BNL-111732-2016-JA

Engineering a monolignol 4-O-methyltransferase with high selectivity for the condensed lignin precursor coniferyl alcohol

Y. Cai, M-W Bhuiya, J. Shanklin, C-J Liu

Brookhaven National Laboratory, Upton, NY 11973 USA

Submitted to the Journal of Biological Chemistry

September 2015

Biology Department

Brookhaven National Laboratory

U.S. Department of Energy
DOE Office of Basic Energy Sciences

Notice: This manuscript has been co-authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

This preprint is intended for publication in a journal or proceedings. Since changes may be made before publication, it may not be cited or reproduced without the author's permission.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Engineering a monolignol 4-*O*-methyltransferase with high selectivity for the condensed lignin precursor coniferyl alchohol

Yuanheng Cai, Mohammad-Wadud Bhuiya[ǂ], John Shanklin and Chang-Jun Liu*

Biological, Environmental & Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11973, USA

Running Title: Engineering a coniferyl alcohol specific MOMT

To whom correspondence should be addressed: Chang-Jun Liu, Biological, Environmental & Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11973, USA; Tel: 631-344-2966

Key words: Monolignol, 4-*O*-methyltransferase, Lignin, Crystal Structure

Background: Guaiacyl lignin dominates the polymer's condensation, therefore, negatively affecting plant cell wall's digestibility. **Results:** A promiscuous *O*-methyltransferase was incrementally evolved to constrain its substrate specificity to guaiacyl lignin precursors.
Conclusion: The enzyme with nine amino acid substitutions is functionally specialized for selectively methylating the condensed lignin precursors.
Significance: The obtained enzyme is a candidate for specifically altering lignin structure to improve plant cell wall's digestibility.

ABSTRACT

Lignin, a rigid biopolymer in plant cell walls, is derived from the oxidative polymerization of three monolignols. The composition of monolignol monomers dictates the degree of lignin condensation, reactivity, and thus the degradability of plant cell walls. Guaiacyl lignin is regarded as the condensed structural unit. Polymerization of lignin is initiated through the deprotonation of the *para*-hydroxyl group of monolignols. Therefore, preferentially modifying the *para*-hydroxyl of a specific monolignol to deprive its dehydrogenation propensity would disturb the formation of particular lignin subunits. Here, we test the hypothesis that specific remodeling the active site of a monolignol 4-*O*-methyltransferase would create an enzyme that specifically methylates the condensed guaiacyl lignin precursor coniferyl alcohol. Combining crystal structural information with combinatorial active site saturation mutagenesis, and starting with the promiscuous engineered enzyme, MOMT5 (T133L-E165I-F175I-F166W-H169F), we incrementally remodeled its substrate binding pocket by the addition of four substitutions, i.e., M26H, S30R, V33S, and T319M, yielding a mutant enzyme capable of discriminately etherifying the *para*-hydroxyl of coniferyl alcohol, even in the presence of excess sinapyl alcohol. The engineered enzyme variant has a substantially reduced substrate-binding pocket that imposes a clear steric hindrance thereby excluding bulkier lignin precursors. The resulting enzyme

variant represents an excellent candidate for modulating lignin composition and/or structure *in planta*.

Lignocelluloses or the biomass of plant cell walls represents the most abundant renewable feedstock for producing liquid biofuels or high value bio-based chemicals (1,2). Nonetheless, the recalcitrance of lignin to breakdown has largely impeded the applications of cell wall biomass. Lignin, one of the major structural components in lignocellulosic materials, is a complex aromatic biopolymer composed of polymerized hydroxycinnamyl alcohols, known as monolignols. There are three major monolignols, namely *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol that differ in the number of methoxy groups on their aromatic rings, which lead to correspondingly three distinct structural subunits, the hydroxylphenyl- (H), guaiacyl- (G), and sinapyl- (S) units when they are incorporated into lignin polymer (**Fig. 1A**). In gymnosperms lignin mainly contains G and H subunits; in dicotyledonous species, lignin consists of primarily G and S, with only trace amount of H subunits; grass lignin contains relatively higher H lignin units compared with dicot plants. Lignin monomers link to each other through an oxidative coupling process initiated by the deprotonation of the phenolic *para*-hydroxy group (**Fig. 1**) (3). Owing to the difference in the degrees of methoxylation of the phenyl rings, G and S lignin subunits can form different types and numbers of subunit cross-linkages. The G monomer can be incorporated into lignin by biphenyl and other carbon-carbon linkages, giving rise to highly condensed polymers, whereas S subunits usually form relatively more labile ether bonds through the 4-hydroxy group (4). Therefore, the composition of the monolignol monomers determines the degree of lignin condensation, the complexity of the lignin structure, and thereby, the degradability of plant cell walls (4). In general, the high S/G ratio of lignin is often associated with increased pulping yields (5,6), and enzymatic sugar release (7), while a high content of G-units is associated with poor lignin degradation owing to the presence of C-C bonds at the free C5 position (8). Lignin engineering has emerged as an efficient tool to produce biomass with better digestibility (9). Since the growth and development of the plants usually is compromised by a dramatically reduced overall lignin content, while plants seem to tolerate a wide range of lignin structure alteration (10), perceivable interest in lignin engineering has centered on tuning the monomeric composition of lignin. In particular, engineering plants with high S to G ratio, i.e., enriched in syringyl monomers and/or with reduced guaiacyl lignin, has been targeted for efficient utilization of plant cell wall biomass.

Previously, we evolved a number of monolignol 4-*O*-methyltransferase variants (MOMTs) that can trans-methylate the *para*-hydroxyls of monolignols, which prevents the incorporation of modified precursors into the lignin polymer by disrupting oxidative radical formation (11,12) (**Fig. 1B**). However, the engineered enzyme is biochemically active against both coniferyl- and sinapyl-alcohol with similar catalytic turnover rates (e.g., MOMT3 and 4 in Table 1 and 2). Therefore, the transgenic plants with overexpressed MOMT maintained similar monomer composition and lignin structure as the wild-type plants (12). We thus hypothesized that if a variant enzyme could preferentially 4-*O*-methylate coniferyl alcohol with respect to sinapyl alcohol, it should specifically interfere with the incorporation of G monomers into lignin, and limit the amount of the condensed subunits. This could potentially lead to improvement in cell wall digestibility.

Protein engineering is an effective tool to modify protein functionality by mimicking and accelerating natural evolution at a

laboratory scale (13). The difference between S- and G- lignin monomer results from the numbers of methoxyl groups in their phenyl rings. S-monomer possesses two methoxyl moieties at its 3- and 5- position of phenyl ring thus is structurally bulkier (**Fig. 1A**). Based on such structural disparity, we hypothesized that remodeling the substrate binding pockets of the previously obtained non-specific MOMTs would allow the enzyme to discriminate between lignin monomers thereby constraining its substrate specificity to coniferyl alcohol. We tested this hypothesis by using a structure-guided combinatorial saturation mutagenesis approach to reshape the binding pocket of a substrate promiscuous MOMT in a step-wise fashion and successfully obtained a functionally specialized variant, MOMT9, that shows a substantially altered substrate preference for coniferyl alcohol. Crystallization and structural determination revealed that MOMT9 possesses a significantly smaller substrate binding pocket compared to its parent enzyme, which impedes the binding of the bulkier sinapyl alcohol, explaining its strong specificity for coniferyl alcohol.

EXPERIMENTAL PROCEDURES

Chemicals—[methyl-^{14}C] *S*-Adenosyl-L-methionine (SAM) was purchased from American Radiolabelled Chemicals, Inc. The BugBuster and Bradford solution, respectively, were from Novagen (Madison, WI) and Bio-Rad. All other chemicals were obtained from Sigma.

Library construction—The MOMT5 coding sequence was cloned into a pET28a(+) vector, using *Nde*I and *Bam*HI sites to generate pET28aMOMT5 plasmid. The plasmid was then used as the template for constructing the mutant library. To identify the residues that potentially are involved in substrate recognition and discrimination, we inspected the residues within a 10 Å distance around the bound monolignols in the crystal structures of MOMT5-monolignol complexes. Four groups of sites of mutagenesis, i.e., M26-S30-V33, P129-L133, A134-L139, and W166-F169, were initially identified according to their geometrical proximity to the bound monolignol, and the principles of CAST (combinatorial active-site saturation test) library construction, i.e. if one member of the pair of amino acids at position "n" then the second one being chosen should be at (n+1) in a loop, (n+2) in a β sheet, (n+3) in a 3_{10} helix, and (n+4) in a α helix (14).

