Evolution and molecular basis of substrate specificity in a 3-ketoacyl-CoA synthase gene cluster from *Populus trichocarpa*

Jeff Y. Chen[1,2] **, Arishba Mumtaz**[1] **, and Eliana Gonzales-Vigil**[1,2,*]

From the [1]Department of Biological Sciences, University of Toronto Scarborough, Toronto, Canada; [2]Department of Cell and Systems Biology, University of Toronto, Toronto, Canada

Edited by Joseph Jez

Very long chain fatty acids (VLCFAs) are precursors to sphingolipids, glycerophospholipids, and plant cuticular waxes. In plants, members of a large 3-ketoacyl-CoA synthase (KCS) gene family catalyze the substrate-specific elongation of VLCFAs. Although it is well understood that KCSs have evolved to use diverse substrates, the underlying molecular determinants of their specificity are still unclear. In this study, we exploited the sequence similarity of a KCS gene cluster from *Populus trichocarpa* to examine the evolution and molecular determinants of KCS substrate specificity. Functional characterization of five members (PtKCS1, 2, 4, 8, 9) in yeast showed divergent product profiles based on VLCFA length, saturation, and position of the double bond. In addition, homology models, rationally designed chimeras, and site-directed mutants were used to identify two key regions (helix-4 and position 277) as being major determinants of substrate specificity. These results were corroborated with chimeras involving a more distantly related KCS, PtCER6 (the poplar ortholog of the *Arabidopsis* CER6), and used to show that helix-4 is necessary for the modulatory effect of PtCER2-like5 on KCS substrate specificity. The role of position 277 in limiting product length was further tested by substitution with smaller amino acids, which shifted specificity toward longer products. Finally, treatment with KCS inhibitors (K3 herbicides) showed varying inhibitor sensitivities between the duplicated paralogs despite their sequence similarity. Together, this work sheds light on the molecular mechanisms driving substrate diversification in the KCS family and lays the groundwork for tailoring the production of specific VLCFAs.

Very long chain fatty acids (VLCFAs) have a backbone of 20 or more carbons and are precursors for important cellular lipids such as sphingolipids, glycerophospholipids, and triacylglycerols (1–3). VLCFAs and their derivatives regulate a variety of physiological and developmental processes in eukaryotes (4–7). In higher plants, VLCFAs play an additional role in the biosynthesis of cuticular waxes, a hydrophobic film deposited on aerial surfaces that provides protection against environmental stressors such as desiccation, infection, and herbivory (8–10). Among most characterized plants, the predominant components of cuticular waxes are aliphatic, functionalized hydrocarbons derived from VLCFAs (11).

VLCFAs are biosynthesized by the iterative two-carbon elongation of C16-C18 long chain fatty acyl-CoAs (12). In plants, a key enzyme in VLCFA elongation is the β- or 3-ketoacyl-CoA synthase (KCS), an integral membrane protein that forms parts of the microsomal fatty acid elongation (FAE) complex (13, 14). The FAE complex consists of four core enzymes that each catalyze one of the four steps in VLCFA elongation (15). The first reaction, catalyzed by KCS, condenses a fatty acyl-CoA with malonyl-CoA to generate a 3-ketoacyl-CoA. Subsequent reduction of the carbonyl by ketoacyl-CoA reductase, dehydration of the hydroxyl group by hydroxyacyl-CoA dehydratase, and reduction of the double bond by enoyl-CoA reductase yield a fatty acyl-CoA extended by two carbons (16). This extended very long chain fatty acyl-CoA can be used for another round of elongation; however, this depends on the substrate specificity of the condensing enzyme.

Previous work has shown that KCS catalyzes the only substrate-specific step in the FAE complex, whereas the other three enzymes are generalists (17). Because of this, KCSs have expanded into a large, multigenic family with diverse substrates that have yet to be fully explored (18–20). To date, the substrates of characterized KCSs include fatty acids with lengths ranging from C16-C36, containing double bonds or hydroxyl groups (20–25). The spatiotemporal expression of KCSs with different substrate specificities modulates the VLCFA profiles of different tissues. For example, expression of KCS16 in *Arabidopsis* trichomes results in the production of exceptionally long VLCFAs (up to C38), whereas seed-specific expression of KCS18/FAE1 results in the accumulation of shorter unsaturated gondoic (C20:1) and erucic (C22:1) acid in storage triacylglycerols (17, 25). The epidermis expressed *Arabidopsis* KCS6/CER6/CUT1 produces VLCFAs longer than C26 that are important precursors to cuticular waxes, particularly in the stem (26, 27). The substrate specificity of CER6 has also been shown to be modulated by a binding partner, CER2, which modifies the elongation capability of CER6 to even longer VLCFAs (14, 28, 29). CER2s and their paralogs (termed CER2-likes) are part of the functionally diverse BAHD-acyltransferase family and have been characterized in Arabidopsis, rice, maize, sacred lotus, and poplar (28–33).

---

* For correspondence: Eliana Gonzales-Vigil, e.gonzalesvigil@utoronto.ca.

Substrate specificity of a KCS gene cluster

While studies on other fatty acid biosynthetic and modifying enzymes have explored the molecular basis of substrate specificity, such studies in KCS are limited by the lack of structural resources and commercially available very long chain fatty acyl-CoA substrates, and insolubility of membrane proteins (34–36). A way around this is to heterologously express KCS enzymes in yeast (*Saccharomyces cerevisiae*) and compare their VLCFA profiles to the empty control strain (37). The differences in VLCFA product profiles are indicative of the distinct substrate specificities of the expressed KCS. The yeast elongase complex is compatible with plant KCSs and can complete the other three reactions of the FAE complex (*i.e.*, reduction, dehydration, and reduction). Previous work has shown that substrate preference from *in vitro* assays with purified enzymes and assays comparing yeast profiles *in vivo* are comparable, indicating that heterologous expression in yeast can be used to assess activity toward different VLCFAs (12, 37, 38). With this approach, chimeric proteins expressed in yeast were used to compare the substrate specificity between the FAE1 orthologs from *Arabidopsis thaliana* and *Brassica napus*. The first 173 residues of the N-terminal region were shown to convert the product ratio of C20:1 to C22:1 from one KCS to another (37). It is presently unknown whether these results are generalizable to other KCSs, and regions that affect specificity toward unsaturated substrates or substrates longer than C22 have yet to be identified. Interestingly, distant members of the KCS family show differential inhibitor specificity toward K3 herbicides, although the molecular basis of this inhibition is also still unclear (38).

Recently, a large screen of cuticular waxes from a collection of wild genotypes of *Populus trichocarpa* led to the identification of a KCS with substrate preference toward mono-unsaturated VLCFAs (*PotriKCS1/PtKCS1*) (24). This study found an alkene homologous series (ranging in length from C_{23} to C_{31}) as major components of the abaxial wax of *P. trichocarpa* leaves. Moreover, striking variation in the content of alkenes across samples was found, with most genotypes showing an increase in alkene accumulation as leaves expanded and a few genotypes failing to accumulate this compound class in their wax. When grown in a common garden, alkene-accumulating genotypes exhibit larger weight and diameter and greater resistance to the leafspot, suggesting that this trait confers an important advantage to *P. trichocarpa*. Comparison of the transcriptomes of contrasting genotypes pointed at *PtKCS1* as the candidate gene responsible for the differences. The expression of *PtKCS1* was found to be associated with the accumulation of very long chain alkenes in leaf cuticular waxes. When expressed in yeast, PtKCS1 showed substrate preference toward monounsaturated VLCFAs, consistent with a proposed pathway where these VLCFAs are decarboxylated to yield alkenes. Interestingly, *PtKCS1* is part of a tandem duplicated cluster of eight KCS-encoding genes on chromosome (Chr) 10, which share a high degree of homology (24). Unlike *PtKCS1*, expression of the closely related *PtKCS2* in yeast (*S. cerevisiae*) showed higher relative preference toward the elongation of saturated VLCFAs. Notably, an evolutionary hallmark of specialized metabolism is the duplication of biosynthetic genes, after which sub-functionalization or neofunctionalization can result in diversified gene functions (39).

Although the activity of several KCSs has been previously characterized, little is known about how substrate specificity evolves immediately following gene duplication and which amino acid substitutions drive these changes. To address these questions, we examined the Chr 10 cluster of KCSs from *P. trichocarpa*. First, we explored the evolutionary history of this gene cluster, revealing ancient and recent duplication events. Then, elongation assays of five paralogs (PtKCS1, 2, 4, 8, 9) were performed to compare substrate preferences toward saturated or monounsaturated VLCFAs of varying lengths. Based on these results, chimeric proteins and site-directed mutagenesis were used to identify two key regions (helix-4 and residue 277) affecting substrate specificity in this clade. Furthermore, these regions were confirmed to alter substrate specificity in a more distantly related KCS, PtCER6, and affect the synergistic relationship between PtCER6 and a CER2 (PtCER2-like5). We used these findings to rationally engineer longer VLCFA products by substituting smaller amino acids at position 277. Finally, we showed that the paralogs of the gene cluster exhibited differential inhibition by K3 herbicides despite their sequence similarities. Our results shed light on the molecular basis of KCS substrate specificity, highlighting the plasticity and epistatic nature of substrate binding in KCSs.

