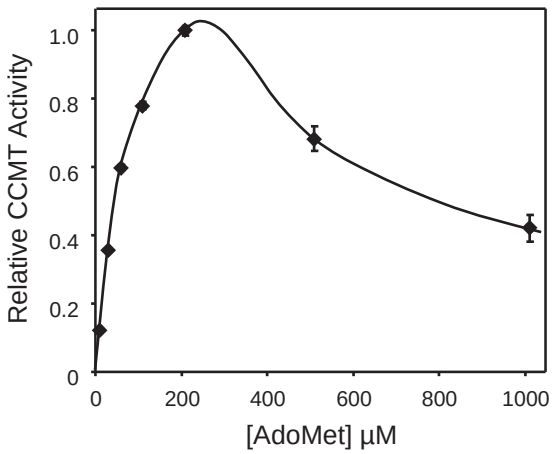
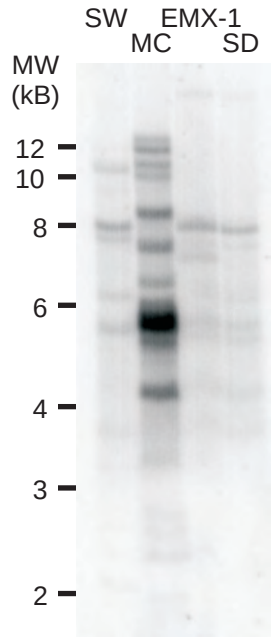
Supplemental Data. Kapteyn et al. 2007. Evolution of cinnamate/*p*-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate.

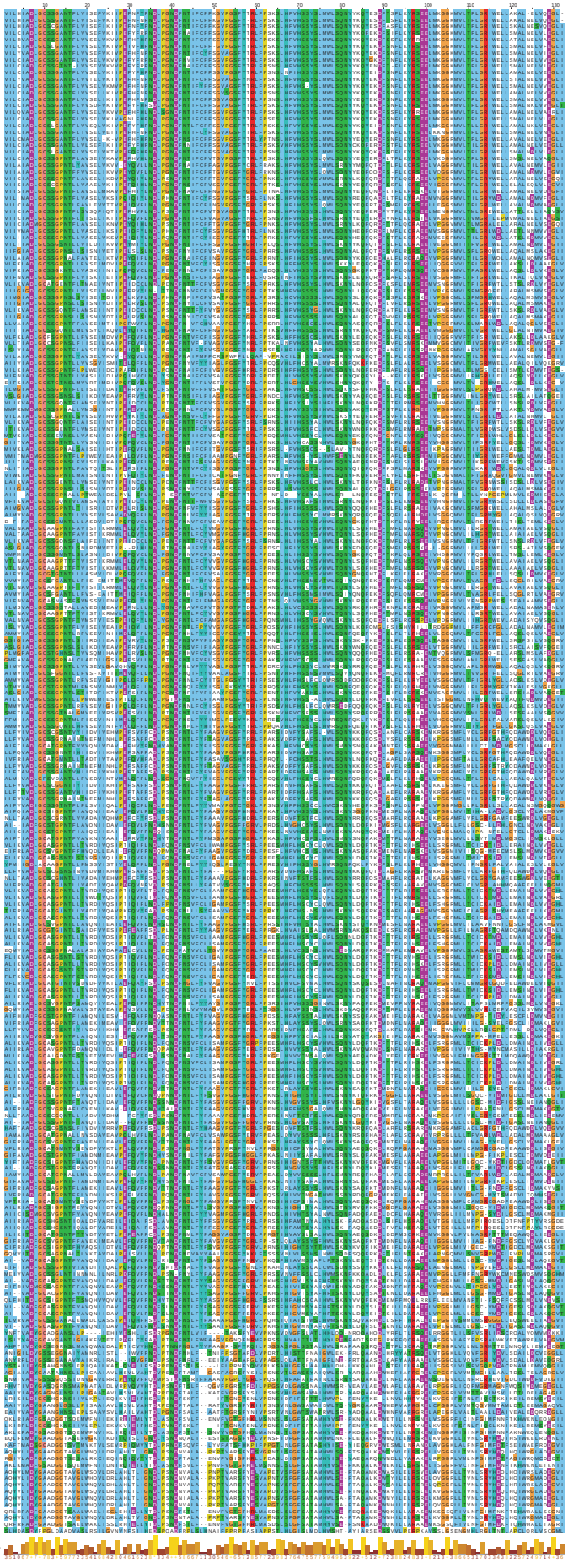
Supplemental Figure 1. AdoMet inhibition of CCMT activity in MC glandular trichomes. Crude protein extracts from young leaves were assayed at constant cinnamate concentration and varying AdoMet concentration. Activity/inhibition curves obtained for protein extracts from isolated trichomes were essentially identical.