To construct the defined four CAST libraries and two additional saturated-mutagenesis libraries at sites of F179 and T319, we adopted the strategy of QuickChange site-directed mutagenesis using the primers listed in **Table 1**. NDT (D: adenine/guanine/thymine) degenerate codon was used, which encodes for the reduced 12 amino acid alphabet, so that we could generate focused yet highly enriched libraries (15).

For additional construction of a W143-Y326 CAST library, we employed a dual-tube megaprimer strategy (16). Briefly, megaprimers randomized at sites 143 and 326 were first PCR-amplified using the primer pair MOMT_r4D3W143XF, and IEMT-326R (**Table 1**). Then, the megaprimer PCR products were purified and subsequently applied in constructing the CAST library using the QuickChange strategy.

Library screening and enzyme assay—For screening enzyme activity, we used an established 96-well plate-radioactivity assay (11) with minor modifications. Briefly, the *E. coli* cells harboring the library plasmids were cultured, induced, and lysed with BugBuster solution as previously described (11). The lysate was directly used for initial functional screening. In the screening of MOMT5 mutant variants, equal amount of coniferyl alcohol or sinapyl alcohol (at 500 μM) together with ^{14}C-SAM (at 500 μM) were used in the assays; while in the screening of MOMT8 variants, we used 20 μM coniferyl alcohol and 300 μM

sinapyl alcohol, respectively, in the assays to impose more stringent screening for obtaining enzymes preferential for coniferyl alcohol. After incubation at 30 °C for 10 min, each well of reaction was partitioned with 150 µl water-saturated ethyl acetate and 50 µl ethyl acetate solvent containing the extracted product from each well was transferred into a new microplate and then taken for counting radioactivity with a microplate scintillation counter (TopCount NXT, Packard). The conversion rates of enzymes for coniferyl- and sinapyl-alcohols were then calculated. We selected transformants with a higher conversion rate of coniferyl alcohol than of sinapyl alcohol. The substrate preferences of selected mutant variants were further validated in the assays via the same method but using the purified enzymes.

For assaying enzyme-specific activity, the purified enzymes were incubated in 50 mM Tris-Cl buffer (pH 7.5) containing 1 mM DTT, 500 µM SAM, with 500 µM phenolic substrates. The reaction proceeded at 30 °C for 5~10 min. To determine the kinetic parameters, a series of monolignol concentrations ranging from 5- to 500-µM were used, while maintaining the SAM concentration at 500 µM. The products were analyzed by HPLC, using a reverse-phase C18 column (Luna 5u C18(2), 250x 4.6 mm, Phenomenex). Samples were resolved in a mobile phase of 0.2% acetic acid (A) with an increasing concentration gradient of acetonitrile containing 0.2% acetic acid (B) at 0 to 2 min, 5% (B); 2 to 30 min, 5- to 50 % (B); 30 to 32 min, 50 to 100% and then 100% for 2 min at a flow rate of 1 ml/min. UV absorption was monitored at 254-, 280-, 310-, and 330-nm using a multiple-wavelength photodiode array detector.

Crystallization and structural determination—Crystals of both MOMT5 and MOMT9 were grown via a hanging-drop vapor diffusion method in 22% (w/v) PEG 4000, 0.3 M $Mg(NO_3)_2$ (pH 7.2), and 1 mM DTT with 1 mM SAH and 1 mM monolignols. MOMT5 was co-crystallized with S-adenosyl homocysteine (SAH), coniferyl alcohol, or sinapyl alcohol; in contrast, MOMT9 was co-crystallized with SAH and coniferyl alcohol only. Diffraction data were collected at the X29- and X25-beamlines of the National Synchrotron Light source. The initial phase was determined by molecular replacement using the MOMT3 crystal structure (PDB: 3TKY) (12) with the CCP4i program suite (17). The model building and refinement respectively, were done with coot (18) and refmac5 (19).

The geometry of the substrate binding pocket was analyzed with the channel finding feature of the 3V website (http://3vee.molmovdb.org/) (20). The radii of the outer and inner probes were set at 3.5- and 0.9-Å, respectively.

Homology modeling and Substrate Docking analysis— Homology modeling of the mutant MOMT8 was accomplished by using the Swiss-model web server (21-23) and employing the crystal structure of MOMT5-C as a template. The docking of coniferyl alcohol to its active site was inspected manually to fit the orientation that monolignol bound in the substrate-binding pocket. The residues within 10 Å distance from the substrate were selected, and potential hot spots were analyzed further and grouped according to CAST rules (14).

RESULTS

Evolving MOMT5 variant and its crystal structure determination—Previously, we have evolved a monolignol 4-*O*-methyltransferase (MOMT3, T133L-E165I-F175I) from an isoeugenol 4-*O*-methyltransferase (11) (Table 1 and 2). The engineered MOMT3 confers regiospecific methylation activity on the *para*-hydroxyls of both coniferyl and sinapyl alcohols (11). Further optimizing MOMT3 activity via iterative saturation mutagenesis generated a set of tetra- and penta-mutant

variants, which include the one carrying two additional substituted amino acids at its substrate binding pocket, i.e., F166W and H169F, designated MOMT5 (T133L-E165I-F166W-F175I-H169F). Enzyme activity assay and kinetic analysis revealed that MOMT5 differs from the reported MOMT3 and MOMT4 mutants (11,12). It not only possesses a considerable 4-*O*-methyaltion activity to both coniferyl and sinapyl alcohols (**Table 2**), but also exhibited a discernible substrate preference towards coniferyl alcohol (**Table 3**). Its specificity constant (k_{cat}/K_m) for coniferyl alcohol is about 2.7-fold higher than that for sinapyl alcohol (**Table 3**).

To gain insights into its substrate binding, catalysis, and the emerging substrate specificity, we co-crystallized MOMT5 with both coniferyl alcohol (MOMT5-C), and sinapyl alcohol (MOMT5-S) in the presence of SAH, the de-methylation product of the methyl donor, SAM. Crystallization of the MOMT5-phenolic substrate generated diffracted crystals with monoclinic space-group $P2_1$ (**Table 4**). The diffraction data were phased by molecular replacement using the previously determined crystal structure MOMT3 (PDB: 3TKY). After refinement, the structures of the MOMT5-substrate complexes were solved to 1.8 Å, and to 1.7 Å, respectively, for the coniferyl alcohol- and sinapyl alcohol-complexes.

The overall tertiary and quaternary structures of MOMT5 are almost identical to the MOMT3 structures (12), and are reminiscence of classic plant phenolic *O*-methyltransferases, particularly, the caffeic acid/5-hydroxy ferulic acid 3/5 *O*-methyltransferases (COMT) (24,25) (**Fig. 2**). In each asymmetric unit cell, the crystal contains four MOMT5 molecules. Each two monomers form a biological dimer along a non-crystallographic two- fold axis, in which each monomer comprises a small N-terminal domain that is involved in mediating dimerization, and contributes to the building of a portion of the substrate binding pocket, and a larger Rossmann-fold containing C-terminal domain that is mainly responsible for substrate binding and catalysis (**Fig. 2A**).

