Published in final edited form as:

Biochemistry. 2016 November 15; 55(45): 6314–6326. doi:10.1021/acs.biochem.6b00887.

Dynamic conformational states dictate selectivity toward native substrate in a substrate-permissive acyltransferase

Olesya Levsh^{1,2}, Ying-Chih Chiang³, Chun Fai Tung³, Joseph P. Noel⁴, Yi Wang^{3,*}, and Jing-Ke Weng^{1,2,*}

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

³Department of Physics, The Chinese University of Hong Kong, Shatin, Hong Kong

⁴Howard Hughes Medical Institute, Jack H. Skirball Center for Chemical Biology & Proteomics, The Salk Institute for Biological Studies, La Jolla, CA

Abstract

Hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT) is an essential acyltransferase that mediates flux through plant phenylpropanoid metabolism by catalyzing a reaction between *p*-coumaroyl CoA and shikimate, yet it also exhibits broad substrate permissiveness *in vitro*. How do enzymes like HCT avoid functional derailment by cellular metabolites that qualify as non-native substrates? Here, we combine X-ray crystallography and molecular dynamics to reveal distinct dynamic modes of HCT under native versus non-native catalysis. We find that essential electrostatic and hydrogen-bonding interactions between the ligand and active site residues, enabled by active site plasticity, are elicited more effectively by shikimate than by other non-native substrates. This work provides a structural basis for how dynamic conformational states of HCT favor native over non-native catalysis by reducing the number of futile encounters between the enzyme and shikimate.

GRAPHICAL ABSTRACT

*To whom correspondence should be addressed. yiwang@phy.cuhk.edu.hk and wengj@wi.mit.edu.

AUTHOR CONTRIBUTION

O.L. and J.K.W. designed the research; O.L., J.P.N., J.K.W. collected X-ray diffraction data and solved crystal structures. Y.C.C., C.F.T., and Y.W. performed MD simulations. O.L. performed mutagenesis and kinetic assays. O.L., Y.C.C., Y.W., and J.K.W. interpreted the results and wrote the paper.

Keywords

enzyme promiscuity; enzyme dynamics; phenylpropanoid metabolism; BAHD acyltransferase; HCT; catalytic mechanism

Enzyme promiscuity, which can be more precisely described as substrate permissiveness, mechanistic elasticity, and concomitant product diversity, has long been regarded as a facilitator in the evolution of novelty¹. In 1976, Jensen first theorized that primordial enzymes were generalists capable of multiple catalytic tasks and later evolved into more specialized catalysts². Since then, the theory of enzyme recruitment, which posits that enzyme promiscuity provides a starting point for positive selection and leads to neofunctionalization via cycles of duplication and selection, has garnered experimental support^{3, 4}. Generally, mutations that are neutral or nearly neutral to the primary function of the ancestral enzyme may alter its non-native functions, rendering new “raw materials” for selection to act upon, as long as they are not selected against^{5, 6}. Furthermore, the catalytic properties of discrete enzymes are subjected to selection in the context of an *in vivo* metabolic system to maximize organismal fitness – catalytic machinery must be efficient enough to produce essential molecules when they are necessary, but not so efficient that it drains vital metabolite pools. As a result, natural enzymes – even those from the same enzyme family – exhibit kinetic parameters (e.g. K_m and k_{cat} determined against native substrates) that can span orders of magnitude but in general agree with the measured metabolite pools and fluxes *in vivo*^{7, 8}. Previous studies also showed that catalytic specificity and efficiency of many natural enzymes could be further improved through directed evolution^{9–11}. Together, these observations suggest that systems-level constraints imposed on natural enzymes might have prevented them from evolving toward the highest specificity and efficiency limits set by physicochemical and thermodynamic constraints, thereby setting the stage for substrate permissiveness. However, it remains largely unknown how enzymes with relaxed substrate specificity maintain their primary catalytic functions in the face of competing non-native substrates widely present *in vivo*.