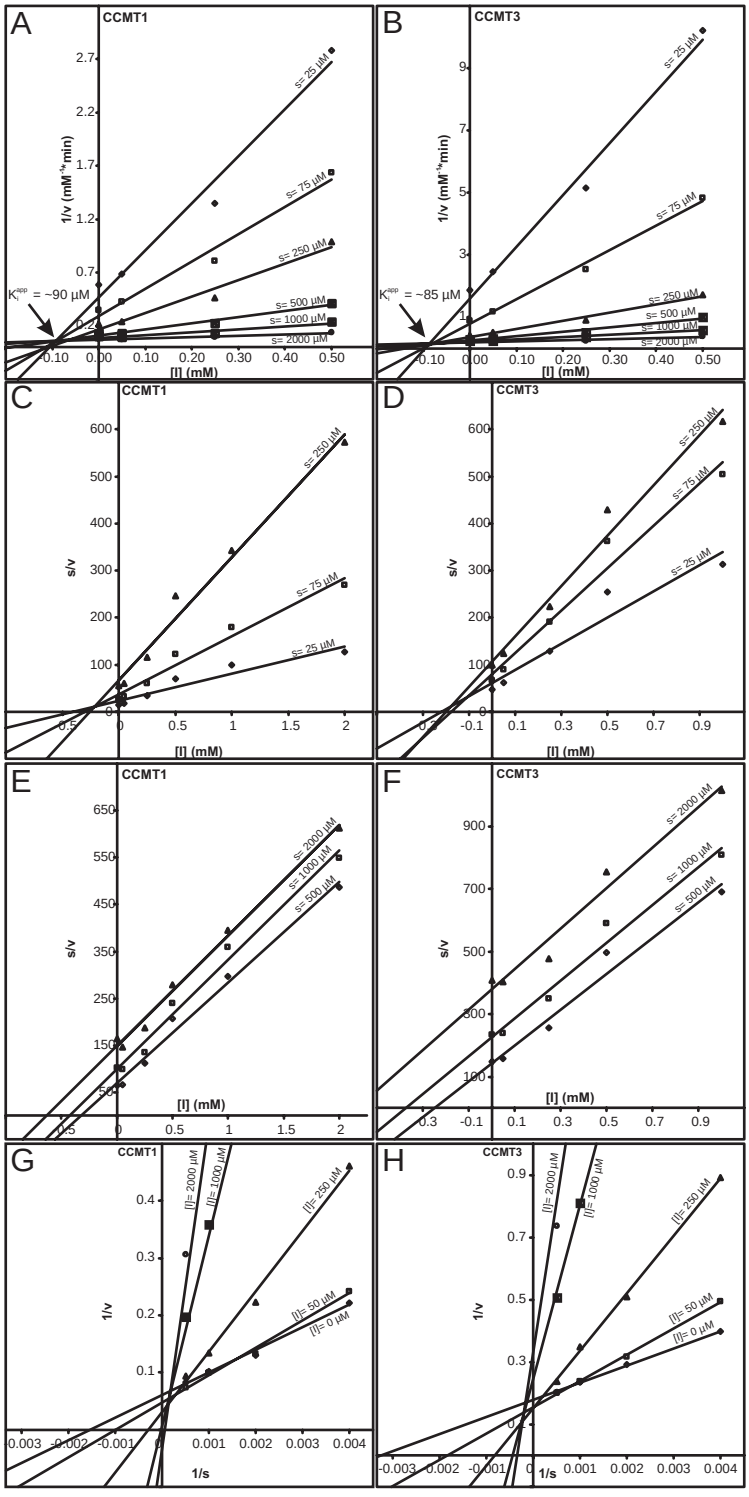
Supplemental Figure 2. DNA blot analysis of CCMT copy number and organization in the genomes of the four basil lines, SW, MC and EMX-1 and SD. Total genomic DNA was analyzed as described in the Methods section. Note that EMX-1 and SD have almost identical banding patterns, and share many bands with line SW, whereas line MC's banding pattern is much more complex and very different from the other lines, suggesting that gene duplication events are likely to have occurred in this line, relative to the others.