In both the MOMT5-confieryl alcohol (MOMT5-C) and -sinapyl alcohol (MOMT5-S) complexes, the monolignols are bound to the substrate binding pocket in an almost identical manner, except that the end group of propene tail is positioned slightly variable. This positioning, however, substantially differs from that of coniferyl alcohol bound in the structure of MOMT3 (**Fig. 3A-C**). Compared to the monolignol bound in MOMT3 (**Fig. 3A**), the 3-methoxyl moiety of the bound monolignol in either MOMT5-C and MOMT5-S complexes fits more snugly into a hydrophobic core formed by an array of nonpolar residues, including Ile 165, Trp 166, Phe 169, Phe 179, and Tyr 326 (**Fig. 3B** and **C**). This optimized positioning most likely is attained by the increased hydrophobicity and the corresponding geometric alteration of the nonpolar core with respect to the substitutions of Phe 166 to Trp, and His 169 to Phe, the two additional mutations leading to MOMT5 from MOMT3. This orients the entire rigid hydroxycinnamyl alcohol in ~30 degree rotation in the binding pocket, relative to that in MOMT3 structure (**Fig. 3A-C**). Under this binding mode, the plane of phenyl ring is sandwiched by the side chains of a pair of Met residues (Met 183 and Met 323), a feature commonly found in other phenolic *O*-methyltransferase structures (24-26). The 5-methoxyl of sinapyl alcohol in MOMT5-S complex protrudes to a hydrophobic cavity flanked by Trp 269, Phe 130, Thr 319, and the Met 26 from the dimerized MOMT molecule (**Fig. 3C**). The propene tail of monolignol occupies a hydrophobic pocket bordered by residues Phe 130, Leu 133, Leu 139, Leu 322, and Tyr 326. Because of the rigid body rotation, the 9-hydroxyl moiety of the tail does not form hydrogen bonds with the side chain of Ser 30 (from the dyad molecule), seen in

the MOMT3-coniferyl alcohol complex (**Fig. 3A-C**). This altered binding mode promotes the close proximity of the 4-hydroxyl of the bound monolignol to the catalytic base, His 272, favoring its deprotonation. In addition, the proximity of 4-hydroxyl of monolignol to the side chain of Asp 273 may permit the formation of a hydrogen-bond to further stabilize its positioning to enable efficient nucleophile formation and the subsequent trans-methylation.

Directed evolution of MOMT5 to alter its substrate preference—Sinapyl alcohol bears both 3- and 5- methoxyl moieties on its phenyl ring, and thus is bulkier than coniferyl alcohol (**Fig. 3D**). The structure elucidation of the complexes of MOMT5-C and MOMT5-S suggests that while both monolignols bind identically in the active site, the potential hindrance of the binding pocket to 5-methoxyl moiety of sinapyl alcohol might explain the slight preference of MOMT5 for coniferyl alcohol over sinapyl alcohol. Therefore, we hypothesized that enhancing such steric hindrance may result in more profound substrate discrimination. Site-directed saturated mutagenesis (SM) has been proven to be effective in protein directed evolution (27) and we successfully adopted it in our previous mutagenesis studies to evolve and enhance MOMT activity (11,12). However, using this approach is difficult to explore the synergistic effect of two or more point mutations. To circumvent this obstacle, we implemented CAST library construction method that allows simultanous mutation of a few explicitly defined geometrically-correlated amino-acid sites (14,28). Our previous work in engineering MOMT3 and 4 mainly focused on individually substituting the residues involved in binding the phenyl rings of monolignols (11,12), whereas residues contributing to building the hydrophobic cores accommodating the propene tail and the 5-methoxyl moiety of monolignol were not investigated. Therefore, based on their structural proximity to the phenolic substrate, and the principles of CAST design, we defined three groups of residues, the M26-S30-V33, P129-L133, and A134-L139, which directly or indirectly participate in the interaction with 5-methoxyl and the propanoid tail of sinapyl alcohol. In addition the pair residues of W166-F169 that are involved in constituting a hydrophobic core for holding the 3-methoxyl of the phenolic compound was revisited based on CAST strategy (**Fig. 4**).

Adopting the NDT degeneracy strategy (where N = A, T, G or C and D = A, G or T), we constructed four saturated mutant libraries corresponding to each group of defined amino acid residues. Using ^{14}C labeled S-adenosyl-L-methionine (SAM) as the co-substrate, the mutant variants were functionally screened with a 96-well plate format. After screening about 500 colonies for each library, one mutant carrying substitutions M26H-S30R-V33S, designated as MOMT8, was identified with a substantial improvement in substrate preference towards coniferyl alcohol. With the purified proteins, we found that the MOMT8 mutant variant exhibited the highest specific activity to coniferyl alcohol among the range of phenolic substrates we examined (**Table 2**). Kinetically, it displayed more than 8-fold higher conversion rate to coniferyl alcohol than to sinapyl alcohol with a k_{cat} of 0.049 s^{-1} versus 0.006 s^{-1}. In addition, the binding affinity of MOMT8 to coniferyl alcohol was also increased by 60% in comparison with that to sinapyl alcohol. Overall, the MOMT8 mutant exhibited 13.4-fold increased specificity constant to coniferyl alcohol with respect to sinapyl alcohol (**Table 3**).

Homology modeling of MOMT8 revealed that simultaneous substitutions of Met 26 to His, Ser 30 to Arg, and Val 33 to Ser substantially altered the size and shape of the substrate binding site; in particular, the substitution of Ser 30 to Arg offers a longer, bulkier side chain potentially protruding into

the substrate binding pocket, which imposes further steric constraints to the space that positions the propene tail and 5-methoxyl substituent of sinapyl alcohol (**Fig. 5**). This change further impairs the binding of sinapyl alcohol compared with the MOMT5 structure (**Fig. 5A** and **B**).

Since coniferyl alcohol lacks a 5-methoxy substituent, we rationalized that further reducing or shaping the substrate binding pocket at the sites directly interacting with this moiety, e.g., T319 (**Fig. 4** and **5**), or modulating the residues interfering with the positioning of the phenyl ring of the substrate, e.g., F179, W143-Y326 (**Fig. 4**), might further eliminate the potential in binding sinapyl alcohol, and thus enhancing the enzyme's discrimination to two predominant monolignols. Therefore, the mutant libraries were created at the sites of T319, F179 and W143-Y326. In screening the generated saturation mutation libraries, we used 15 times lower concentration of coniferyl alcohol than sinapyl alcohol, i.e., 20 µM versus 300 µM in the assays, respectively, to enable a more stringent screening condition to identify variants with preference for coniferyl alcohol. A new variant, MOMT9, was identified, that has an additional mutation, T319M. The MOMT9 variant exhibited greater substrate preference for coniferyl alcohol among a set of hydroxycinnamyl alcohols and aldehydes that we examined (**Table 2**). Kinetically, MOMT9 showed nearly a 7-fold tighter binding affinity and more than 25-fold higher specificity constant for coniferyl alcohol than to sinapyl alcohol (**Table 3**).

To verify MOMT9's effectiveness in selectively methylating monolignols, we incubated it with a mixture of coniferyl- and sinapyl- alcohol substrates at three ratios that mimic the ratios of G and S lignin monomers reportedly presented in Arabidopsis stems (29), poplar stems (30) and leaves (31) (**Fig. 6**). When both monolignols were presented to the bifunctional control MOMT5, the correspondingly 4-*O*-methylated products accumulated in proportion to the substrate mixing ratios (**Fig. 6A-C**). This finding demonstrates that MOMT5 is functionally unable to discriminate between the two monolignols. In contrast, under the same assay conditions, for all the ratios of substrates, MOMT9 exclusively yielded only one methylated product, the 4-*O*-methylated coniferyl alcohol, even when the amount of the sinapyl alcohol substrate was twice that of coniferyl alcohol (**Fig. 6D-E**). These data demonstrate that MOMT9 essentially recognizes coniferyl alcohol as the sole substrate, even when sinapyl alcohol is in excess.

Structural determination of MOMT9—To visualize the structural basis underlying the significantly increased substrate specificity of MOMT9, we co-crystallized it with coniferyl alcohol (MOMT9-C), in the presence of SAH, the demethylation product of the methyl donor SAM. The crystal structure was solved to 1.82 Å (**Table 4**).

As seen in Figure 5, the most significant structural difference between MOMT5 and MOMT9 is the size of their substrate binding pockets. In MOMT5, the set of hydrophobic residues form a spacious substrate binding pocket large enough to accommodate bulky substrates such as sinapyl alcohol with substituent of the 3- and 5-methoxy groups (**Fig. 5A and C**). In contrast, in MOMT9, the corresponding binding pocket is largely occluded due to the protruding bulky side chains of Arg 30 and Met 319. Together with the geometric coordination of the mutations of His 26 and Ser 33, these four additional substitutions, relative to MOMT5, preclude the binding of bulky phenolic substrates, thus shifting the enzyme preference to solely act on coniferyl alcohol (**Fig 5C** and **D**). Measuring the volume and surface area of the substrate binding pockets of the complexes MOMT3, MOMT5-C, MOMT5-S, and MOMT9-C revealed that the volume of the pocket

changed significantly from 661 $Å^3$ in MOMT5-S complex to 478.5 $Å^3$ in MOMT9, which corresponds to an approximately 30% volume reduction (**Table 5**).

