from pBLAST search against select species in Phytozome v.12.1.6 proteome database. Bootstrap statistics (200 replicates) are indicated at the tree nodes. The scale measures evolutionary distance in substitutions per amino acid. Sequences from species that belong to non-angiosperms, early-angiosperms, monocots, eudicots, and core eudicots are colored as yellow, orange, purple, red, and green, respectively. The arrow points to the F6′H1-specific clade highlighted in light-purple which includes homologous sequences from early-angiosperms, eudicots, and core eudicots.



Supplementary Fig. 38 | Structural alignment of apo-*At*COSY and predicted structures of *At*COSY, *Gm*COSY, *St*COSY, and *Sb*COSY. AlphaFold2.0 was used to predict the structures. The RMSD values were calculated by the align function on PyMOL; *At*COSY (blue) = 0.992, *St*COSY (cyan) = 0.952, *Gm*COSY (pink) = 0.932, and *Sb*COSY (NCBI Accession # XM_002437812.2) (beige) = 1.389. The crystallized Apo-*At*COSY structure is portrayed in pale green. Active-site-lining residues, Phe40, Tyr42, His161, Cys164, and Trp371 numbered according to *At*COSY, exhibit similar positioning and are shown in sticks. *Sb*COSY harbors a Y42T substitution.

Supplementary References

1. Swift, M. L. GraphPad Prism, Data Analysis, and Scientific Graphing. *Journal of Chemical Information and Computer Sciences* vol. 37 411–412 Preprint at https://doi.org/10.1021/ci960402j (1997).
2. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
3. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–4 (2014).
4. Vanholme, R. *et al.* COSY catalyses trans–cis isomerization and lactonization in the biosynthesis of coumarins. *Nature Plants* vol. 5 1066–1075 Preprint at https://doi.org/10.1038/s41477-019-0510-0 (2019).