Plant specialized metabolism represents a rich network of rapidly evolving metabolic enzymes. Evolutionary exploitation of ancestral enzyme promiscuity has allowed new metabolic traits to emerge in diverse taxa of the plant kingdom¹². The BAHD family of acyltransferases, named after the first four enzymes discovered from this family (BEAT, AHCT, HCBT, and DAT), are an important family of plant specialized enzymes responsible for the biosynthesis of a huge diversity of ester- and amide-containing natural products¹³. The BAHD family has radiated extensively throughout land plant evolution, with a small number of BAHD genes contained in the basal bryophytes, and the majority of BAHD homologs arising independently within major lineages of vascular plants^{14, 15}. Members of the BAHD acyltransferase family can be identified by sequence homology and the universally conserved HXXXD and DFGWG motifs, although the biochemical functions of individual BAHD acyltransferases remain largely uncharacterized¹⁶.

Hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT) is one of the few BAHD acyltransferases that exhibits an orthologous phylogenetic pattern across all land

plants (Supplementary Figure S1)¹³. HCT transfers the *p*-coumaroyl group from the acyl donor *p*-coumaroyl CoA onto the 5-hydroxyl of the acyl acceptor shikimate to produce *p*-coumaroylshikimate, which in turn serves as a key intermediate for the first aromatic ring *meta*-hydroxylation step in the plant phenylpropanoid metabolism¹⁷. Important phenylpropanoids downstream of HCT include monolignols, building blocks of the polymer lignin in vascular plants, as well as a plethora of pigments, flavor compounds, UV protectants, and phytoalexins¹⁸. Since shikimate is a precursor for the biosynthesis of aromatic amino acids in plants, it has also been proposed that the involvement of HCT in phenylpropanoid metabolism implements a metabolic flux control in the pathway that is responsive to the availability of cytosolic shikimate¹⁹.

Despite its essential role in plant phenylpropanoid metabolism, HCT displays inferior kinetic parameters. Previous studies reported a K_m of 323 μ M measured using *Coleus blumei* HCT (CbHCT) against its native acyl acceptor shikimate, a value one or two orders of magnitude greater than that of some of the other BAHD acyltransferases characterized previously^{20–23}. Moreover, CbHCT can utilize a diverse array of non-native acyl donor and acceptor substrates to produce a variety of esters and amides *in vitro*²⁴. Generally, substrate-permissive enzymes have been considered to be more evolvable than highly specific enzymes^{5, 25}. Indeed, rosmarinic acid synthase (RAS), a BAHD associated with the Lamiaceae family, has apparently neofunctionalized from HCT after an ancestral gene duplication event^{24, 26} (Supplementary Figure S1). Rather than using shikimate as in HCT, RAS utilizes an alternative acyl acceptor, 4-hydroxyphenyllactate, and produces *p*-coumaroyl-4'-hydroxyphenyllactate, a key precursor for the specialized rosmarinic acid biosynthesis²⁶.

We were intrigued by the fact that HCT, as an essential enzyme in plant phenylpropanoid metabolism, exhibits low substrate specificity. How does HCT maintain its specific metabolic function *in vivo* in the presence of other reactive metabolites that are able to access its active site and derail native catalysis? By combining approaches such as structural biology, site-directed mutagenesis, enzyme kinetics, and molecular dynamics (MD) simulation, we sought to gain a deeper understanding of the molecular interactions between HCT and its native substrate in addition to a range of non-native substrates. Based on our findings, we propose a model where specific conformational substates of the enzyme support efficient shikimate turnover but do not facilitate productive catalytic encounters with a range of non-native substrates.

EXPERIMENTAL PROCEDURES

Cloning and Site-directed Mutagenesis of HCTs

The open reading frame (ORF) of AtHCT was PCR amplified from Arabidopsis cDNA, digested with Nco I and Xho I, and ligated into Nco I and Xho I-digested pHIS8–3 *E. coli* expression vector. The ORF of CbHCT (GenBank accession version CBI83579.1) was codon optimized for expression in *E. coli*, and synthesized as a gBlocks® Gene Fragment (Supplementary Data S2). The fragment was cloned into Nco I-digested pHIS8–4 *E. coli* expression vector using Gibson Assembly. Site-directed mutagenesis was performed according to the QuikChange II Site-Directed Mutagenesis protocol (Agilent Technologies).