Results

Evolution of a KCS gene cluster in P. trichocarpa

In *P. trichocarpa* (v3.0), Chr 10 harbors a tandem duplicated gene cluster encoding eight KCS ORFs within a 215 kb region (Fig. 1A). To examine the origin of the gene cluster and relationship between the paralogous genes, we first conducted a phylogenetic analysis of all KCSs from *P. trichocarpa*. This analysis identified another closely related KCS on Chr 8 (Fig. 1, A and B). Since Chr 8 and 10 arose from a whole genome duplication (WGD) event approximately 60 million years ago, the most parsimonious explanation is that following WGD, the copy on Chr 10 expanded by tandem duplication (40). The phylogenetic tree indicates that the PtKCS1 clade can be further divided into two subclades based on sequence similarity: subclade A (*PtKCS1, 3, 5, 7, 9, 11*) and subclade B (*PtKCS2, 4, and 8*). Furthermore, subclade B genes are more similar to the copy on Chr 8 than to adjacent subclade A genes on Chr 10 in terms of coding sequence, promoter sequence, and expression profiles, indicating that the subclade divergence preceded the WGD event (Figs. 1B, S1, and S2). Therefore, phylogenetic evidence supports a model in which tandem duplication of an ancestral KCS, followed by WGD, gene loss on Chr 8, and additional independent tandem duplications on Chr 10 have led to the current architecture of the gene cluster in *P. trichocarpa* (Fig. 1C). In addition to multiple KCS copies, this region has a high density of repetitive elements (Fig. 1D). Notably, we found that the gene cluster is disproportionally rich in Helitrons (DNA-transposons) relative to adjacent regions on Chr 10 (Fig. S3). Additional evidence is
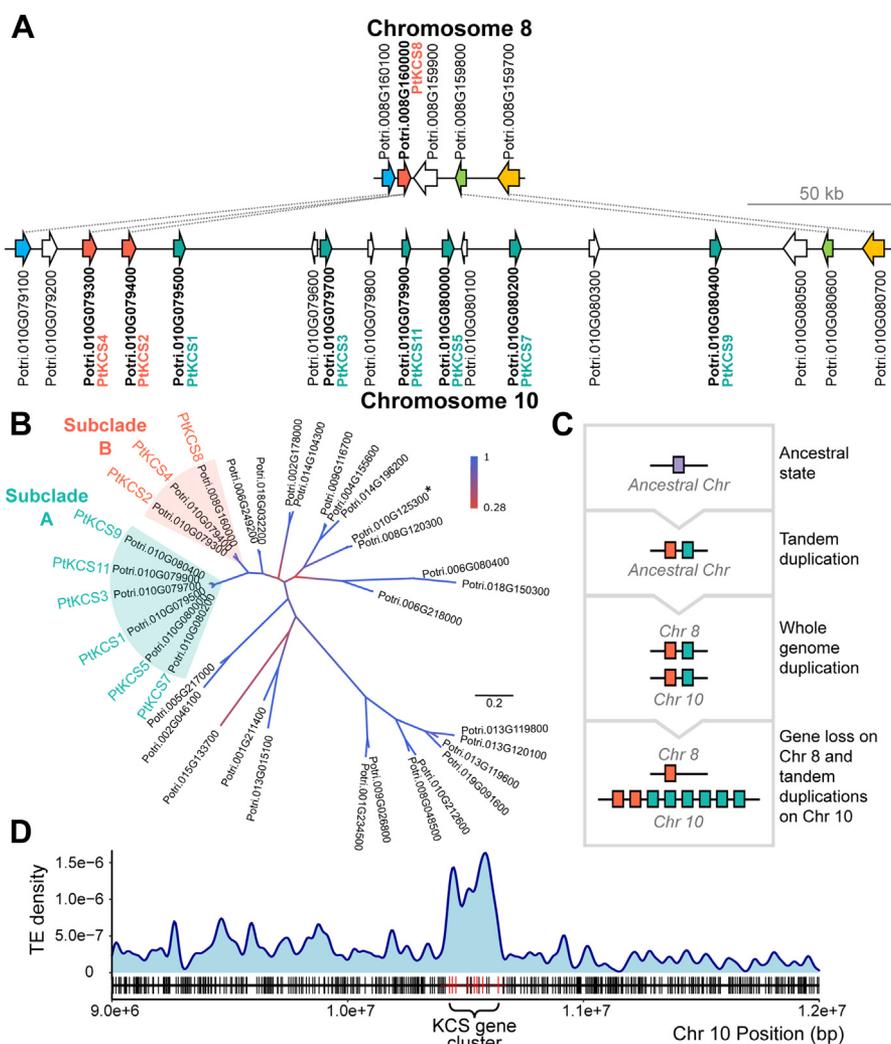
Figure 1. *PtKCS1* gene cluster in *P. trichocarpa*. *A*, diagram of the locus on Chr 8 and 10 containing *PtKCS1* clade members (bolded). *B*, maximum-likelihood phylogenetic tree of all KCS protein sequences in *P. trichocarpa*, excluding sequences shorter than 200 amino acids. Bootstrap values are represented by line color. Scale bar for branch length represents 0.2 amino acid substitutions per site. The asterisk denotes the gene corresponding to *PtCER6*. *C*, proposed evolutionary model of the PtKCS1 gene cluster from an ancestral KCS. *D*, kernel density of annotated transposable elements across a 3 Mb region spanning Chr 10: position 9,000,000...1,2000,000, on the *P. trichocarpa* v3.0 genome. A gene map is shown beneath the graph, with *red bars* representing KCS-encoding genes. Chr, chromosome.

needed to distinguish whether the high density of Helitrons is due to active transposition, by-products of uneven crossovers during cell division, or a combination of both events. Nevertheless, these observations suggest that this locus is prone to retaining repetitive elements.

As a result of the independent duplication events, members of the PtKCS1 clade have varying degrees of relatedness, sharing 74% to 99% amino acid identity between the nine copies (Table S1). Previous work has demonstrated that PtKCS1 has a higher preference toward unsaturated VLCFAs relative to PtKCS2, evidence of functional divergence in this clade (24). Thus, we hypothesized that other PtKCS1 clade members could serve as a model to understand the evolution of KCS substrate specificity on different timescales. Additionally, their high sequence similarity could help narrow down the molecular determinants of KCS substrate specificity.

PtKCS1 clade members have divergent VLCFA substrate specificities

KCS enzymes specialize in the substrates they can use for the condensing reaction. Hence, the VLCFA product profile observed in yeast will be reflective of the differences in substrates of the heterologously expressed KCS. To investigate the extent of substrate specificity divergence in the PtKCS1 clade, we chose to compare the elongation activity of five members: PtKCS1, 2, 4, 8, 9 (Fig. S4). Elongation activity was assayed using heterologous expression in yeast followed by fatty acid methyl ester (FAME) analysis (37). Yeast accumulates endogenous VLCFAs (predominantly C26 and minor levels of C20 and C20:1) as a result of the native ELO2 and ELO3 enzymes that share no homology to KCSs (Fig. 2*A*) (12, 41). However, KCS activity can be detected by the increase in accumulated VLCFAs relative to the empty vector strain. In yeast, the five enzymes synthesized varying amounts of saturated and
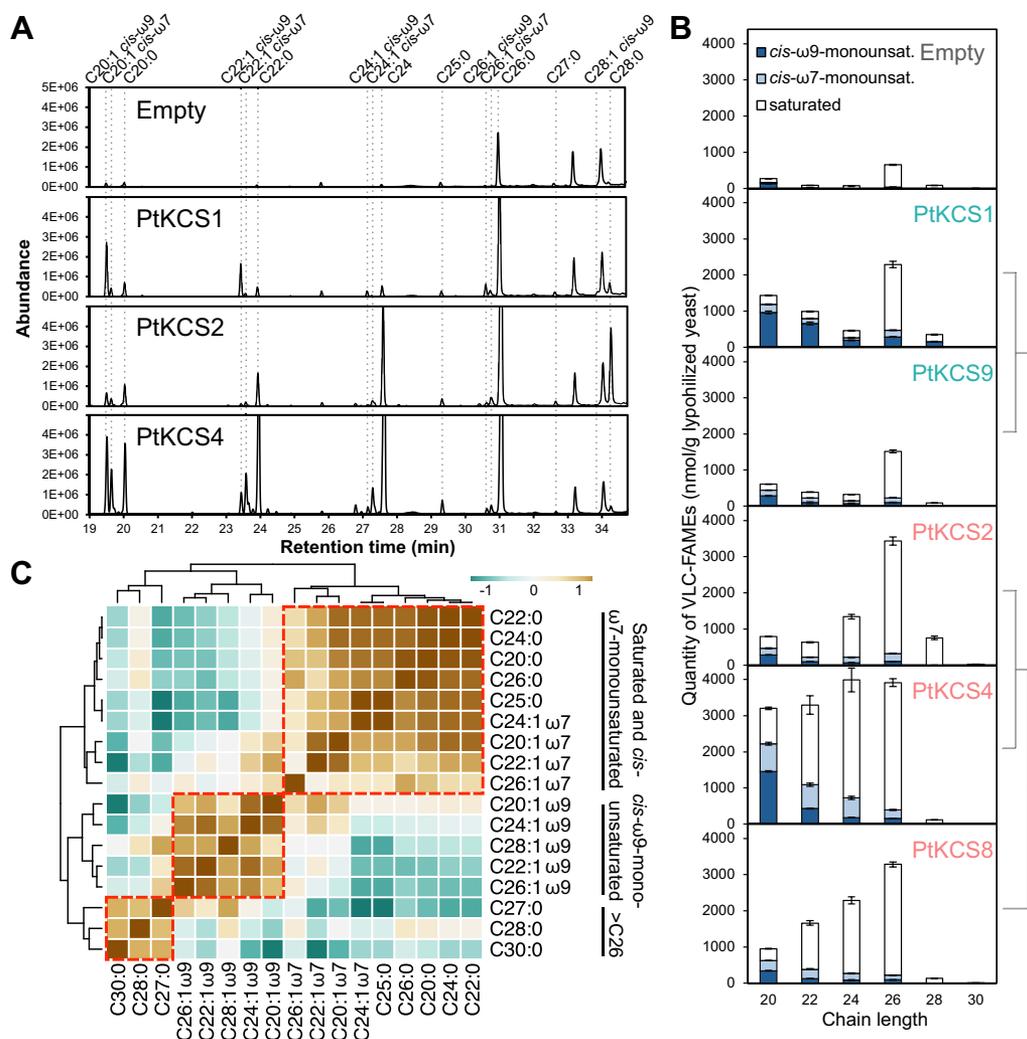
Figure 2. Substrate specificity of poplar KCSs from the PtKCS1 clade. *A*, representative GC-MS total ion chromatograms of VLC-FAMEs extracted from yeast expressing the empty vector, PtKCS1, PtKCS2, and PtKCS4. *B*, quantities of even chained VLC-FAMEs in nmol per gram of lyophilized yeast expressing the empty vector, PtKCS1, 9, 2, 4, and 8. Data represent means of five biological replicates ± SEM. The cladogram indicates the phylogenetic relationship between the five proteins. *C*, metabolite-metabolite correlation matrix of the elongation products, using nonparametric Spearman's ranking. VLC-FAME, very long chain fatty acid methyl ester.