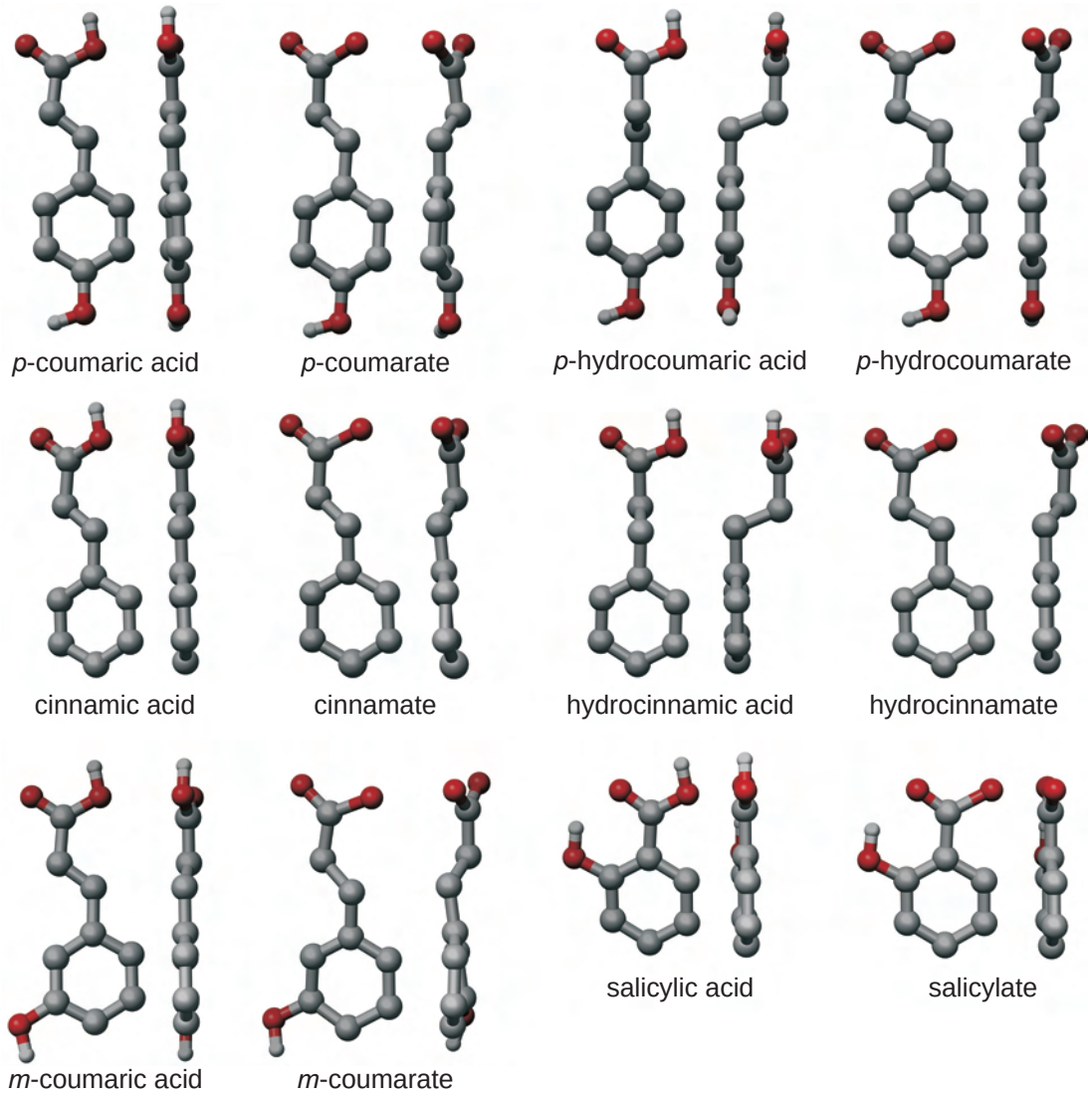
Supplemental Figure 3. Sequence alignment used to generate maximum likelihood tree displayed in Figure 9.