Recombinant Protein Expression and Purification

Protein expression constructs were transformed into *E. coli* BL21(DE3) strain and purified with affinity followed by size-exclusion chromatography, as previously described²⁷. Detailed procedures of the protein purification protocol are provided in the Supplementary Experimental Procedures.

Protein Crystallization

AtHCT crystals were grown by hanging drop vapor diffusion at 4 °C by mixing 2 µL 10 mg/mL protein with 1 µL of a reservoir solution containing 0.5 M ammonium acetate, 0.1M MOPSO-NaOH, pH 7.24, 18% PEG 8000. CbHCT crystals were grown by hanging drop vapor diffusion at 4 °C by mixing 1 µL 15 mg/mL protein with 1 µL of a reservoir solution containing 0.1 M Bis-Tris:HCl, 25% PEG 3350, pH 7.5. For soaking AtHCT crystals, soaking drops were prepared by mixing 1.8 µL of reservoir solution with either 0.2 µL of 100 mM *p*-coumaroyl CoA in water or 0.2 µL of 100 mM *p*-coumaroylshikimate in DMSO. For soaking CbHCT crystals, a soaking drop was prepared from 1.6 µL of reservoir solution mixed with 0.2 µL of 500 mM 3-hydroxyacetophenone in DMSO, and 0.2 µL of 100 mM *p*-coumaroyl CoA in water. Crystals were transferred into soaking solution and incubated for 1 hour before transfer to a cryoprotection solution of 17% glycerol and reservoir solution. Apo crystals were cryoprotected without being soaked. Single crystals were mounted in a cryoloop and flash frozen in liquid nitrogen.

X-ray Diffraction and Structure Determination

X-ray diffraction data was collected at beamlines 8.2.1 and 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory on ADSC Quantum 315 CCD detectors for AtHCT crystals, and at beamline 24-ID-E of the Advanced Photon Source at Argonne National Laboratory on a CCD-based ADSC Quantum 315 detector for CbHCT crystals. Diffraction intensities were indexed and integrated with iMosflm²⁸, and scaled with Scala under CCP4^{29, 30}. The phase of the AtHCT SeMet apo structure was determined by single-wavelength anomalous diffraction (SAD). The phases for the other HCT structures were determined with molecular replacement using Phaser under Phenix³¹. Further structural refinement utilized Phenix programs³¹. Coot was used for manual map inspection and model rebuilding³². Crystallographic calculations were performed using Phenix. The Protein Data Bank accession numbers for the apo-, *p*-coumaroyl CoA-bound, and *p*-coumaroyl shikimate-bound AtHCT, and the apo- and 3-hydroxyacetophenone-bound CbHCT are 5KJS, 5KJT, 5KJU, 5KJV, and 5KJW, respectively.

Comparative Sequence and Structure Analysis

AtHCT, CbHCT, SbHCT, SmHCT, PpHCT, and CbRAS protein sequences were derived from NCBI³³. The PaHCT sequence was derived from ConGenIE³⁴. The remaining sequences were derived from 1KP³⁵. In all cases, AtHCT was used as the search query. Amino acid alignment of HCT orthologs was created using T-COFFEE with default settings³⁶. ESPript 3.0³⁷ was used to display the multiple sequence alignments as shown in Figure 2F and Supplementary Data S1. Atomic coordinates of apo (PDB ID: 4KE4) and *p*-coumaroylshikimate and CoA-bound (PDB ID: 4KEC) SbHCT were retrieved from the

Protein Data Bank³⁸. All structures were aligned based on topology with the SSM superpose function in Coot^{32, 39}. KVFinder⁴⁰ was used to calculate enzyme active site volumes (additional details can be found in Supplementary Experimental Procedures). All structural figures were created with the PyMOL Molecular Graphics System, version 1.3 (Schrödinger, LLC)⁴¹.

Phylogenetic analysis

Evolutionary analyses were conducted in MEGA7⁴². A multiple sequence alignment was built using MUSCLE with the UPGMB clustering method. The analysis involved 38 amino acid sequences. All positions with less than 95% site coverage were eliminated. There were a total of 421 residue positions in the final dataset. Eight iterations were performed with a gap open penalty of -2.9 and a gap extend penalty of 0⁴³. The phylogenetic tree was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model⁴⁴. The tree with the highest log likelihood (-12321.8762) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 0.6543)].