monounsaturated VLCFAs ranging from C20-C30 in length (Fig. 2, *A* and *B*). Using authentic standards, we identified two series of monounsaturated products: one corresponding to the *cis*-ω9 double bond configuration and the other to the *cis*-ω7 configuration (Fig. S5). The substrates and products of the PtKCS1 clade members can be described in terms of three properties: carbon length, unsaturation, and position of the double bond. Subclade A members (PtKCS1 and 9) generally showed higher accumulation of shorter products relative to PtKCS2 and 8 from subclade B (Fig. 2*B*). In terms of unsaturation, PtKCS2 and 8 showed preference toward the production of saturated VLCFAs and *cis*-ω7-monounsaturated substrates. Conversely, subclade A members (PtKCS1 and 9) showed higher production of *cis*-ω9-monounsaturated fatty acids. The only exception was PtKCS4, which accumulated both shorter and *cis*-ω9-monounsaturated VLCFAs (like subclade A) but also high levels of saturated products (like subclade B). Remarkably, due to their high sequence similarity (99%) and homology, PtKCS2 and PtKCS4 only differ at five

amino acids. Four of these residues account for higher activity toward monounsaturated VLCFAs, demonstrating that KCS substrate specificity can evolve rapidly following gene duplication.

To better understand KCS activity toward the different classes of VLCFAs, we generated a metabolite-metabolite correlation map, showing that the products of these KCSs segregated into three main clusters: (1) saturated and *cis*-ω7 monounsaturated, (2) *cis*-ω9 monounsaturated, and (3) saturated VLCFAs with a chain length > C26 (Fig. 2*C*). These results suggest that the substrate specificity of these KCSs is predominantly determined by the presence and position of the double bond as opposed to chain length, apart from VLCFAs longer than C26. Furthermore, the tight correlation of *cis*-ω7 monounsaturated products with saturated products likely results from the double bond being closer to the terminal methyl end of the VLCFA; this is opposed to *cis*-ω9 VLCFAs which contain a kink closer to the middle of in the hydrocarbon tail and thus may be more sterically constrained.

Homology modeling and ligand docking of PtKCS1, 2, and 4 show differences in the putative substrate-binding site

To gain insight into the putative substrate-binding site of the KCS, we built a homology model of the cytosolic portion of PtKCS1 using Alphafold2 with templates from the closely related Type III polyketide synthases (PKSs; Fig. 3A, Tables S2 and S3) (42, 43). The predicted PtKCS1 structure shows the characteristic five layer αβαβα fold that is conserved between Type III PKSs, 3-ketoacyl-ACP synthases, and thiolases (44–46). To identify putative substrate-binding domains, we docked docosanoic acid (a C22:0 fatty acyl chain) and malonyl-CoA (the extender substrate) onto the homology model (Fig. 3A). Malonyl-CoA and docosanoic acid bind to distinct regions of the protein but converge on the catalytic triad as expected for the

condensation reaction. Though a free fatty acid is not generated in the proposed KCS reaction mechanism, docosanoic acid represents the fatty acyl chain that is covalently attached to the active site cysteine (47). The fatty acid docks to a large hydrophobic tunnel that extends from the active site and is only accessible through the core of the protein. Comparison with models of PtKCS2 and 4 shows that the binding tunnel in PtKCS1 is kinked compared to PtKCS2 and 4, which could explain the relative preference of PtKCS1 for *cis*-monounsaturated compared to saturated substrates (Fig. 3B). The binding tunnels of PtKCS2 and 4 are similarly shaped but differ near the end of the tunnel around helix-4 (residues 131–141).

The amino acid sequences of PtKCS1, 2, and 4 show distinct regions with varying degrees of conservation (Fig. 3C). For

Figure 3. Structural and sequence comparisons of PtKCS1, 2, and 4. *A*, homology model of the cytosolic portion of PtKCS1 docked with docosanoic acid (*black*) and malonyl-CoA (*magenta*), with the catalytic triad shown as *yellow sticks*. Alpha helix 4 and position 277 are highlighted by an *orange dotted line*. *B*, comparison of the binding tunnel in PtKCS1, 2, and 4 homology models, shown by surface representation and colored according to hydrophobicity. Docosanoic acid (*black*) is docked to each model with residues of the catalytic triad shown as *yellow sticks*. *C*, sequence alignment of PtKCS1, 2, and 4 including predicted secondary structures from PtKCS1. The five amino acid differences between PtKCS2 and 4 are highlighted in *turquoise*, and the three catalytic residues are highlighted in *yellow*. Predicted transmembrane helices are labeled (TM).

example, the C-terminal half of the proteins is more conserved compared to the N terminus (Fig. S6). When plotted on the homology model of PtKCS1, most substitutions line the putative VLCFA binding tunnel by helices-4, 6, 7, and 11 (Fig. S7). Conversely, helix-5 and 12, which are positioned by the CoA moiety, contain few substitutions between paralogs. As recognition of CoA thioesters is a conserved feature across KCSs and even Type III PKSs, mutations are more likely to be deleterious in this region.

Site-directed mutagenesis of two residues switches the substrate preference of PtKCS4 to 2

PtKCS2 and 4 have divergent substrate preferences despite only having four amino acid differences in the cytosolic domain and one in the transmembrane domain (Fig. 3C). These substitutions are sufficient to confer relative preferential elongation of shorter, monounsaturated VLCFAs in PtKCS4 and relative preferential elongation of longer saturated VLCFAs in PtKCS2, especially C28 (Fig. 2A). To determine the effect on substrate specificity, we used site-directed mutagenesis to convert each of the four different residues on PtKCS4 to the corresponding residue from PtKCS2 (Fig. 4A). Mutagenesis of R283C or T137K reduced the production of monounsaturated VLCFAs, while mutagenesis of T137K on helix-4 additionally increased production of saturated C28 VLCFA (Fig. 4, A and B). To test whether the combined effect of two substitutions could convert the overall product profile to that of PtKCS2, we generated a double mutant, $\text{PtKCS4}_{\text{T137K, R283C}}$. By changing these two residues, the overall substrate preference of PtKCS4 was almost completely converted to that of PtKCS2 (Fig. 4 and Table S4). Notably, there was a synergistic increase in C28 in the double mutant compared to either mutant alone, showing these positions have epistatic effects on substrate specificity.

Plotting the amino acid differences between PtKCS2 and 4 on the three-dimensional models shows that they are clustered around the binding tunnel of the fatty acyl ligand, with residues 137 and 283 positioned on opposite sides of the ligand (Fig. 4B). In PtKCS4, both the polar threonine (at residue 137) and arginine (at residue 283) point into the binding tunnel to effectively constrict and decrease its hydrophobicity, which could explain why PtKCS4 is restricted to synthesizing shorter VLCFAs compared to PtKCS2.

Identifying the molecular determinants of substrate specificity between PtKCS1 and 2

Compared to PtKCS2 and 4, PtKCS1 has a higher relative preference toward *cis*-monounsaturated compared to saturated VLCFAs (Fig. 2). To test whether the two sites we identified in PtKCS2 and 4 are also necessary for substrate specificity between PtKCS1 and 2, we generated site-directed mutants at these positions, corresponding to residues 131 and 277 in PtKCS1 (Fig. 5, Group I). The activity toward *cis*-monounsaturated fatty acids was reduced in the single and double mutants, showing that these sites also affect substrate specificity between PtKCS1 and 2 (Fig. 5 and Table S5). The combined effect of the double mutant on accumulation of *cis*-monounsaturated VLCFAs was greater than either single mutant alone, further indicating that these two positions have synergistic effects. Though the $\text{PtKCS1}_{\text{W277C}}$ mutant only had a minor effect on monounsaturated VLCFAs, there was a large increase in C28, characteristic of PtKCS2 activity. In contrast, mutating residue 275 ($\text{PtKCS1}_{\text{R275L}}$) resulted in a product profile identical to PtKCS1, showing that not all divergent positions near the binding tunnel affect substrate specificity.