DISCUSSION

Directed evolution of a functionally specialized enzyme with a new substrate often proceeds via generalist intermediates, mimicking Darwinian evolution, in which a progenitor enzyme with broad substrate specificity often serves as the starting point to elaborate new functionality (32). Previously a phenylpropene 4-*O*-methyltransferase from a wild flower *Clarkia breweri* (Fairy fans) (denoted IEMT) (33,34) was chosen for evolving monolignol 4-*O*-methyltransferase activity (11). IEMT specifically recognizes allylphenol eugenol and isoeugenol as the substrates. It arose from a gene duplication and the subsequent accumulation of mutations within its ancestor, the caffeic acid/5-hydroxy ferulic acid 3/5 *O*-methyltransferases (COMT) that is involved in monolignol biosynthesis (**Fig. 7**). During this process, distinct substrate specificity and regiospecific methylation properties evolved (34). As Figure 7 illustrates, our previous engineering MOMT from IEMT essentially invoked a "reverse evolution" path, in which IEMT's substrate-recognition constraint was re-invoked within the evolutionarily plastic active sites. The resulting variants (i.e., MOMTs) regained the ability to bind lignin biosynthetic precursors, coniferyl- and sinapyl- alcohols, while retaining the ability to perform 4-*O*-methylation (11). This directed evolution effort essentially broadened IEMT's substrate recognition. The promiscuous MOMT variants represent an ideal starting point for evolving the functional specialization. As guaiacyl lignin has been a target for modulating lignin condensation, an enzyme variant with capacity to specifically modify G-lignin precursors represents an interesting candidate for perturbing G-lignin synthesis. In the present study, a monolignol 4-*O*-methyltransferase variant, MOMT5, was incrementally evolved with focusing on the sites lining its binding pocket, which yielded a new variant that effectively discriminate the major lignin precursors and preferentially etherify guaiacyl lignin precursor coniferyl alcohol (**Fig.7**).

In evolving MOMT from IEMT, increased hydrophobicity of the binding pocket such as the replacement of Glu165 with Ile, Phe or Val correlates with the gain of monolignol methylating activity (11). In contrast, when tuning MOMT5 into a more specialized enzyme for G-lignin precursors, steric hindrance of the binding pocket determines its substrate preference. The substitutions of Ser 30 and Thr 319 for the bulkier residues Arg and Met, respectively, prevented the binding of bulkier sinapyl alcohol, making enzyme in favor of coniferyl alcohol (**Table 6**).

Evolving a specific enzyme with a constrained substrate specificity or a novel function is more difficult than to relax the enzyme's substrate constraint (35). In contrast to our previous single site iterative saturation mutagenesis, we created mutant libraries based on the CAST principle. CAST was designed to exploit the synergistic geometric effects of saturating small portions of a given protein's active site (36). It was applied successfully in broadening substrate specificity (14), or altering stereo- and/or regio-selectivity (28, 37, 38). The current study provides an additional example how CAST effectively constrains enzyme's substrate specificity. In this study, we also adopted a NDT-degeneracy strategy, instead of the previous NNK strategy for simultaneously randomizing two or three sites to construct the more focused libraries. NDT facilitates the substitution of an amino acid to one of 12 different amino acids, mostly with dissimilar properties (15). With combinatorial randomization at three sites (e.g., M26-S30-V33), NDT degeneracy limits the library size to $\sim 1.7\times10^3$, while retaining high-quality

coverage. To ensure 95% library coverage, usually threefold oversampling is required, i.e., the number of variants to be screened in the NDT library for 95% coverage is about 500 for a two-position randomized mutant and 5,000 for a three-position mutant library (15). Notably, in our screening of the three-position mutant library, the beneficial mutant, i.e., MOMT8, with desired improved substrate specificity was obtained twice after screening only 500 mutants, i.e., only ~10% of the required sampling size for 95% coverage. This result validates the effectiveness and robustness of our structure-based position selection and evolution design.

We had previously targeted T319 using single position saturated mutagenesis in evolving IEMT to MOMT3, but it did not lead to obvious beneficial improvements in terms of changing catalytic efficiency or substrate selectivity of the parent enzyme (11). However, after the accumulation of eight mutations in MOMT8, an additional substitution at this site, T319M, resulted in approximately 5-fold decrease of the binding affinity to sinapyl alcohol and a substantial increase in its catalytic efficiency to coniferyl alcohol (**Table 3**). These data demonstrate that the evolution trajectory is important for constructing an enzyme with desired properties, i.e., that the functional effects of a mutation can depend on the presence of other mutations.

The evolution of a specialized enzyme from a generalist can be accomplished in different ways, either with positive selection of high activity on one substrate, or negative selection to remove those variants having the undesired activity (35). Designing proper screening procedure is essential to attain desired enzyme properties. This was particularly exemplified in our study. In our first round mutagenesis, we screened the mutant variants of MOMT5 for activity preferring for coniferyl alcohol by using the same concentration of coniferyl alcohol and sinapyl alcohol in the assays. Although the resulting enzyme MOMT8 catalytically distinguished two monolignol substrates, its discrimination was the result of a compromised turnover rate (k_{cat}) for sinapyl alcohol, while the binding affinities (K_m) for both monolignols are comparable to those of the parent enzyme MOMT5 (**Table 3**), because no substrate availability pressure was imposed in the functional screening. In contrast, during evolving MOMT9 from MOMT8, a strategy was designed to maintain an order-of-magnitude higher concentration of the non-desired substrate sinapyl alcohol in the assay than that of the desired substrate coniferyl alcohol. As a consequence, the new variant MOMT9 displays a 7-fold higher binding affinity for coniferyl alcohol than for sinapyl alcohol; and its catalytic efficiency for coniferyl alcohol is 25- fold higher than in using sinapyl alcohol as the substrate.

Determination of the crystal structure of the evolved variant enzymes MOMT5 and MOMT9 clearly illustrates the structural basis for the incremental evolution of the substrate binding pocket. Reduction in size of the substrate binding pocket of MOMT9 confirms that steric hindrance on the binding and catalysis of bulkier lignin precursor sinapyl alcohol underlies the enzyme's substrate preference.

In conclusion, with the judicial use of structural information along with carefully designed laboratory evolution, we successfully engineered an enzyme variant, MOMT9, that exhibited the desired substrate specificity and reaction selectivity toward etherifying the *para*-hydroxyl of the guaiacyl lignin precursor, conifery alcohol. *Para*-etherification would efficiently block the oxidative coupling of those modified precursors, thereby, impeding their incorporation into lignin. The engineered unique biochemical characteristic of MOMT9 makes it a potential candidate to interfere the synthesis of more condensed guaiacyl lignin subunits in the cell walls of plants and thus

may be useful for biotechnological improvement of cellulosic biomass.

Supporting Information (Fore reviewer only)
Supplemental Data 1. MOMT5-Coniferyl alcohol structural complex
Supplemental Data 2. MOMT5-Sinapyl alcohol structural complex
Supplemental Data 3. MOMT9-Coniferyl alcohol structural complex
Supplemental Data 4. MOMT9-Coniferyl alcohol structural complex validation report
Supplemental Data 5. MOMT-Sinapyl alcohol structural complex validation report
Supplemental Data 6. MOMT5-Coniferyl alcohol structural complex validation report

Data deposition
The atomic coordinates and structure factors for MOMT5-coniferyl alcohol-SAH (code 5CVJ), MOMT5-sinapyl alcohol-SAH (code 5CVU) and MOMT9-coniferyl alcohol-SAH (code 5CVV) have been deposited in the Protein Date Bank (http://www.rcsb.org)

Acknowledgements
We thank Dr. Eran Pichersky, University of Michigan, for sharing the initial parental enzyme *C. breweri* IEMT clone. This work is supported by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy (DOE) through Grant DEAC0298CH10886 to C.J.L. The use of the X-ray beamline X29 and X25 of the National Synchrotron Light source in Brookhaven National Laboratory was supported by the Office of Basic Energy Sciences, DOE, under Contract No. DEAC0298CH10886.

Conflict interest
The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions
Y. C and C.J.L designed experiments; Y.C and M.W.B performed experiments; Y.C and C.J.L analyzed data and wrote paper. JS contributed to experimental design, data interpretation and writing. All the authors analyzed the results, and edited and approved the final version of the manuscript.