Supplemental Figure 4. Additional plots demonstrating the parabolic nature of SA inhibition of CCMT1 (**A**, **C**, **E**, and **G**) and CCMT3 (**B**, **D**, **F** and **H**) activities. **A** and **B**, 1/v vs. [I] plots used to calculate apparent Ki for SA. **C** and **D**, s/v vs. [I] plots at low [substrate] suggest that the inhibition approximates competitive inhibition at low [s]. **E** and **F**, s/v vs. [I] plots at high [substrate] suggest that the inhibition approximates mixed inhibition at high [s]. **G** and **H**, 1/v vs. 1/s plots clearly show classical parabolic inhibition, where the plots do not intersect at a single point, but rather form a parabolically shaped intersection curve.



Supplemental Figure 5. Energy minimized structures of potential substrates (in free acid and carboxylate forms) for CCMT enzymes used in docking experiments, displayed from two sides.



Supplemental Table 1. Peptide identification of CCMT isoforms from MC and SW glandular trichome soluble protein samples using proteomics approach. Residues highlighted in bold show differences between isoforms.

Peptide sequence	# of spectra	Peptide location	CCMT isoform match	Protein source
FVVADLGCSSGRN	1	64-76	conserved	MC
KQFQSDL**GV**FLRS	3	197-209	CCMT1	MC
KQFQSDL**GA**FLRS	2	197-209	CCMT2	MC
KQFQSDL**VS**FLRS	1	197-209	CCMT3	MC
RTSPDPA(E)DQGAWILTFSTRY	1	227-246	CCMT1,2	MC
RYQDAWNDLVQEGLISSEKR	1	245-264	conserved	MC
KRDTFNIPIYTPSLEEFKE	2	263-281	conserved	SW
RDGAFIINKL	3	285-294	conserved	MC
KLQLFHGGSALIIDDPNDAVEISRA	9	293-317	conserved	MC,SW
RSLTGGLVDAHIGDQLGHELFSRL	17	323-346	conserved	MC,SW
RLLS**Q**(R)AV**D**(A)QAKE	11	345-356	CCMT1	MC,SW

Supplemental Methods

Energy Minimization of Potential Substrates' Structures

Potential substrates for CCMT enzymes were energy minimized to obtain the most stable conformation(s) of the molecules, i.e., those having the least steric hindrance, for use in docking experiments with the modeled CCMT and SAMT proteins and the two mutants (see Supplemental Figure 5 online for examples). Both the free acid and the carboxylate forms of all potential substrate molecules were analyzed in this manner. The free (hydroxy)cinnamic acids (protonated forms of the molecules) existed as rigid planar or almost planar conformations in the most minimized state. An alternate conformation, where the carboxyl group lied perpendicular to the plane of the rest of the molecule, was also found as a local minimum for *p*-coumaric acid, cinnamic acid and *m*-coumaric acid, but only if the initial molecular conformation used for the minimization run was similar to this conformation. Slight perturbation away from this conformation in subsequent energy minimization runs led to formation of the (near) planar conformation for the protonated forms of these molecules, suggesting that the (near) planar conformers are the most stable (lowest energy) for the free (hydroxy)cinnamic acids. With the carboxyl group protonated, resonance structures can form, leading to a partial positive charge on the phenolic group (for *p*-coumaric acid) or one of the ring carbons (for *m*-coumaric and cinnamic acids) along with a partial negative charge on the carbonyl oxygen of the carboxyl groups. Such resonance structures, also stabilized by interaction of the pi bonds of the ring with those of the two side chain carbons and also with those of the double bond between the carbonyl carbon and oxygen of the (hydroxy)cinnamic acids, would keep the molecules that are in the protonated (free acid) state near planar and prevent free rotation of the carboxyl group. As long

as the non-carbonyl carbon of the carboxyl group is protonated, the carboxyl group would be restricted in its ability to freely rotate with respect to the rest of the molecule.