Residues 131 and 277 affected substrate preference but were not sufficient to entirely convert the substrate specificity of
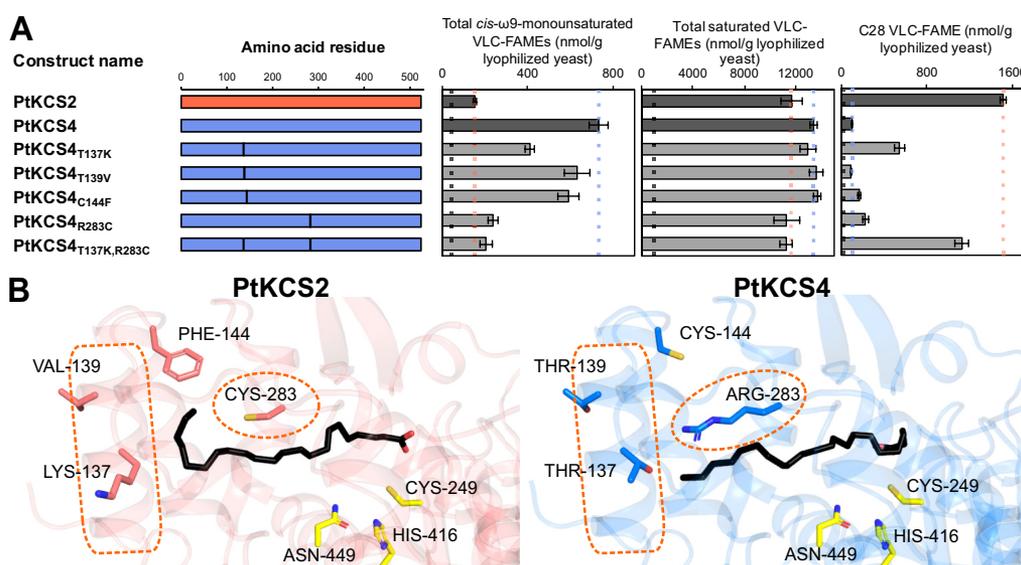
Figure 4. Substrate specificity of site-directed mutants between PtKCS4 and PtKCS2. *A*, diagrams of the site-directed mutants and respective production of *cis*-ω9-monounsaturated, saturated, and C28 VLC-FAMEs expressed in nmol per gram of lyophilized yeast. Reference lines indicate levels of VLC-FAMEs in the empty vector (*gray*), PtKCS2 (*salmon*), and PtKCS4 (*blue*). Data represent means of three biological replicates ± SEM. *B*, close-up view of the region in PtKCS2 and 4 where docosanoic acid (*black*) is docked. Amino acid differences between PtKCS2 and 4 are labeled and shown in stick representation. The residues of the catalytic triad are shown as *yellow sticks*. Alpha helix 4 and position 283 are highlighted by an *orange dotted line*. VLC-FAME, very long chain fatty acid methyl ester.
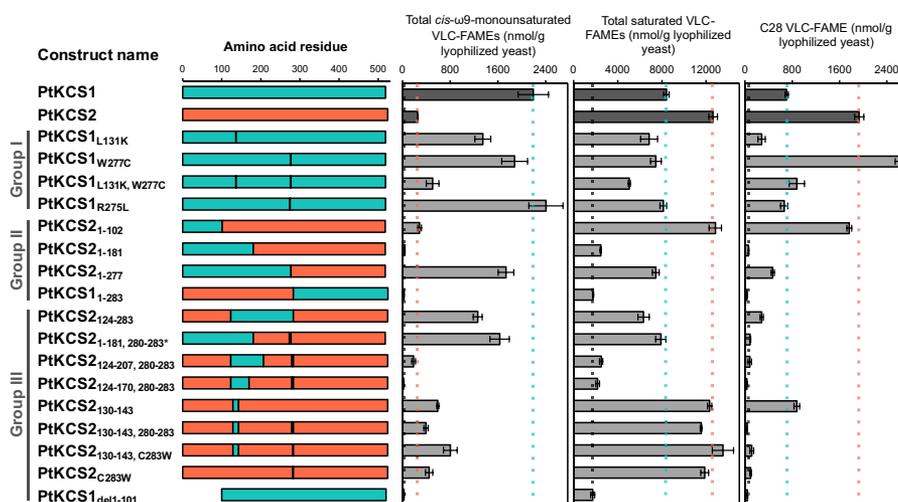
Figure 5. Substrate specificity of fusion proteins and site-directed mutants between PtKCS1 and PtKCS2. Diagrams of the fusion proteins and site-directed mutants and respective production of *cis*-ω9-monounsaturated, saturated, and C28 VLC-FAMEs expressed in nmol per gram of lyophilized yeast. Reference lines indicate levels of VLC-FAMEs in the empty vector (*gray*), PtKCS1 (*teal*), and PtKCS2 (*salmon*). The asterisk on PtKCS2$_{1-181, 280-283}$ indicates that the actual position of the second swapped region is 274 to 277, but this corresponds to position 280 to 283 on the rest of the Group III fusions. Data represent means of three biological replicates ± SEM. VLC-FAME, very long chain fatty acid methyl ester.

PtKCS1 to PtKCS2. This is expected since PtKCS2 and 4 share 99% identity, whereas PtKCS1 and 2 only share 74% identity. Therefore, we looked to identify the molecular basis of substrate specificity between PtKCS1 and 2 using a fusion protein approach guided by natural polymorphisms between paralogs, hydrophobicity plots, and homology models (Fig. S8). We rationally designed chimeric proteins by fusing varying lengths of the PtKCS1 N terminus to the PtKCS2 C terminus (Fig. 3A; Group II). PtKCS2$_{1-102}$, which only contains the transmembrane domain from PtKCS1, showed identical activity to PtKCS2, indicating that this region does not affect substrate specificity, consistent with previous reports that substrates are delivered cytosolically and not through the membrane (37) (Fig. 5). The first 129 residues swapped between PtKCS1 and 2 were also not sufficient to modify substrate specificity (as shown by PtKCS2$_{1-129}$). Conversely, PtKCS2$_{1-277}$ showed almost identical activity to PtKCS1, demonstrating that the first 277 residues from PtKCS1 are sufficient for converting the substrate specificity of PtKCS2 to PtKCS1. We generated a reciprocal fusion with the N terminus of PtKCS2 fused to the C terminus of PtKCS1 (PtKCS1$_{1-283}$), but this abolished the elongation activity in yeast, which could be due to unfavorable interactions in the tertiary structure.

To further narrow down the specific regions required for PtKCS1 substrate specificity, we generated chimeras to swap smaller regions of PtKCS2 with the corresponding region from PtKCS1 (Fig. 5A; Group III). PtKCS2$_{123-283}$ was still sufficient to confer PtKCS1 substrate specificity though there was a minor loss of activity. Surprisingly, taking the fusion with reduced activity (PtKCS2$_{1-181}$) and replacing residues 274 to 277 by helix-11 (generating PtKCS2$_{1-181, 280-283}$) was sufficient to convert the substrate specificity to PtKCS1. Narrowing the region down to only 25 amino acid differences between PtKCS1 and PtKCS2 in the PtKCS2$_{123-207, 280-283}$ caused a large reduction in activity, but this fusion was still sufficient to increase the relative preference toward monounsaturated substrates. This fusion contains

substitutions that span helices-4, 5, 6, and 11. Though the transmembrane domain does not affect substrate specificity, it appears to be essential for KCS activity in yeast, which was further confirmed by a transmembrane domain deletion (PtKCS1$_{del1-101}$) that showed no elongation activity.

Interestingly, only nine substitutions around helix-4 (PtKCS2$_{129-143}$) were sufficient to increase the production of monounsaturated fatty acids (predominantly C20:1 and C22:1) without affecting activity toward saturated VLCFAs. Therefore, helix-4 is important in conferring activity toward monounsaturated substrates, but additional regions by helices-5, 6, and 11 are important in discriminating against saturated products. These combined analyses identify the key regions (in the N terminus and at position 277) that contribute to substrate preference between PtKCS1 and 2.

The N-terminal region controls substrate specificity in a more distantly related KCS

The results aforementioned show that helix-4 and site 277 are key determinants of substrate specificity in the PtKCS1 clade. To test whether these results can be extrapolated to more distantly related KCSs, we cloned PtCER6 (Potri.010G125300), sharing 59% amino acid identity with PtKCS1, for subsequent fusion and site-directed mutagenesis studies (Fig. 1B). Based on sequence similarity, PtCER6 is the putative poplar ortholog of AtCER6, a KCS from *Arabidopsis* which produces mainly C28 VLCFAs in yeast (28). Functional characterization of PtCER6 showed similar accumulation of predominantly C28 VLCFAs, markedly different from the broad product profile of PtKCS1 (Fig. 6A and Table S6). Reciprocal swaps of the N-terminal region between PtKCS1 and PtCER6 (PtCER6$_{1-279}$ and PtKCS1$_{1-260}$) resulted in empty control levels of VLCFA accumulation, likely due to the disruption of epistatic interactions on the protein tertiary structure as observed in the PtKCS1$_{1-283}$ mutant between PtKCS1 and 2 (Figs. 6, A–C and 5). However, swapping helix-4
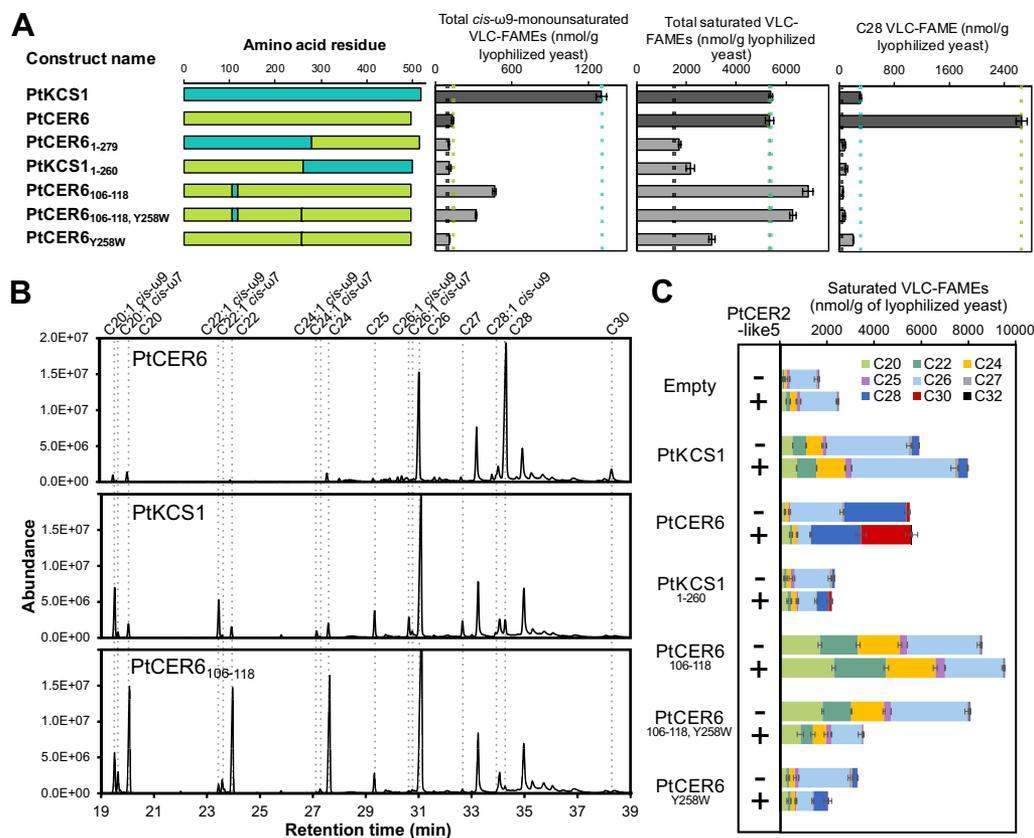
Figure 6. Effect of helix-4 and site 277 on the substrate specificity of PtCER6. *A*, diagrams of chimeras and site-directed mutants between PtKCS1 and PtCER6 and their respective production of *cis*-ω9-monounsaturated, saturated, and C28 VLC-FAMEs expressed in nmol per gram of lyophilized yeast. Data represent means of three biological replicates ± SEM. *B*, representative total ion chromatograms of PtCER6, PtKCS1, and PtCER6106-118. *C*, chain length distribution of saturated VLC-FAMEs produced by chimeras expressed with or without PtCER2-like5. Data represent means of three biological replicates ± SEM. VLC-FAME, very long chain fatty acid methyl ester.

with the homologous region from PtKCS1 (PtCER6106-118) was sufficient to increase production of monounsaturated fatty acids by 3.4-fold and levels of C20-C24 VLCFAs at the expense of C28 VLCFAs, resembling the product profile observed in the PtKCS1 gene cluster (Fig. 6B). A single mutation in PtCER6 (PtCER6Y258W) also caused a shift toward shorter VLCFAs characteristic of PtKCS1 but was not sufficient to increase specificity toward unsaturated products. Our results show that helix-4 appears to be a major determinant for specificity across KCSs. Furthermore, this region could be changed without deleterious effects on activity, which may be explained by the placement of helix-4 near the surface of the homology model and relative isolation from other secondary structures (Fig. 3A).