Enzyme Kinetics

A coupled assay was used for kinetic analysis. 917 nM of *Arabidopsis thaliana* 4CL1 (NCBI accession version NP_175579.1) in 50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 2 mM ATP, 2 mM *p*-coumaric acid, and 3 mM coenzyme A was allowed to equilibrate for 24 hours to generate *p*-coumaroyl CoA at ~ 2 mM final concentration. 75 µL of this mastermix was aliquoted into Eppendorf tubes, and the assay is initiated by addition of 5 µL of substrate followed by 20 µL of enzyme. For shikimate kinetics, 21.22 nM of wild-type CbHCT, and 64 µM of R356A, R356D, and R356E were assayed against 10, 50, 100, 250, 500, 1000, 2500, and 5000 µM of shikimate. The wild-type HCT reaction was timed for 5 minutes, and the mutant reactions for 15 minutes. For 3,4-dihydroxybenzylamine kinetics, 4.24 µM of wild-type CbHCT, 8.5 µM of R356A, 1.49 µM of R356D, and 4.25 µM of R356E were assayed against 1, 2.5, 5, 10, 25, 50, 75, and 100 mM of 3,4-dihydroxybenzylamine for 5 minutes. For dopamine kinetics, 21.2 nM of wild-type and mutant CbHCT enzymes were assayed against 1, 2.5, 5, 10, 25, 50, 75, and 100 mM of dopamine for 15 minutes. Absolute quantification of initial rates was achieved by normalizing with a standard curve, obtained by running reactions with small amounts of substrate to completion. All velocities measured were in initial rate conditions. All reactions were quenched by addition of 10 µL of 10 M acetic acid. The assays were centrifuged, and analyzed directly by liquid chromatography-mass spectrometry (LC-MS). Detailed procedures of the LC-MS analysis are provided in the Supplementary Experimental Procedures. K_m and V_{max} were determined by fitting raw data to the Michaelis-Menten equation using nonlinear regression in Prism, version 6.0f (GraphPad Software).

Molecular Dynamics

MD simulation systems were constructed by solvating the protein (or protein-substrate complex) in a water box with at least 12.5 Å water padding on each side. Following

neutralization, a concentration of 0.15 M NaCl was added using the autoionize plugin of VMD⁴⁵. All simulations were performed using NAMD 2.10⁴⁶ with the CHARMM36 and CHARMM General force field^{47–50}. For each system, three or five replicas of 100-ns production runs were performed (see Table S2). For more details on the MD simulation and analysis protocols, please see the Supplementary Experimental Procedures.

RESULTS

Structural features and catalytic mechanism of *Arabidopsis thaliana* HCT

To shed light on the structural basis for substrate recognition and the catalytic mechanism of HCT, we determined several crystal structures of *Arabidopsis thaliana* HCT (AtHCT) in its apo, *p*-coumaroyl CoA-bound, and *p*-coumaroylshikimate-bound forms (Supplementary Table S1). Similar to several previously reported crystal structures of BAHD acyltransferases^{38, 51, 52}, AtHCT consists of two quasi-symmetric N-terminal (residues 1–171, 374–394) and C-terminal (residues 223–373, 395–431) domains, connected by a long loop (residues 172 to 222) (Figure 1A). Each of the N- and C-terminal domains features a β -sheet core flanked by α -helices with similar spatial arrangement, suggesting the BAHD acyltransferase fold might have evolved from an ancient domain duplication event, although no significant sequence similarity could be detected between the amino acid sequences of the two domains. The active site is situated at the interface of the domains.