References
1. Somerville, C. (2006) Biofuel. *Current Biol.* **17**(4), R115-R119
2. Liu, C. J., Cai, Y., Zhang, X., Gou, M., and Yang, H. (2014) Tailoring lignin biosynthesis for efficient and sustainable biofuel production. *Plant Biotechnol. J.* **12**(9), 1154-1162
3. Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., Marita, J. M., Hatfield, R. D., Christensen, J. H., and Boerjan, W. (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochem. Rev.* **3**, 29-60
4. Boerjan, W., Ralph, J., and Baucher, M. (2003) Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546
5. Huntley, S. K., Ellis, D., Gilbert, M., Chapple, C., and Mansfield, S. D. (2003) Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. *J. Agric. Food. Chem.* **51**(21), 6178-6183
6. Lapierre, C., Pollet, B., PetitConil, M., Toval, G., Romero, J., Pilate, G., Leple, J. C., Boerjan, W., Ferret, V., DeNadai, V., and Jouanin, L. (1999) Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid *O*-methyltransferase

activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiol.* **119**(1), 153-163
7. Studer, M. H., DeMartini, J. D., Davis, M. F., Sykes, R. W., Davison, B., Keller, M., Tuskan, G. A., and Wyman, C. E. (2011) Lignin content in natural Populus variants affects sugar release. *Proc. Natl. Acad. Sci. U S A* **108**(15), 6300-6305
8. Neutelings, G. (2011) Lignin variability in plant cell walls: contribution of new models. *Plant Sci.* **181**(4), 379-386
9. Vanholme, R., Morreel, K., Ralph, J., and Boerjan, W. (2008) Lignin engineering. *Curr. Opin. Plant Biol.* **11**(3), 278-285
10. Bonawitz, N. D., and Chapple, C. (2013) Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty? *Curr. Opin. Biotechnol.* **24**(2), 336-343
11. Bhuiya, M. W., and Liu, C. J. (2010) Engineering monolignol 4-*O*-methyltransferases to modulate lignin biosynthesis. *J. Biol. Chem.* **285**(1), 277-285
12. Zhang, K. W., Bhuiya, M. W., Pazo, J. R., Miao, Y., Kim, H., Ralph, J., and Liu, C.-J. (2012) An engineered monolignol 4-*O*-methyltransferase represses lignin polymerization and confers novel metabolic capability in Arabidopsis. *Plant Cell* **24**, 3122–3139
13. Cobb, R. E., Sun, N., and Zhao, H. M. (2013) Directed evolution as a powerful synthetic biology tool. *Methods* **60**, 81-90
14. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D., and Vogel, A. (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angewandte Chemie (International ed* **44**(27), 4192-4196
15. Reetz, M. T., Kahakeaw, D., and Lohmer, R. (2008) Addressing the numbers problem in directed evolution. *Chembiochem* 9(11), 1797-1804
16. Zhang, W., and Mannervik, B. (2013) An improved dual-tube megaprimer approach for multi-site saturation mutagenesis. *World J. Microbiol. biotechn.* **29**(4), 667-672
17. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**(Pt 7), 1131-1137
18. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**(Pt 4), 486-501
19. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**(Pt 4), 355-367
20. Voss, N. R., and Gerstein, M. (2010) 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38,** W555-562
21. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**(2), 195-201
22. Guex, N., Peitsch, M. C., and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30** Suppl 1, S162-173
23. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37**(Database issue), D387-392
24. Louie, G. V., Bowman, M. E., Tu, Y., Mouradov, A., Spangenberg, G., and Noel, J. P. (2010) Structure-function analyses of a caffeic acid O-methyltransferase from perennial ryegrass reveal the molecular basis for substrate preference. *Plant Cell* **22**(12), 4114-4127
25. Zubieta, C., Kota, P., Ferrer, J.-L., Dixon, R. A., and Noel, J. (2002) Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase. *Plant Cell* **14**(6), 1265-1277
26. Liu, C.-J., Deavours, B. E., Richard, S. B., Ferrer, J.-L., Blount, J. W., Huhman, D., Dixon, R. A., and Noel, J. P. (2006) Structural basis for dual functionality of isoflavonoid *O*-methyltranserases in the evolution of plant defense responses. *Plant Cell* **18**, 3656–3669

27. Reetz, M. T., and Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**(4), 891-903
28. Reetz, M. T., Carballeira, J. D., Peyralans, J., Hobenreich, H., Maichele, A., and Vogel, A. (2006) Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chem. (Weinheim an der Bergstrasse, Germany)* **12**(23), 6031-6038
29. Li, X., Bonawitz, N. D., Weng, J. K., and Chapple, C. (2010) The growth reduction associated with repressed lignin biosynthesis in Arabidopsis thaliana is independent of flavonoids. *Plant Cell* **22**(5), 1620-1632
30. Robinson, A. R., and Mansfield, S. D. (2009) Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. *Plant J.* **58**(4), 706-714
31. Cabane, M., Pireaux, J. C., Leger, E., Weber, E., Dizengremel, P., Pollet, B., and Lapierre, C. (2004) Condensed lignins are synthesized in poplar leaves exposed to ozone. *Plant Physiol.* **134**(2), 586-594
32. Khersonsky, O., Roodveldt, C., and Tawfik, D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508
33. Wang, J., and Pichersky, E. (1998) Characterization of S-adenosyl-L-methionine: (Iso)eugenol *O*-methyltransferase involved in floral scent production in *Clarkia breweri*. *Archiv. Biochem. Biophys.* **349**, 153-160
34. Wang, J., and Pichersky, E. (1999) Identification of specific residues involved in substrate discrimination in two plant *O*-methyltransferases. *Archiv. Biochem. Biophy.* **368**, 172-180
35. Tracewell, C. A., and Arnold, F. H. (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin.Chem. Biol.***13**, 1–7
36. Bommarius, A. S., Blum, J. K., and Abrahamson, M. J. (2011) Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* **15**(2), 194-200
37. Acevedo-Rocha, C. G., Hoebenreich, S., and Reetz, M. T. (2014) Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol. Biol.* **1179**, 103-128
38. Kille, S., Zilly, F. E., Acevedo, J. P., and Reetz, M. T. (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nature Chem.* **3**(9), 738-743

Figure legends

Figure 1. The scheme of lignin polymerization process and monolignol 4-*O*-methyltransferase (MOMT)-catalyzed reaction. (A) Lignin monomeric precursors, (B) MOMT-catalyzed methylation reaction, (C) The dehydrogenation of monolignols, and (D) the subsequent polymerization process.

Figure 2. Crystal structures of MOMTs with the substrate monolignol and the methyl donor product SAH. (A) Overall structure of dimerized MOMT5-SAH-coniferyl alcohol complex, the two molecular chains are in green and magenta; the phenolic substrate (coniferyl alcohol) bound in the active site is depicted as stick-ball representation, and the methyl donor product SAH is color-coded as C in yellow, N in blue, and O in red; N: N-termini, C: C-termini. (B) Superimposition of the tertiary structures of MOMT5 (magenta), MOMT9 (orange), MOMT3 (3TKY, cyan), and the closed form of LpCOMT1 (3P9I).

Figure 3. The close-up view of substrate binding in crystal structures of MOMTs. (A-C) Substrate binding modes in a previously determined MOMT3-coniferyl alcohol (A), MOMT5-coniferyl alcohol (B), and MOMT5-sinapyl alcohol (C) structure complexes. The phenolic substrate is colored in magenta. Doted black lines in (B) and (C) illustrate the rotation of the bound monolignols in structure complex. The doted red line indicates hydrogen bonds. Water molecules are depicted as the green balls. (D) The electron density associated with monolignol as observed crystallographically in the MOMT5-coniferyl alcohol, and MOMT5-sinapyl alcohol structure complexes, respectively. The calculated 2Fo-Fc electron density map was contoured at 1s.

Figure 4. CAST library design based on MOMT5-coniferyl alcohol crystal structure complex. The initially selected four groups of mutation sites are depicted with surface representation. Other targeted sites are color-coded with C in purple and O in red. The bound coniferyl alcohol is depicted as ball-stick representation.

Figure 5. Surface representation of the substrate binding pockets of the structure complexes of MOMT5-sinapyl alcohol (A), MOMT8 homology model (B), MOMT5-coniferyl alcohol (C), and MOMT9- coniferyl alcohol (D). The substitutions of amino acid residues in MOMT8 and 9 are labeled in red.