The elongation activity of certain KCSs is modulated by a binding partner, CER2, through an unknown biochemical mechanism (29). In *Arabidopsis*, the current model proposes that AtCER2/CER2-like proteins can extend the chain length of AtCER6 products beyond C28. To test whether PtCER2s could extend the chain length specificity of different poplar KCSs and identify which regions on the KCS are important for this synergistic interaction, we coexpressed a previously characterized PtCER2-like5 with different KCSs and chimeras (Fig. 6A) (33). When PtCER2-like5 was coexpressed with PtCER6, the product specificity was shifted toward C30 VLCFAs (Fig. 6C and Table S6). Coexpression with PtKCS1 did not shift the products

toward longer VLCFAs. Remarkably, the product specificity of PtKCS11-260 and PtCER6Y258W was partly modified by PtCER2-like5 toward longer VLCFAs, indicating that the first N-terminal 1 to 260 residues of PtCER6 are sufficient for the modulatory effects of PtCER2-like5 on KCS substrate specificity. Furthermore, constructs containing helix-4 from PtKCS1 (PtCER6106-118 and PtCER6106-118, Y258W) abolished the effect of PtCER2-like5 on extending the length of PtCER6 products, indicating this structure is necessary for the modulatory effect of PtCER2like-5 on PtCER6. Interestingly, coexpression of PtCER2-like5 increased the total production of VLCFAs in PtKCS1, PtCER6106-118, and the empty vector control, an effect which was reversed in PtCER6 chimeras containing the Y258W mutation. Although the increase could be due to interactions of PtCER2-like5 with the endogenous ELO enzymes, these results suggest an additional role of CER2s in regulating VLCFA biosynthesis that is independent of the role in extending elongation beyond C28 (Fig. 6C).

Engineering longer VLCFAs by substituting smaller amino acids at position 277

Since position 277 (position 283 in PtKCS2 and 4) appears to have an important effect on the length of KCS products, we tested whether products could be rationally engineered toward

longer VLCFAs by substituting smaller amino acids at this position. We mutated the amino acids at this position in PtKCS1 (PtKCS1$_{W277V}$) and PtKCS4 (PtKCS4$_{R283G}$). While both mutants had a similar substrate specificity to the WTs, the levels of C28 increased at the expense of shorter VLCFAs, suggesting that the residue at 277 can act as a molecular ruler to regulate product length (Figs. 7*A* and S9).

Next, we tested whether these mutants had activity toward VLCFAs with additional functional groups. The fatty acids in yeast that are available to KCS as substrates are saturated and monounsaturated fatty acids (predominantly C18:0 and C18:1Δ9), but KCSs could have activity toward other substrates, such as polyunsaturated or hydroxylated fatty acids. To test this, yeast expressing KCSs were exogenously fed with linoleic, linolenic, and ricinoleic acid, corresponding to ω-6, ω-3, and hydroxylated fatty acids, respectively (Fig. 7*B*). Most KCSs, including the PtKCS1$_{W277V}$ mutant, had little to no activity toward these substrates beyond the background activity of the empty control (Fig. 7*C*). Remarkably, the PtKCS4$_{R283G}$ mutant could uniquely elongate diunsaturated fatty acids up to C26:2 and triunsaturated fatty acids up to C24:3, corresponding to four and three cycles of elongation, respectively. Together, these results show that KCSs can be engineered to tailor VLCFA biosynthesis and introduce novel substrate specificities.

Members of the PtKCS1 gene cluster are differentially inhibited by K3 herbicides

Acetamides, chloroacetamides, oxyacetamides, and tetrazolinones are part of the widely used class K3 herbicides that target KCSs to inhibit VLCFA biosynthesis in plants (48, 49). Since the duplicated KCSs exhibited divergent substrate preferences, we wanted to test whether they could also be differentially inhibited by K3 herbicides, which has been shown for distantly related KCSs from *Arabidopsis* (38). Since endogenous yeast elongases are not affected by K3 herbicides, we tested the effect of four inhibitors (alachlor, anilofos, fentrazamide, and flufenacet) on the elongase activity of PtKCS1, 2, and 4 using the yeast system (38). The substrate preferences of PtKCS1, 2, and 4 cover the major differences in activity observed in our studies, which is why they were chosen for the herbicide assay. The effect of inhibition varied between the three paralogs (Fig. 8*A*). Flufenacet and fentrazamide had the greatest effect on reducing VLCFA elongation in all three KCSs, whereas alachlor and anilofos slightly reduced activity. PtKCS1 was the most sensitive to inhibition, although there were also minor differences between PtKCS2 and 4. Since PtKCS2 and 4 only differ by five amino acids, these results imply that at least one of these residues is involved in conferring sensitivity to alachlor and anilofos.

To assess the inhibitor sensitivity of the different KCSs, we assayed PtKCS1, 2, and 4 in the presence of increasing concentrations of fentrazamide. PtKCS2 and 4 showed similar dose-response curves, whereas PtKCS1 was more sensitive to fentrazamide at lower concentrations (Fig. 8*B*). Since PtKCS1 and PtKCS2 have a different sensitivity to fentrazamide, we wanted to test whether KCS inhibition could be uncoupled from substrate specificity. Therefore, we repeated the experiment with PtKCS2$_{1-277}$, a chimera between PtKCS1 and PtKCS2. The dose-response curve of PtKCS2$_{1-277}$ was more like PtKCS2, despite having nearly identical substrate specificity to PtKCS1 (Figs. 8*C* and 5). This shows that it is possible to uncouple inhibitor sensitivity from substrate specificity. Comparing PtKCS1, PtCER6, and PtCER6$_{106-118}$ also showed differential inhibitor sensitivity of each enzyme to fentrazamide, although PtCER6$_{106-118}$ was more similar to PtCER6 (Fig. S10).

Discussion

Gene duplication plays an important role in the evolution of new gene functions, particularly in the diversification of specialized metabolites (39). In *P. trichocarpa*, PtKCS1 is part of a tandem duplicated gene cluster on Chr 10 that was recently identified as a key enzyme in the biosynthesis of alkenes for cuticular waxes (24). In this study, we used the PtKCS1 gene cluster as a model to study the molecular mechanisms driving evolution of substrate specificity. We first examined the origin of the gene cluster to uncover a history of tandem duplication events, WGD, and gene loss (Fig. 1*C*).

Using this gene cluster as a model for KCS evolution, we compared the substrate preference of five paralogs that share varying degrees of relatedness. The paralogs were promiscuous

Figure 7. Elongation activities of rationally designed mutants PtKCS1$_{W277V}$ and PtKCS4$_{R283G}$. *A*, chain length distribution of saturated VLCFAs produced by mutants compared to WT enzymes. Data represent means of three biological replicates ± SEM. *B*, structures of exogenously fed fatty acids. *C*, detected elongation products after feeding substrates to empty vector control, PtKCS1 and 4 WT enzymes, and mutants. N.D.= not detected. VLCFA, very long chain fatty acid.

Figure 7A — Quantity of saturated VLC-FAMEs (nmol/g of lyophilized yeast). Categories: C20, C22, C24, C25, C26, C27, C28, C30. Bars for PtKCS1, PtKCS1 W277V, PtKCS4, PtKCS4 R283G.

Figure 7B — Structures: Linoleic Acid 18:2 *cis* 9,12; Linolenic Acid 18:3 *cis* 9,12,15; Ricinoleic Acid (R)-12-OH- 18:1 *cis* 9.

Figure 7C

Substrate	Detected elongation products				
	Empty	PtKCS1	PtKCS1 W277V	PtKCS4	PtKCS4 R283G
18:2 *cis* **9,12**	20:2	20:2	20:2	20:2, 22:2	20:2, 22:2, 24:2, 26:2
18:3 *cis* **9,12,15**	20:3	20:3	20:3	20:3	20:3, 22:3, 24:3
(R)-12-OH-18:1 *cis* **9**	N.D.	N.D.	N.D.	N.D.	N.D.