In contrast, removal of the proton from the carboxyl group, which is what occurs at physiological pH, would prevent this resonance-based stabilization by preventing formation of the partial negative charge on the carbonyl oxygen. When the deprotonated forms of the (hydroxy)cinnamic acids, i.e., the carboxylates, were analyzed, less planar and more flexible conformations were found to be the most energetically favorable. Importantly, the restrictions on carboxyl group rotation relative to the rest of the molecule were reduced for *p*-coumarate, cinnamate and *m*-coumarate. Thus, the carboxylates of these molecules were much more flexible than the free acids. The protonated and deprotonated forms of hydrocinnamic and *p*-hydrocoumaric acids were also very flexible, because these compounds lack a double bond between carbons 8 and 9 (see Supplemental Figure 5 online), and the energy minimized forms of these molecules possessed much more bent conformations.

All of the free acids of the benzenoid potential substrates tested (benzoic acid, salicylic acid, anthranilic acid) only existed as the fully planar conformers after energy minimization. Moreover, the most stable conformations for benzoate, anthranilate and salicylate still kept the plane of the carboxylate group close to the plane of the aromatic ring. For anthranilate and salicylate, this is easily explained by intramolecular hydrogen bonds that form between the carbonyl oxygen in the carboxylate moiety and the amino and hydroxyl groups, respectively, at the ortho position on the aromatic rings of these two molecules, maintaining very rigid planar structures. Benzoate, on the other hand, is not constrained by intramolecular hydrogen bonds. However, because the double bond of the carbonyl group is proximate to the aromatic ring,

significant interaction between the pi bonds of the ring and of the carbonyl group would be expected, and would stabilize the planar conformer.

Docking Experiments Explain Differences in Substrate Preferences of CCMTs and SAMTs

As briefly outlined in the Results section, the energy minimized structures of potential substrates were used in docking experiments by GOLD version 3.1.1 (Cambridge Crystallographic Data Centre) using the modeled structures of CCMT1, CCMT2, CCMT3 proteins and the CCMT1 H160M mutant, as well as AmSAMT native protein and M163H mutant. In these docking experiments, active site residues were allowed to have flexible side chains, and various hydrogen bond constraints were applied. When flexible side chains were not allowed, no reasonable docking results could be obtained (i.e., all substrates were docked by GOLD to regions of the active site in conformations that would make methyl transfer from AdoMet impossible). However, when active site residue side chains (presence and position in active site based on Modeller results) were allowed to be flexible in the docking runs, the substrates (see Table 1) for particular CCMT isoforms or for SAMT or the mutant proteins were docked by GOLD in their respective enzymes in conformations that would allow methyl transfer from AdoMet. On the other hand, non-substrate compounds did not dock in a manner that would allow for favorable methyl transfer in particular enzymes. These results suggest that although the modeled CCMT structures may not be completely accurate representations of the structures of the proteins in solution, they represent good working models to generate hypotheses regarding which residues in the active site might be involved in substrate preference determination.

In these docking experiments, for example, *p*-coumarate formed three hydrogen bonds with active site residues of CCMT1: one oxygen of the carboxylate was hydrogen bonded to the

epsilon N of His-160, the other carboxylate oxygen was hydrogen bonded to the amide N of Gln-36, and the phenolic hydroxyl group hydrogen bonded to the hydroxyl sidechain of Ser-232. This was an interesting finding, because His-160 replaced Trp-161 as the residue involved in carboxylate hydrogen bonding, compared to what was proposed for CbSAMT. In the models of CCMT1, Trp-161 is too far away from the carboxyl group and Gln-36 to form a strong hydrogen bond. CCMT1 preferred *p*-coumarate as substrate and appeared to bind the carboxylate form preferentially. Docking of pCA and CA as the protonated free acids led to apparent restricted flexibility of the substrate, leading to a more rigid molecule trying to fit into the active site of the CCMT1 enzyme. Nevertheless, pCA still formed three hydrogen bonds, with His-160, Gln-36 and Ser-232. However, in order for the carboxyl group to attack the methyl on AdoMet, which is required for methyl transfer, it (the carboxyl group) must be present in the deprotonated form. Similar results were found for the other CCMT isoforms. These results all suggest that binding of the deprotonated forms of the substrates is required for catalysis to occur, and that the enzymes preferentially bind these forms.