Figure 6. *In vitro* reaction of MOMT5 (A-C) and MOMT9 (D-F) with mixed monolignol coniferyl alcohol and sinapyl alcohol substrate at G to S ratios of 60.8:37.8 (A and D), 31.23:68.4 (B and E), and 76.2:17.8 (C and F). 4OMeCA: 4-*O*-methylated coniferyl alcohol, 4OMeSA: 4-*O*-methylated sinapyl alcohol.

Figure 7. Scheme of directed evolution path for coniferyl alcohol 4-*O*-methyltransferase, MOMT9. The dot-line box indicates a nature evolution of IEMT from COMT.

Table 1. Degenerate primers used in saturation mutagenesis

Name	5'-3' sequence	
PT166-169F	ggaatgaatatcndtgattacndtggaacagaccac	this work
PT166-169R	gtggtctgttccahngtaatcahngatattcattcc	this work
PT134-139R	ccagggctccaaahngaccttgtcggtahncaggagcaaaaa	this work
PT134-139F	tttttgctcctgndtaccgacaaggtcndtttggagccctgg	this work
PT129-133F	gtttctcttgctndtttttgctcndtgctaccgacaag	this work
PT129-133R	cttgtcggtagcahngagcaaaaaahnagcaagagaaac	this work
PT26-30-33R	ggccatggggagahnggcggcahnggccagctgahnggcgaagaggtt	this work
PT26-30-33F	aacctcttcgccndtcagctggccndtgccgccndtctccccatggcc	this work
MOMT_r4D3W143XF	gtccttttggagcccnnkttttacttgaaagat	this work
MOMT_r7F5F179XF	aacaaggtgnnkaacaagggaatgtcc	this work
MOMT_r7F5F179XR	tcccttgttmnncaccttgttaattct	this work
IEMT-319F	ACCAAGGTAGTCATCCATNNKGACGCCCTCATGTTG	(11)
IEMT-319R	GGCCAACATGAGGGCGTCMNNATGGATGACTACCTT	(11)
IEMT-326R	GCCCTCATGTTGGCCNNKAACCCAGGCGGCAAAGAA	(11)

Table 2. Substrate profiles of MOMT variants
The enzymatic activity was measured from the reaction at 30 °C for 10 min. The data represent the mean ± S.D. of two replicates.

Substrate	MOMT3[a,b]	MOMT4[a]	MOMT5	MOMT8	MOMT9
	(nmol·mg^{-1}·min^{-1})				
coniferyl alcohol	346.3 ± 10.7	66.8 ± 0.7	80.6 ± 0.4	51.4 ± 0.7	68.3 ± 0.3
sinapyl alcohol	192.9 ± 1.3	79.6 ± 0.9	39.6 ± 3.7	7.7 ± 0.1	9.8 ± 0.1
coumaryl alcohol	ND	5.5 ± 0.1	9.6 ± 0.1	2.5 ± 0.2	3.7 ± 0.1
caffeyl alcohol	ND	1.7 ± 0.0	6.5 ± 0.2	10.8 ± 0.2	12.9 ± 0.03
coniferyl aldehyde	127.9 ± 2.1	28.7 ± 0.8	10.9 ± 0.2	3 ± 0.1	5.8 ± 0.2
sinapyl aldehyde	58.7 ± 1.4	14.6 ± 0.5	1.8 ± 0.1	0.4 ± 0.04	5 ± 0.1
coumaryl aldehyde	ND	1.9 ± 0.0	2.3 ± 0.04	0.4 ± 0.1	1 ± 0

[a]Data were from previous reports (11,12)
[b] Data were obtained based on radioactivity assays.
ND: not determined

Table 3. **Kinetic parameters of MOMT variants.**
Kinetic parameters were obtained from the reactions at 30 °C for 10 min. The data represent the mean ± S.D. of two replicates.

Enzyme	Substrate	k_{cat} (x10^{-3} s^{-1})	K_m (μM)	k_{cat} / K_m	Catalytic ratio (coni/sina)
MOMT3[a]	coniferyl alc.	247±10	198.5±23.7	1246	0.85
	sinapyl alc.	173±4	118.7±11.0	1463	
MOMT4[a]	coniferyl alc.	528±2	192.6±24.5	2739	0.68
	sinapyl alc.	271±11	68.1±11.1	3999	
MOMT5	coniferyl alc.	69±9	11.4±1.6	6053	2.7
	sinapyl alc.	33±1	14.7±2.3	2245	
MOMT8	coniferyl alc.	49±1	16.1±1.3	3043	13.4
	sinapyl alc.	6±0.1	26.3±4.1	228	
MOMT9	coniferyl alc.	63±1	16.1±1.4	3913	25.3
	sinapyl alc.	17±0.2	109.5±4	155	

[a]Data were from previous reports (11,12), in which they were obtained based on radioactivity assays.

Table 4. Structure statistics of MOMT5 and MOMT9

	MOMT5-C	MOMT5-S	MOMT9-C
Wavelength (Å)	1.1	1.1	1.1
Space group	$P2_1$	$P2_1$	$P\ 2_1$
Resolution (Å) [a]	65.94-1.68 (1.77-1.68)	50.81-1.60(1.69-1.60)	49.9-1.73(1.82-1.73)
Cell parameters			
a, b, c (Å)	66.07, 150.36, 67.93	66.54, 151.63, 68.39	68.29, 67.48, 74.07
β	93.19	92.55	93.18
Observed reflections[a]	838684(73068)	877559(73869)	410628(56591)
Unique reflections[a]	146779(18641)	156919(17226)	70249(10169)
Average I/(I) [a]	14.6(2.86)	13.7(3.11)	11.3(3.15)
Completeness (%)[a]	97.8(85.1)	88.6(67.2)	100.0(100.0)
R_{merge} (%) [a,b]	7.2(36.0)	7.0(35.1)	8.0(34.5)
R factor[b]	0.1678	0.177	0.1909
R_{free}[b]	0.2075	0.2152	0.2415
No. of amino acid residues	1428	1429	697
Number of water molecules	842	475	287
RMSD			
bond length	0.021	0.023	0.019
bond angle	2.245	2.36	2.103

[a] Numbers in parentheses refer to the highest shell.
[b] Rmerge = $|I_h - \langle I_h \rangle|/I_h$, where $\langle I_h \rangle|$ is the average intensity over symmetry equivalent reflections.
[c] R factor = $\sum|F_{obs} - F_{calc}|/ \sum F_{obs}$, where summation is over the data used for refinement, R_{free} was calculated using 5% of data excluded from refinement.

Table 5. Substrate binding pocket (SBP) geometry of MOMT variants

	MOMT3-C	MOMT5-C	MOMT5-S	MOMT9-C
SBP volume ($Å^3$)	624.3±99.1	578.3±42.9	661.0±40.1	478.5±9.2
SBP surface area ($Å^2$)	656.8±115.1	627.3±16.8	661.5±22.0	551.5±53.0

Table 6. Summary of amino acid substitutions in the obtained MOMT variants

Variant	Amino acid change
IEMT	Wild type
MOMT1	E165I
MOMT3	T133L-E165I-F175I
MOMT4	T133L-E165I-F175I-H169F
MOMT5	T133L-E165I-F175I-F166W-H169F
MOMT8	T133L-E165I-F175I-F166W-H169F-M26H-S30R-V33S
MOMT9	T133L-E165I-F175I-F166W-H169F-M26H-S30R-V33S-T319M