Figure 8. Differential inhibition of PtKCS1 clade members with K3 herbicides. *A*, proportion of total VLC-FAMEs in the respective herbicide treatments (2 μM) relative to the uninhibited and empty vector controls. Proportion represents the mean of 3 biological replicates ± SE. The structures of the four herbicides, alachlor, anilofos, flufenacet, and fentrazamide are shown above. *B*, log-inhibitor response curves of PtKCS1, 2, and 4 in the presence of increasing fentrazamide concentration. *C*, log-inhibitor response curves of PtKCS1, 2, PtKCS2$_{1-277}$ in the presence of increasing fentrazamide concentration. VLC-FAME, very long chain fatty acid methyl ester.

to products between C20-C30 in length, showing distinct substrate preferences based on length, unsaturation, and position of the double bond. In general, subclade A members had a higher relative preference for monounsaturated substrates compared to subclade B, apart from PtKCS4. Previous work on the closest orthologs to the PtKCS1 clade in *A. thaliana* (AtKCS2/20) and *Helianthus annus* (HaKCS1) found a broad product specificity characteristic of the poplar clade, accumulating predominantly saturated C20-C26 VLCFAs (38, 50, 51). These studies provide insight into the ancestral substrate preference of these genes, corroborating the promiscuity of this clade toward shorter VLCFAs. In poplar, the gene cluster provides an example of specialization after gene duplication, where subclade A enzymes have increased their relative preference toward unsaturated VLCFAs compared to subclade B. This specialization toward unsaturated fatty acids has occurred again in subclade B enzyme PtKCS4 after diverging from PtKCS2. On the other hand, PtKCS2 has specialized toward

longer products, up to C30 VLCFAs (Fig. 2*B*). Members of the gene cluster appear to have undergone different variations of subfunctionalization and neofunctionalization, specializing in the production of shorter *versus* longer, saturated *versus* monounsaturated VLCFAs.

Despite the extensive literature on characterizing different KCSs using plant mutants and heterologous expression, little is known about the molecular basis of KCS substrate specificity. Studies have been partly hindered by the recalcitrance of the *in vitro* assay, owing to both the limited availability of commercially available very long chain acyl-CoA substrates and insolubility of the membrane-bound KCSs. Furthermore, we found that the addition of a C-terminal His-tag to KCSs caused a significant reduction in activity (Fig. S11). As such, our results focus on the comparison of product profiles as opposed to enzyme kinetics.

Generating fusion proteins to identify the molecular determinants of substrate specificity revealed that the first 277 residues of the N-terminal domain were sufficient to switch the substrate specificity from PtKCS2 to PtKCS1 (Fig. 5). A prior genome survey of KCSs from 28 plant species showed higher diversity on the N-terminal region, suggesting that this region is under more relaxed purifying selection (19). The previous study on KCS substrate determinants compared orthologs of FAE1 from *A. thaliana* and *B. napus*, also finding that the N-terminal region (up to helix-4) was key (37). Using different KCSs (PtKCS1, 2, 4, and PtCER6), we narrowed down the substrate determining regions in the N-terminal domain and identified an additional region at position 277 affecting substrate specificity. Site-directed mutants and fusion constructs showed that substitutions by helix-4 and at site 277 had major effects on activity toward VLCFAs based on unsaturation and length. Using this knowledge, the rational substitutions of smaller amino acids at site 277 were able to shift the products toward longer VLCFAs in PtKCS1 and 4, as well as introduce novel activities toward elongation of the linolenic acid, a triunsaturated ω-3 fatty acid (Fig. 7*C*).

It is worth noting that during the process of constructing the chimeras between PtKCS1 and PtKCS2, a few constructs failed to modify the VLCFA profile of yeast (Fig. 5). Notably, PtKCS1$_{1-283}$ lost activity on saturated and unsaturated fatty acids, whereas the reciprocal PtKCS2$_{1-277}$ was active. Similarly, PtKCS2$_{1-181}$, PtKCS2$_{124-207, 280-283}$ and PtKCS2$_{124-170, 280-283}$ were inactive, suggesting that there are key interactions between the region around 124 to 170 amino acids with the C-terminal half that are disrupted in these constructs. When this region was narrowed down to helix-4 (130–143 amino acids), constructs were active. This suggests that there are incompatible combinations between the two halves of PtKCS1 and PtKCS2 and that helix-4 is unusual in its modular property, that is, it can be replaced without destabilizing these interactions. Another interesting observation was that though helix-4 and residue 277/283 had clear effects on substrate specificity, combination mutants of these two regions for PtKCS1/PtKCS2 and PtKCS1/PtCER6 did not necessarily have an additive or synergistic effect (Figs. 5 and 6). These results again show that other regions on the protein are seemingly

epistatic and contribute to the structure of the substrate-binding tunnel. This makes sense considering the size of the binding site spans almost half of the entire protein (Fig. 3).

By modifying the more distantly related PtCER6 (59% amino acid identity to PtKCS1) with residues from PtKCS1, we further supported the role of helix-4 and site 277 in other KCSs. Furthermore, the coexpression of PtCER2-like5 with chimeras of PtCER6 and PtKCS1 showed that PtCER2-like5 requires the N-terminal region of PtCER6 for its modulatory effect on VLCFA elongation. Though these data likely suggest that PtCER6 interacts with PtCER2-like5 through the N-terminal domain (particularly helix-4), additional structural and interaction studies are required to test this. The interaction between CER2-likes with condensing enzymes is not fully understood; yet at this point, it seems that the ability of CER2-likes to shift the chain length profile is limited to AtCER6 and phylogenetically related KCSs (AtCER60, WLS4 in rice, and PtCER6 in poplar) (29, 38). Structural models for CER2-like show that the N-terminal 100 residues that differentiate the modulatory activity of Arabidopsis CER2-likes on AtCER6 are positioned on the surface of the protein (2). However, it remains unknown if this region physically interacts with the FAE complex, binds any of the substrates of KCS, or whether CER2-like physically interacts with other KCS whose elongation it does not modulate. Future studies should consider physical interaction assays to determine if the increase in total VLCFAs observed in PtKCS1, PtCER6$_{106-118}$, and the empty vector control depend on interactions of PtCER2-like5 with yeast ELO or poplar KCSs.

KCSs have been shown to be differentially inhibited by K3 herbicides (38). Interestingly, the tandem duplicates also showed patterns of differential inhibition when treated with K3 herbicides, despite their recent divergence. As with substrate specificity, KCS inhibition also seemed to be easily modified by few substitutions. Using a fusion between PtKCS1 and 2 (PtKCS2$_{1-283}$), we showed that substrate specificity could be uncoupled from inhibitor sensitivity to fentrazamide (Fig. 8C). Considering that the most closely related enzymes to KCSs (chalcone synthases) are covalently inhibited by K3 herbicides, it is likely that KCSs are also inactivated in this manner, and the properties of the active site pocket could change without affecting the terminal end of the large binding tunnel (52). Understanding the molecular basis of this inhibition could be useful for engineering transgenic crops that are resistant to certain herbicides or developing herbicides that target specific KCSs.

Current models suggest that the binding site of the KCS is flexible to accommodate for different substrates, which generates the broad product profile observed for certain KCSs (37). We propose that this "flexible model" can be refined to include a recognition domain for certain motifs on functionalized VLCFAs. For example, the carbon series of monounsaturated VLCFAs observed for PtKCS1 likely results from preferential recognition of the *cis*-ω9 motif (Fig. 2). Furthermore, fusions of PtKCS2 and PtCER6 containing helix-4 from PtKCS1 shows increased production of *cis*-ω9 products, suggesting that helix-4 confers partial recognition of this motif. Feeding 18:2 and 18:3 fatty acids to yeast expressing

PtKCS4$_{R283G}$ also generated a carbon series of diunsaturated or triunsaturated VLCFAs, respectively. Therefore, a flexible region between the active site and the recognition domain could explain how functionalized VLCFAs are extended to varying lengths (Fig. S12). Another explanation we propose here is the "fixed model" whereby the flexible substrates of varying lengths conform to fit inside a fixed binding tunnel containing the recognition domain. Supporting this model is the shape of the cavity in the PtKCS homology models, showing a large cavity at the active site that could potentially accommodate folding of the hydrocarbon chain (Fig. 3B). Solved structures with bound substrates would help distinguish between these two models and further elucidate the mechanisms of KCS substrate specificity.

In general, KCS substrate preference was easily altered by perturbations to the substrate-binding domain. This is likely because substrate specificity of hydrocarbon chains is largely determined by the size and hydrophobicity of the binding tunnel unlike polar ligands, which are stabilized by specific ionic interactions (53). Substrate promiscuity is a key trait of enzymes involved in specialized metabolism (39). Using different mutants, we show that even a single substitution can drastically change KCSs substrate specificity. These minor changes, coupled with their promiscuity, demonstrate how KCSs might have rapidly evolved to accept new substrates. KCSs are closely related to Type III PKSs such as chalcone synthases, which also use diverse substrates. Unlike PKSs, however, the KCS has been largely overlooked in terms of their molecular mechanisms and biocatalytic potential. There is recent evidence that certain KCSs can condense elongation cycle intermediates in a PKS-like manner, suggesting they are more versatile than previously thought (23). Using directed evolution or rational approaches to alter the substrate determining regions of KCSs has the potential to engineer novel, noncanonical substrate specificities.

Here, we show that the evolution of KCSs can happen rapidly following gene duplication, resulting in subfunctionalization and neofunctionalization. Using the paralogs of a duplicated KCS gene cluster as the basis for mutational analyses, we identified major determinants of specificity toward unsaturation and chain length. The PtKCS1 gene cluster is part of the larger KCS gene family that has unique and diverse substrate specificities. This work lays the foundation toward understanding the molecular mechanisms behind their incredible diversity.

Experimental procedures

Materials

The INVSc1 yeast strain, PYESDEST52 vector, and PYES3/CT:CER2-like5 construct were generously shared by Shawn D. Mansfield at the University of British Columbia. The herbicides (alachlor, anilofos, fentrazamide, and flufenacet) were purchased from Sigma–Aldrich as PESTANAL analytical standards. External standards methyl heptadecanoate, methyl oleate, SUPELCO 37 Component FAME Mix, 11(E)-eicosenoic acid, 13(Z)-eicosenoic acid, and 14(Z)-eicosenoic acid

were purchased from Sigma–Aldrich, Cayman Chemical, Acros Organics, and Tokyo Chemical Industry at ≥95% purity. The internal standard (nonadecanoic acid) was purchased from Agilent. Fatty acids used for feeding were purchased from Cayman Chemical and Sigma–Aldrich at ≥98% purity.