One of the most interesting of the docking experiment results was the docking of salicylate to the CCMT isoforms. As for CA and pCA, SA would need to bind to CCMT as the carboxylate (salicylate) in order for the carboxyl group to attack the methyl of AdoMet. In this state, binding of SA to CCMT1 leads apparently to a non-productive complex, where the epsilon N of His-160 forms a hydrogen bond to the 2-hydroxyl group of SA and the amide N of Gln-36 forms a strong H-bond with the carboxylate group of SA. This effectively moves SA's carboxyl group far from the reactive methyl of AdoMet, thus preventing methyl transfer and readily explaining why SA is not a substrate for CCMT1. Indeed, based on these results we expected SA to be a good competitive inhibitor of CCMT1, as outlined in the Results section.

Because of these results, we evaluated the activity of CCMT1-H160M with the same battery of substrates shown in Table 1. Although SA was not the best substrate for this mutant enzyme, it was a very good substrate, with 61% specific activity compared to CA. When SA was docked into CCMT1-H160M, the 2-hydroxyl group formed a hydrogen bond to the backbone carbonyl oxygen of Phe-157, stabilizing the molecule in a position and conformation where the carboxyl group, stabilized by hydrogen bonds from the amide N of Gln-36 and the epsilon N of Trp-161, was placed in a very good position for attack on the reactive methyl of AdoMet. Furthermore, the best substrate for CCMT1-H160M turned out to be BA, with 162% of the activity of CA with this enzyme. Docking of BA into this mutated form of the enzyme demonstrated that the carboxylate group was held by two hydrogen bonds (to the epsilon N of Trp-161 and the amide N of Gln-36) in an ideal position and orientation for attack on the reactive methyl of AdoMet. In contrast, docking of BA into CCMT1 showed formation of hydrogen bonds to Gln-36 and His-160 in such a way that the carboxyl group of BA was relatively far from AdoMet's reactive methyl. This supported the observation that BA was a much poorer substrate for CCMT1 than was cinnamate.

Anthranilate was only a substrate for CCMT1-H160M, and not for any of the non-mutagenized CCMT enzymes. Anthranilate bound in docking experiments to the active site of CCMT1 in two orientations with similar GOLD fitness scores (33.798 vs 34.110). The docking with the slightly lower fitness score provided two hydrogen bonds for anthranilate to CCMT1 (carboxylate carbonyl O to His-160 epsilon N and amino N to amide N of Gln-36), but positioned the reactive carboxylate O in an unfavorable position for methyl attack. In contrast, the slightly more favorable docking provided stronger hydrogen bonds (carboxylate O to His-160 epsilon N and amide N of Gln-36) to stabilize binding, but which positioned the reactive

carboxylate O even further from AdoMet's reactive methyl, suggesting why this compound is not a substrate for native CCMTs. On the other hand, docking experiments with CCMT1-H160M suggested that anthranilate formed two hydrogen bonds (amino N to backbone carbonyl O of Phe-157 and carboxyl carbonyl O to amide N of Gln-36), which placed the reactive carboxylate O in a favorable position and orientation for attack on the reactive methyl of AdoMet. This binding was further stabilized by Met-160 and Leu-325 forming hydrophobic interactions with the aromatic ring of anthranilate.