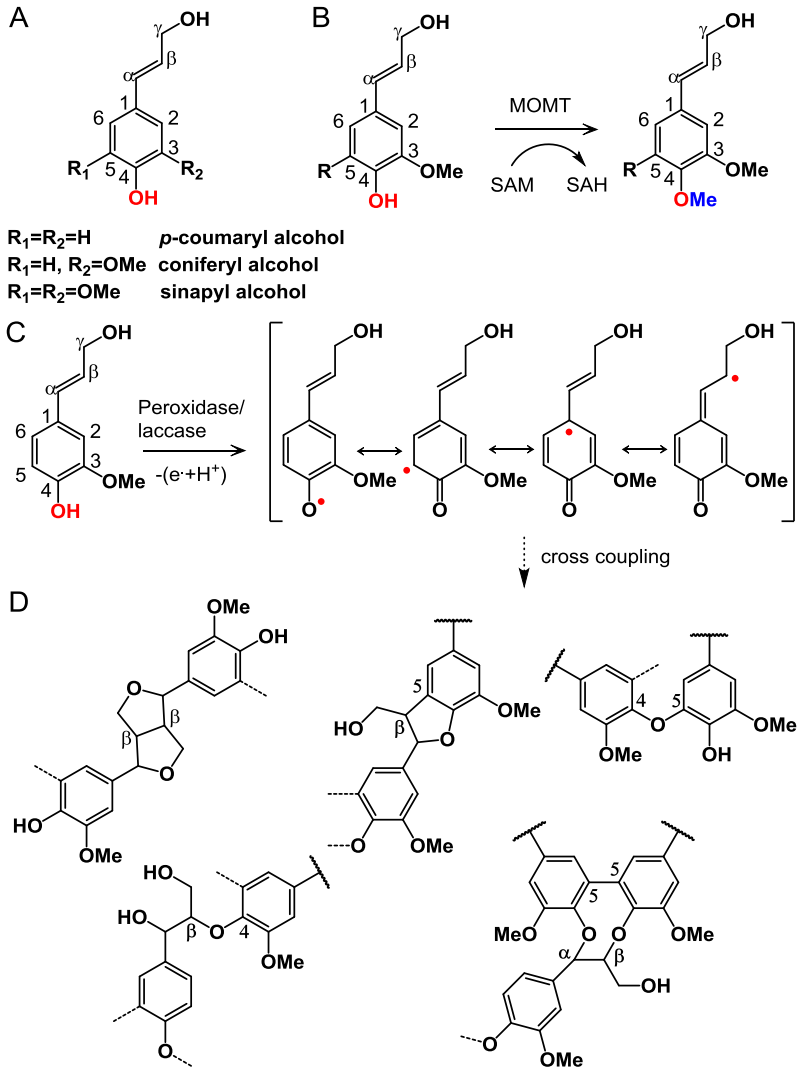

Figure 1. The scheme of lignin polymerization process and monolignol 4-*O*-methyltransferase (MOMT)-catalyzed reaction. (A) Lignin monomeric precursors, (B) MOMT-catalyzed methylation reaction, (C) The dehydrogenation of monolignols, and (D) the subsequent polymerization process.

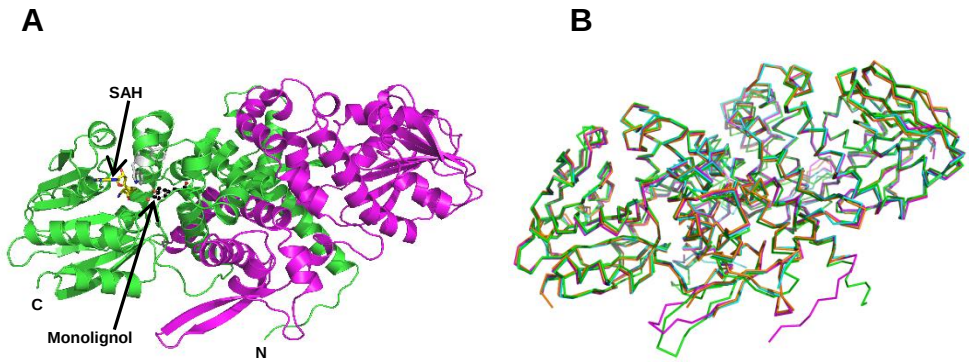

Figure 2. Crystal structures of MOMTs with the substrate monolignol and the methyl donor product SAH. (A) Overall structure of dimerized MOMT5-SAH-coniferyl alcohol complex, the two molecular chains are in green and magenta; the phenolic substrate (coniferyl alcohol) bound in the active site is depicted as stick-ball representation, and the methyl donor product SAH is color-coded as C in yellow, N in blue, and O in red; N: N-termini, C: C-termini. (B) Superimposition of the tertiary structures of MOMT5 (magenta), MOMT9 (orange) with MOMT3 (3TKY, cyan) and the closed form of LpCOMT1 (3P9I).

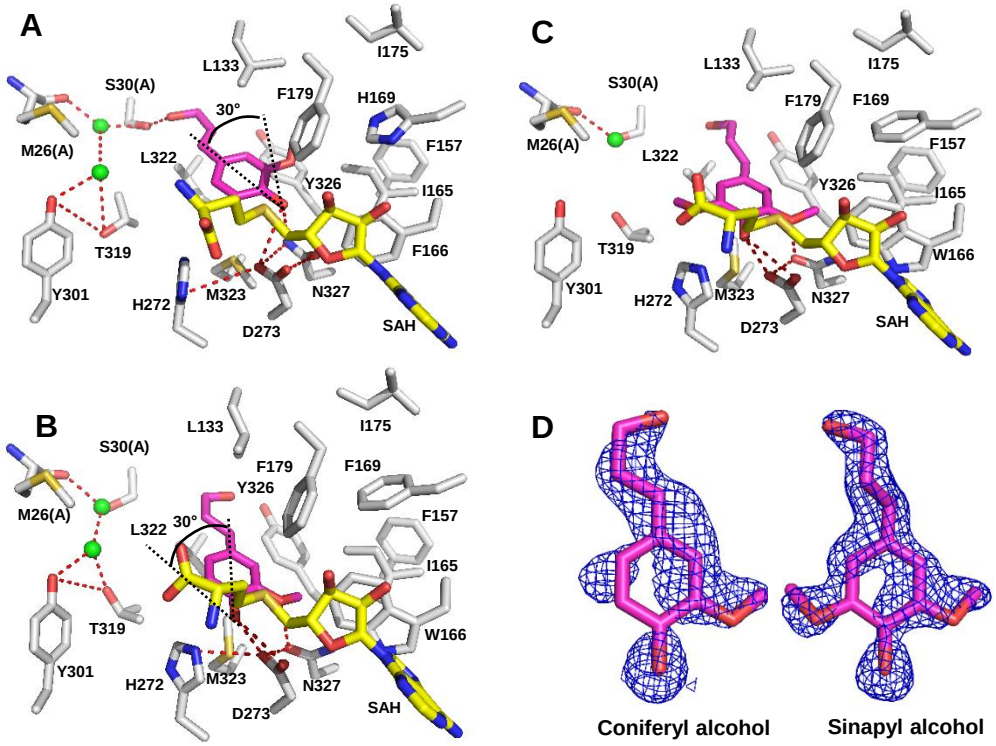

Figure 3. The close-up view of substrate binding in crystal structures of MOMTs. (A-C) Substrate binding modes in a previously determined MOMT3-coniferyl alcohol (A), MOMT5-coniferyl alcohol (B), and MOMT5-sinapyl alcohol (C) structure complexes. The phenolic substrate is colored in magenta. Doted black lines in (B) and (C) illustrate the rotation of the bound monolignols in structure complex. The doted red line indicates hydrogen bonds. Water molecules are depicted as the green balls. (D) The electron density associated with monolignol asobserved crystallographically in the MOMT5-coniferyl alcohol, and MOMT5-sinapyl alcohol structure complexes, respectively. The calculated 2Fo-Fc electron density map was contoured at 1s.