RNA extraction and cloning

Total RNA from *P. trichocarpa* (accessions HALS30-6 and KLNE20-1) leaf tissue was extracted using TRIzol (Invitrogen), then used as a template for complementary DNA (cDNA) synthesis using the OneScript Plus cDNA Synthesis Kit (Applied Biological Materials). The following genes were used in this study: *PtKCS1*, Potri.010G079500; *PtKCS2*, Potri.010G079400; *PtKCS4*, Potri.010G079300; *PtKCS9*, Potri.010G080400; *PtKCS8*, Potri.008G160000; PtCER6, Potri.010G125300; PtCER2-like5, and Potri.005G052200. The KCSs were cloned from cDNA (apart from Potri.010G080400 which was gene synthesized by Bio Basic) for directional Gateway cloning into pENTR/D-TOPO (Invitrogen). All primers used in this study are listed in Table S7. The plasmids were transformed into *Escherichia coli* TOP10 competent cells, screened, and confirmed by Sanger sequencing. Constructs had the same amino acid sequence as the reference *P. trichocarpa* (v3.0) genome, except PtKCS1, which has a serine to proline mutation at position 125. The sequence-confirmed genes were transferred to the PYESDEST52 destination vector *via* the Gateway LR reaction and transformed into *E. coli* TOP10 cells. Vectors obtained from PCR-positive clones were transformed into INVSc1 yeast (Invitrogen) using the SC Easy Comp Kit (Invitrogen) and plated on Synthetic Complete minus Uracil (SC-Ura) media containing 2% (w/v) glucose. Individual transformants were selected and screened by PCR. Yeast containing the PYES3/CT:PtCER2-like5 construct were made competent and transformed with PYESDEST52 vectors as described previously. Colonies were selected on Synthetic Complete minus uracil and tryptophan (SC-Ura-Trp) and screened by PCR.

Chimeric proteins and site-directed mutagenesis

Recombinant protein fusions were generated by overlap extension PCR and cloned into pENTR D-TOPO as described previously (54). Site-directed mutants were generated using mutagenic primers and Phusion High-Fidelity DNA Polymerase (Invitrogen), followed by *DpnI* digestion. All recombinant constructs were confirmed by Sanger sequencing.

Yeast expression assays

Transgenic cultures were grown overnight at 30 °C on SC-Ura plates with 2% glucose. The cultures containing PYESDEST52 vectors were inoculated into SC-Ura liquid media containing 2% raffinose. The cultures containing both PYESDEST52 and PYES3/CT vectors were inoculated into SC-Ura-Trp liquid media containing 2% raffinose. The overnight cultures were grown shaking at 30 °C, 200 rpm to an A_{600} of ∼ 0.5, after which galactose was added to 2% final concentration to induce expression. After 4 days of growth, the liquid cultures (A_{600} ∼ 2.0) were harvested by centrifugation at 4500 rpm for 10 min. The cultures were washed with 1% (w/v) NaCl solution then stored at –80 °C prior to overnight lyophilization.

For the inhibition assays, chemicals were dissolved in 95% ethanol and added at the time of galactose induction to the desired final concentration. Cells were harvested 4 days after growth. Relative activity was calculated by normalizing the VLCFA products from inhibited samples to the uninhibited (0 μM inhibitor) control and appropriate empty vector controls with herbicides added at the same concentration. For substrate feeding assays, fatty acids were prepared in 0.5 M in ethanol and fed using a modified protocol (55). Cells were grown as aforementioned, and fatty acids were added at 2 mM final concentration 1 day after induction, along with Tween-80 (0.1% w/v). The same volume of ethanol and Tween-80 were added to the vehicle controls. Following 4 days of growth, cells were harvested by centrifugation.

FAME analysis

Briefly, 40 mg of lyophilized yeast powder was weighed and transferred to a Teflon capped vial. To each vial, 2 ml of a 1.5 M H_2SO_4 solution containing 50 ng/μl nonadecanoic acid (internal standard) was added, then heated to 90 °C for 1.5 h with periodic mixing. After cooling, 1.5 ml of pentane and 2 ml 0.9% NaCl was added to extract methyl esters. The organic phase was concentrated under N_2 gas, resuspended in pentane, and analyzed by GC-MS.

GC-MS

Samples were analyzed using the Agilent 5977A series GC/MSD system fitted with a 30 m × 320 μM × 1 μM HP-5 column. The initial injection temperature of 150 °C was held for 1 min, ramped to 280 °C at a rate of 4 °C/min, then held at 280 °C for 7 min with a helium gas flow rate of 1.2 ml/min. Compound peaks from the resulting chromatograms were identified using mass spectra and retention time comparison with authentic standards (Fig. S5). Peak areas were normalized to the weight of lyophilized yeast, internal and external standards (methyl heptadecanoate for saturated VLCFAs and methyl octadecenoate for unsaturated VLCFAs) to obtain compound abundance.

Homology modeling and ligand docking

A version of Alphafold2 using MMseq2 homology searching was used to predict the structure of KCSs from amino acid sequences (42, 43). Models were generated using multiple Type III PKS templates (Protein Data Bank IDs: 6dx8, 3e1h, 3wxy, 3wd8, 3wxz, 1xes, 1xet, 3tsy, 4b0n, 5wx4, 2h84, 1u0m, 3vs8, 5hwo, and 5hwq) (Tables S2 and S3). Most sites were modeled at greater than 90% confidence according to the local distance difference test (Fig. S13) (56). The transmembrane domain (residues 1–102 in PtKCS1) anchors the enzyme to the membrane of the endoplasmic reticulum but was not modeled here due to low confidence (47). PyMol (version 2.5.2) was used for visualization (57).

Malonyl-CoA and docosanoic acid were docked to the model using AutodockVina (58). Template structures cocrystallized with their substrates or substrate analogs were used to guide the docking of these ligands, which can be more reliable compared to docking without a reference (59). The dimensions of the grid box were 22 × 10 × 9 Å for docking docosanoic acid and 14 × 13 × 17 Å for docking malonyl-CoA. The scoring for accepted ligand docking conformations was between –5 to –6 kcal/mol, with a docking RMSD of less than 4 Å. The overall structure of the KCS homology model was corroborated using the Phyre2 server (60).

Phylogenetic and correlation analyses

Amino acid sequences from the *P. trichocarpa* (v3.0) genome were obtained from Phytozome (61). MEGAX was used to align sequences (MUSCLE algorithm) and build the maximum-likelihood phylogenetic tree (62). To calculate the density of transposable elements, annotated repeat sequences on Chr 10 were obtained from Phytozome and kernel density was calculated along the sequence with a bin width of 15 kb. The metabolite-metabolite correlation analysis was conducted using R and METAGENassist (63). Levels of metabolites were normalized by the autoscaling method and then a correlation matrix was computed with nonparametric Spearman rank correlation.

Data availability

All data are contained within the article and supporting information.

Abbreviations—The abbreviations used are: cDNA, complementary DNA; Chr, chromosome; FAE, fatty acid elongation; FAME, fatty acid methyl ester; VLCFA, very long chain fatty acid; WGD, whole genome duplication.

References

1. Batsale, M., Bahammou, D., Fouillen, L., Mongrand, S., Joubès, J., and Domergue, F. (2021) Biosynthesis and functions of very-long-chain fatty acids in the responses of plants to abiotic and biotic stresses. *Cells* **10**, 1284
2. Holthuis, J. C. M., Pomorski, T., Raggers, R. J., Sprong, H., and Van Meer, G. (2001) The organizing potential of sphingolipids in intracellular membrane transport. *Physiol. Rev.* **81**, 1689–1723
3. Kim, J., Jung, J. H., Lee, S. B., Go, Y. S., Kim, H. J., Cahoon, R., *et al.* (2013) Arabidopsis 3-ketoacyl-coenzyme A Synthase9 is involved in the synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids, and Phospholipids1[W]. *Plant Physiol.* **162**, 567–580
4. Roudier, F., Gissot, L., Beaudoin, F., Haslam, R., Michaelson, L., Marion, J., *et al.* (2010) Very-long-chain fatty acids are involved in polar auxin transport and developmental patterning in Arabidopsis. *Plant Cell* **22**, 364–375
5. Shang, B., Xu, C., Zhang, X., Cao, H., Xin, W., and Hu, Y. (2016) Very-long-chain fatty acids restrict regeneration capacity by confining pericycle competence for callus formation in Arabidopsis. *PNAS* **113**, 5101–5106
6. Bach, L., and Faure, J.-D. (2010) Role of very-long-chain fatty acids in plant development, when chain length does matter. *Comptes Rendus Biologies* **333**, 361–370
7. Chai, M., Queralta Castillo, I., Sonntag, A., Wang, S., Zhao, Z., Liu, W., *et al.* (2021) A seed coat-specific β-ketoacyl-CoA synthase, KCS12, is critical for preserving seed physical dormancy. *Plant Physiol.* **186**, 1606–1615
8. Todd, J., Post-Beittenmiller, D., and Jaworski, J. G. (1999) KCS1 encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in Arabidopsis thaliana. *Plant J.* **17**, 119–130
9. Serrano, M., Coluccia, F., Torres, M., L'Haridon, F., and Métraux, J.-P. (2014) The cuticle and plant defense to pathogens. *Front. Plant Sci.* **5**, 274
10. Xue, D., Zhang, X., Lu, X., Chen, G., and Chen, Z.-H. (2017) Molecular and evolutionary mechanisms of cuticular wax for plant drought tolerance. *Front. Plant Sci.* **8**, 621
11. Jetter, R., Kunst, L., and Samuels, A. L. (2006) Composition of plant cuticular waxes. In *Annual Plant Reviews Volume 23: Biology of the Plant Cuticle*, John Wiley & Sons, Ltd, Hoboken, NJ: 145–181
12. Paul, S., Gable, K., Beaudoin, F., Cahoon, E., Jaworski, J., Napier, J. A., *et al.* (2006) Members of the Arabidopsis FAE1-like 3-ketoacyl-CoA synthase gene family substitute for the elop proteins of Saccharomyces cerevisiae. *J. Biol. Chem.* **281**, 9018–9029
13. Domergue, F., Chevalier, S., Créach, A., Cassagne, C., and Lessire, R. (2000) Purification of the acyl-CoA elongase complex from developing rapeseed and characterization of the 3-ketoacyl-CoA synthase and the 3-hydroxyacyl-CoA dehydratase. *Lipids* **35**, 487–494
14. Kim, J., Kim, R. J., Lee, S. B., and Chung Suh, M. (2021) Protein-protein interactions in fatty acid elongase complexes are important for very-long-chain fatty acid synthesis. *J. Exp. Bot.* **73**, 3004–3017
15. Haslam, T. M., and Kunst, L. (2013) Extending the story of very-long-chain fatty acid elongation. *Plant Sci.* **210**, 93–107
16. Fehling, E., and Mukherjee, K. D. (1991) Acyl-CoA elongase from a higher plant (Lunaria annua): metabolic intermediates of very-long-chain acyl-CoA products and substrate specificity. *Biochim. Biophys. Acta* **1082**, 239–246
17. Millar, A. A., and Kunst, L. (1997) Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. *Plant J.* **12**, 121–131
18. Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., *et al.* (2008) The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling. *Plant Mol. Biol.* **67**, 547
19. Guo, H.-S., Zhang, Y.-M., Sun, X.-Q., Li, M.-M., Hang, Y.-Y., and Xue, J.-Y. (2016) Evolution of the KCS gene family in plants: the history of gene duplication, sub/neofunctionalization and redundancy. *Mol. Genet. Genomics* **291**, 739–752
20. Blacklock, B. J., and Jaworski, J. G. (2006) Substrate specificity of Arabidopsis 3-ketoacyl-CoA synthases. *Biochem. Biophys. Res. Commun.* **346**, 583–590