Docking experiments with *p*-hydroxyhydrocinnamate (pHydCA) demonstrated that this compound was able to bind very favorably to the active site of CCMT1, with its carboxylate O and phenolic O, respectively, forming hydrogen bonds with the amino N of Gln-36 and with Thr-326's backbone N and sidechain O along with Cys-322's backbone carbonyl O. This binding was possible because pHydCA is able to take on a rather bent conformation, due to its having a reduced sidechain. In the mutant enzyme, CCMT1-H160M, however, it binds such that the carboxylate O is in a favorable orientation but at a rather distant position for attack on the reactive methyl of AdoMet. This may explain the dramatic decrease (25-fold reduction in activity) of this substrate with CCMT1-H160M vs. CCMT1 (see Tables 1 and 2). Hydrocinnamate (HydCA), in contrast, was able to bind in the docking experiments of both CCMT1-H160M and CCMT1 in a position and orientation that allowed attack on AdoMet's reactive methyl. HydCA formed hydrogen bonds from its carboxylate carbonyl O to the S of Cys-322 and the amide N of Gln-36, placing the other carboxylate O in an orientation, albeit at a slightly unfavorable distance, for attack on AdoMet's reactive methyl. These results also support the observed reactivity of this compound with the enzymes tested (see Table 1).

The substrate preference of the mutant CCMT1-H160M was broader than its progenitors. Activity was observed with anthranilate and salicylate in addition to the original set of substrates accepted by CCMT1. Activity with CA was higher by factors of 2.7 and 20 relative to pCA and BA, respectively, for CCMT1-H160M. Because CCMT1 lost 90% of its activity towards CA with the single amino acid change, it is clear that its apparent increase in activity towards benzoate is not only a reflection of a decrease in activity towards CA, but also due to an increased turnover rate with BA. This is also evidenced by the fact that SA is accepted by the mutant enzyme at greater than 60% of the activity with CA. The novel activities towards anthranilate and SA reflect a change in the geometry and composition of the substrate pocket that now allows for functional groups to be present on the phenyl ring, but only in the *ortho* position, and only when the propenyl side chain is shortened. It is notable that the substitution of the *o*-hydroxyl of SA with an amino group reduces activity by nearly 70%, likely due to the increased bulkiness of the amino versus the hydroxyl moiety.

Phylogenetic Analysis

The full length sequences were analyzed to confirm their identity as SABATH homologs by submitting them in searches against the conserved domain database (CDD) of NCBI. CbSAMT protein has a domain architecture consisting of a single domain, Methyltransf_7 ((Position-Specific Scoring Matrix) PSSM-Id: 43418) (Pfam 03492.10), spanning most of the length of the amino acid sequence, which is a conserved and defining feature for the characterized members of the SABATH gene family. RPS-BLAST searches of the full length non-plant sequences revealed that 19 had significant homology spanning most of the domain, with e-values ranging 4e-29 to

3e-12. Twelve additional sequences had similarly significant homology that spanned a smaller portion of the total length of the domain (35-70 % aligned). Two sequences retrieved by PSI-BLAST representing bacterial and fungal species (*Geobacter* and *Cryptococcus*) not otherwise represented were eliminated from further consideration after this analysis. A sequence representing a third species of cyanobacterium, *Gloeobacter*, in addition to *Prochlorococcus* and *Synechococcus*, had a long multidomain sequence consisting of the methyltransfase domain, which has strong homology to the SABATH PSSM, conjugated with a Cfa domain (cyclopropane fatty acid synthase) (PssmID 32411) located just upstream of the SABATH domain. The results of this additional analysis and the significance of the homology of these sequences to the SABATH PSSM provides a strong level of confidence in the identity of these non-plant sequences as potential new members of the SABATH gene family, which until now was thought to be plant-specific.

All identified SABATH family sequences were assembled into a multiple sequence alignment using ClustalX with manual refinement. Longer partial sequence fragments were retained throughout this process to permit evaluation of the consensus of residues of the family as a whole as well as individual sequence subgroups and to evaluate conservation both of active site residues identified from the solved structure of CbSAMT and mutagenesis studies of characterized family members and also to permit identification of strongly conserved residues or motifs not previously implicated as functional active site residues or as residues otherwise conserved as essential to the structure and function of the SABATH gene family proteins. The global alignment was subsequently edited manually to remove regions of high sequence variability as well as partial length sequences that were missing sequence data spanning the included regions of the alignment. The resulting alignment (Supplemental Figure 3 online) was

used to generate a maximum likelihood tree (Figure 9) to permit analysis of the relationships of the sequences of the new SABATH family member genes.