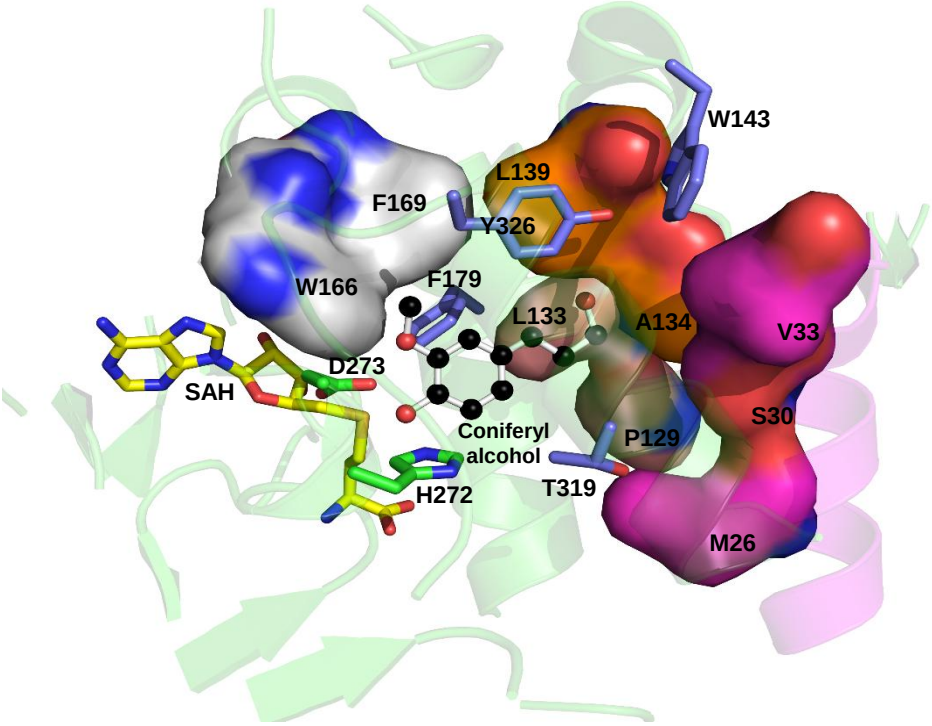

Figure 4. CAST library design based on MOMT-coniferyl alcohol crystal structure complex. The initially selected four groups of mutation sites are depicted with surface representation. Other targeted sites are color-coded with C in purple and O in red. The bound coniferyl alcohol is depicted as ball-stick representation.

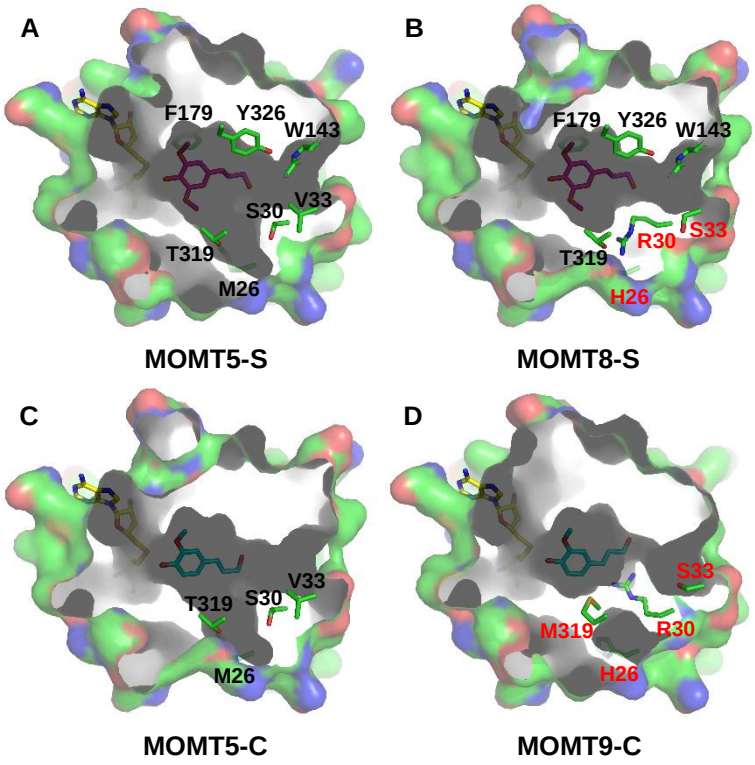

Figure 5. Surface representation of the substrate binding pockets of the structure complexes of MOMT5-sinapyl alcohol (A), MOMT8 homology model (B), MOMT5-coniferyl alcohol (C), and MOMT9- coniferyl alcohol (D). The substitutions of amino acid residues in MOMT8 and 9 are labeled in red.

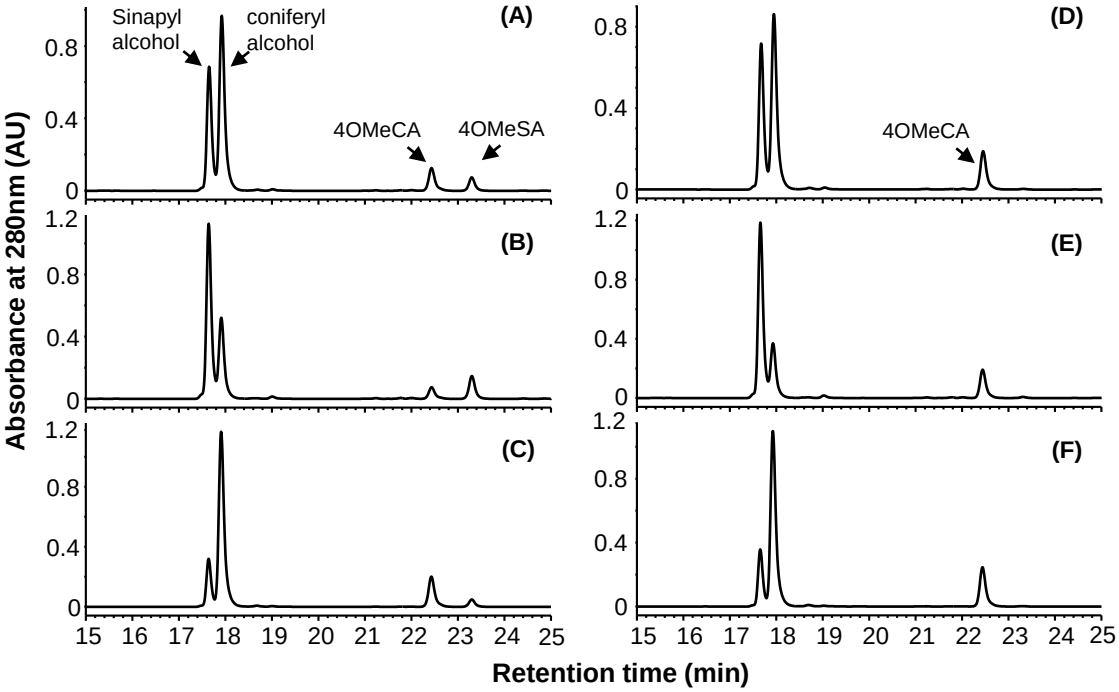

Figure 6. *In vitro* reaction of MOMT5 (A-C) and MOMT9 (D-F) with mixed monolignol coniferyl alcohol and sinapyl alcohol substrate at G to S ratios of 60.8:37.8 (A and D), 31.23:68.4 (B and E), and 76.2:17.8 (C and F). 4OMeCA: 4-*O*-methylated coniferyl alcohol, 4OMeSA: 4-*O*-methylated sinapyl alcohol.

.

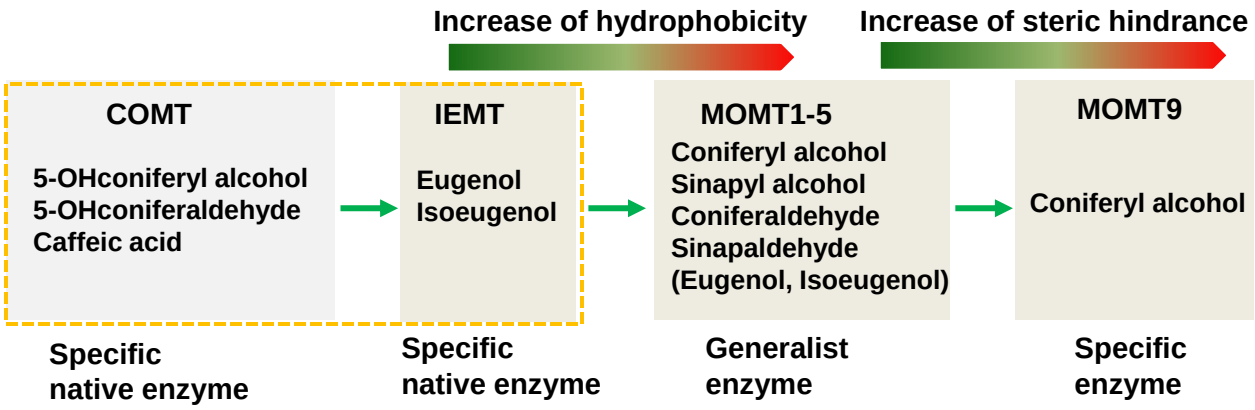

Figure 7. Scheme of directed evolution path of coniferyl alcohol 4-*O*-methyltransferase, MOMT9. The dot-line box indicates a nature evolution of IEMT from COMT.