21. Han, J., Lühs, W., Sonntag, K., Zähringer, U., Borchardt, D. S., Wolter, F. P., *et al.* (2001) Functional characterization of β-ketoacyl-CoA synthase genes from Brassica napus L. *Plant Mol. Biol.* **46**, 229–239

22. Venegas-Calerón, M., Beaudoin, F., Sayanova, O., and Napier, J. A. (2007) Co-Transcribed genes for long chain polyunsaturated fatty acid biosynthesis in the protozoon perkinsus marinus include a plant-like FAE1 3-ketoacyl coenzyme A synthase. *J. Biol. Chem.* **282**, 2996–3003

23. Li, X., Teitgen, A. M., Shirani, A., Ling, J., Busta, L., Cahoon, R. E., *et al.* (2018) Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function. *Nat. Plants* **4**, 711–720

24. Gonzales-Vigil, E., Hefer, C. A., von Loessl, M. E., La Mantia, J., and Mansfield, S. D. (2017) Exploiting natural variation to uncover an alkene biosynthetic enzyme in poplar. *Plant Cell* **29**, 2000–2015

25. Hegebarth, D., Buschhaus, C., Joubès, J., Thoraval, D., Bird, D., and Jetter, R. (2017) Arabidopsis ketoacyl-CoA synthase 16 (KCS16) forms C36/C38 acyl precursors for leaf trichome and pavement surface wax. *Plant Cell Environ.* **40**, 1761–1776

26. Millar, A. A., Clemens, S., Zachgo, S., Giblin, E. M., Taylor, D. C., and Kunst, L. (1999) CUT1, an Arabidopsis gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell* **11**, 825–838

27. Fiebig, A., Mayfield, J. A., Miley, N. L., Chau, S., Fischer, R. L., and Preuss, D. (2000) Alterations in CER6, a gene identical to CUT1, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* **12**, 2001–2008

28. Haslam, T. M., Mañas-Fernández, A., Zhao, L., and Kunst, L. (2012) Arabidopsis ECERIFERUM2 is a component of the fatty acid elongation machinery required for fatty acid extension to exceptional lengths. *Plant Physiol.* **160**, 1164–1174

29. Wang, X., Guan, Y., Zhang, D., Dong, X., Tian, L., and Qu, L. Q. (2017) A β-ketoacyl-CoA synthase is involved in rice leaf cuticular wax synthesis and requires a CER2-LIKE protein as a Cofactor1. *Plant Physiol.* **173**, 944–955

30. Pascal, S., Bernard, A., Sorel, M., Pervent, M., Vile, D., Haslam, R. P., *et al.* (2013) The Arabidopsis cer26 mutant, like the cer2 mutant, is specifically affected in the very long chain fatty acid elongation process. *Plant J.* **73**, 733–746

31. Alexander, L. E., Okazaki, Y., Schelling, M. A., Davis, A., Zheng, X., Rizhsky, L., *et al.* (2020) Maize Glossy2 and glossy2-like genes have overlapping and distinct functions in cuticular lipid Deposition1. *Plant Physiol.* **183**, 840–853

32. Yang, X., Wang, Z., Feng, T., Li, J., Huang, L., Yang, B., *et al.* (2018) Evolutionarily conserved function of the sacred lotus (Nelumbo nucifera Gaertn.) CER2-LIKE family in very-long-chain fatty acid elongation. *Planta* **248**, 715–727

33. Gonzales-Vigil, E., vonLoessl, M. E., Chen, J. Y., Li, S., Haslam, T. M., Kunst, L., *et al.* (2021) Understanding the role of Populus ECER-IFERUM2-likes in the biosynthesis of very-long-chain fatty acids for cuticular waxes. *Plant Cell Physiol.* **62**, 827–838

34. Denic, V., and Weissman, J. S. (2007) A molecular caliper mechanism for determining very long-chain fatty acid length. *Cell* **130**, 663–677

35. Cahoon, E. B., Lindqvist, Y., Schneider, G., and Shanklin, J. (1997) Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4872–4877

36. Chacón, M. G., Fournier, A. E., Tran, F., Dittrich-Domergue, F., Pulsifer, I. P., Domergue, F., *et al.* (2013) Identification of amino acids conferring chain length substrate specificities on fatty alcohol-forming reductases FAR5 and FAR8 from Arabidopsis thaliana. *J. Biol. Chem.* **288**, 30345–30355

37. Blacklock, B. J., and Jaworski, J. G. (2002) Studies into factors contributing to substrate specificity of membrane-bound 3-ketoacyl-CoA synthases. *Eur. J. Biochem.* **269**, 4789–4798

38. Trenkamp, S., Martin, W., and Tietjen, K. (2004) Specific and differential inhibition of very-long-chain fatty acid elongases from Arabidopsis thaliana by different herbicides. *PNAS* **101**, 11903–11908

39. Weng, J.-K., Philippe, R. N., and Noel, J. P. (2012) The rise of chemo-diversity in plants. *Science* **336**, 1667–1670

40. Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., *et al.* (2006) The genome of black cottonwood, Populus trichocarpa (torr. & gray). *Science* **313**, 1596–1604

41. Oh, C.-S., Toke, D. A., Mandala, S., and Martin, C. E. (1997) ELO2 and ELO3, homologues of the Saccharomyces cerevisiae ELO1 gene, function in fatty acid elongation and are required for sphingolipid formation. *J. Biol. Chem.* **272**, 17376–17384

42. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589

43. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2021) ColabFold - making protein folding accessible to all. *Nat. Met.* **19**, 679–682

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

45. Mathieu, M., Zeelen, J. P., Pauptit, R. A., Erdmann, R., Kunau, W. H., and Wierenga, R. K. (1994) The 2.8 A crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of Saccharomyces cerevisiae: a five-layered alpha beta alpha beta alpha structure constructed from two core domains of identical topology. *Structure* **2**, 797–808

46. Austin, M. B., and Noel, J. P. (2003) The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110

47. Ghanevati, M., and Jaworski, J. G. (2002) Engineering and mechanistic studies of the Arabidopsis FAE1 β-ketoacyl-CoA synthase, FAE1 KCS. *Eur. J. Biochem.* **269**, 3531–3539

48. Lechelt-Kunze, C., Meissner, R. C., Drewes, M., and Tietjen, K. (2003) Flufenacet herbicide treatment phenocopies the fiddlehead mutant in Arabidopsis thaliana. *Pest Manag. Sci.* **59**, 847–856

49. Lamberth, C. (2016) Chloroacetamide herbicides. In *Bioactive Carboxylic Compound Classes*, John Wiley & Sons, Ltd, Hoboken, NJ: 293–302

50. Franke, R., Höfer, R., Briesen, I., Emsermann, M., Efremova, N., Yephremov, A., *et al.* (2009) The DAISY gene from Arabidopsis encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds. *Plant J.* **57**, 80–95

51. González-Mellado, D., Salas, J. J., Venegas-Calerón, M., Moreno-Pérez, A. J., Garcés, R., and Martínez-Force, E. (2019) Functional characterization and structural modelling of Helianthus annuus (sunflower) ketoacyl-CoA synthases and their role in seed oil composition. *Planta* **249**, 1823–1836

52. Eckermann, C., Matthes, B., Nimtz, M., Reiser, V., Lederer, B., Böger, P., *et al.* (2003) Covalent binding of chloroacetamide herbicides to the active site cysteine of plant type III polyketide synthases. *Phytochemistry* **64**, 1045–1054

53. Kingsley, L. J., and Lill, M. A. (2015) Substrate tunnels in enzymes: structure-function relationships and computational methodology. *Proteins* **83**, 599–611

54. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59

55. Willey, M., Ochs, M., Busse, C., and McDonough, V. (2020) A transcriptional regulatory system of the S. cerevisiae OLE1 gene responds to fatty acid species and intracellular amount, and not simply membrane status. *J. Lipids* **2020**, e3903257

56. Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728

57. Schrödinger, LLC (2015) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York, NY

58. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461

59. Kroemer, R. T. (2007) Structure-based drug design: docking and scoring. *Curr. Protein Pept. Sci.* **8**, 312–328

60. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858

61. Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucl. Acids Res.* **40**, D1178–D1186

62. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) Mega X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549

63. Arndt, D., Xia, J., Liu, Y., Zhou, Y., Guo, A. C., Cruz, J. A., *et al.* (2012) METAGENassist: a comprehensive web server for comparative metagenomics. *Nucl. Acids Res.* **40**, W88–W95