Close examination of the *p*-coumaroylshikimate-bound AtHCT structure reveals several structural features underlying the catalytic mechanism of HCT (Figure 1A, B)^{38, 52}. For example, the τ -nitrogen of the universally conserved catalytic His153, provided by the N-terminal domain, forms a hydrogen bond with the shikimate 5-oxygen of *p*-coumaroylshikimate. This configuration is consistent with the role of His153 as a general base that deprotonates the 5-hydroxyl of the acyl acceptor substrate shikimate, priming it for nucleophilic attack on the carbonyl carbon of the acyl donor *p*-coumaroyl CoA. Furthermore, the indolic nitrogen of Trp371 from the C-terminal domain is within hydrogen bonding distance of the carbonyl oxygen of *p*-coumaroyl CoA, consistent with its proposed role as an oxyanion hole that stabilizes the negative charge on the tetrahedral intermediate formed after the nucleophilic attack. In the final step of the catalytic cycle, coenzyme A is released from the tetrahedral intermediate as a leaving group to produce the ester product *p*-coumaroylshikimate. Additional residues in AtHCT that contribute to *p*-coumaroylshikimate recognition include Thr369 and Arg356, forming a hydrogen bond with the 3-hydroxyl and a salt bridge with the carboxyl of the shikimate moiety, respectively.

Distinct active site conformational states in HCTs

When comparing the apo, *p*-coumaroyl CoA-bound, and *p*-coumaroylshikimate-bound AtHCT structures, we noticed several distinct active site conformational changes between apo and holo AtHCT structures (Figure 2). Consistent conformational changes in the same active site regions were also observed in a comparison of the previously reported apo and *p*-coumaroylshikimate and CoA-bound ternary *Sorghum bicolor* HCT (SbHCT) structures³⁸ (Figure 2). For ease of reference, all residues are numbered according to their relative position in the AtHCT sequence hereafter.

For example, there is a switch-like conformational shift in the catalytic His153. In structures where *p*-coumaroylshikimate is present, the imidazole side chain of His153 is stabilized in a 180° rotation relative to its apo conformation, a position that facilitates hydrogen bond formation with the shikimate 5-oxygen. Interestingly, in the *p*-coumaroyl CoA-bound AtHCT structure, His153 adopts an intermediate conformation with a side chain rotation of about 90° relative to its apo state (Figure 2A). Furthermore, in the presence of *p*-coumaroylshikimate, the side chain of Arg356 is stabilized in a conformation inside of the active site, compared to its exterior position in the apo structure. This transition is likely mediated by electrostatic attraction between the positively charged guanidinium group of the arginine side chain and the negatively charged carboxyl group of shikimate, and facilitated by the flexibility of the surrounding peptide backbone (Figure 2A-B). In addition, we observed that three loops surrounding the active site, namely L1 (residues 31–34), L2 (residues 357–365), and L3 (residues 392–398), shift inward upon binding of various ligands, resulting in shrinkage of the HCT active site (Figure 2B). A volume calculation of the acyl acceptor binding pocket (referred to as active site volume hereafter) reveals 28% and 53% shrinkage in two AtHCT holo structures in complex with *p*-coumaroyl CoA and *p*-coumaroylshikimate respectively, relative to the AtHCT apo structure, and 39% shrinkage in the SbHCT ternary structure in complex with *p*-coumaroylshikimate and CoA, relative to the SbHCT apo structure (Table 1)⁴⁰. Altogether, these observations highlight the dynamic nature of HCT active site, and implicate the functional importance of specific conformational changes in catalysis.

MD simulations of AtHCT structures corroborate our observations from comparative structural analysis. Five replicas of 100-ns simulations, initiated from the corresponding crystal structures, were performed for AtHCT in the apo state as well as in complex with *p*-coumaroyl CoA and shikimate (Supplementary Table S2). All simulations revealed considerable fluctuations in protein side chains and active site loops, which was reflected in the high variability of the active site volume. For this reason, analysis of structural elements directly involved in catalysis, such as Arg356 and His153, yielded more valuable information. The overall trend of the simulated dynamics underlying these residues is consistent with that derived from our structural analysis. For instance, compared with simulations of apo AtHCT, where a broad conformational distribution of Arg356 is observed, simulations of *p*-coumaroyl CoA and shikimate-bound AtHCT reveal that this residue is stabilized within an optimal distance of the native substrate (Figure 2C). A similar trend is observed for His153 (Supplementary Figure S2). Additionally, clustering analysis reveals that once AtHCT is bound to its substrates, Arg356 adopts a predominantly internal conformation (Figure 2D). In contrast, the apo simulations are dominated by external conformations, including one in which the residue is oriented opposite to the active site (Figure 2E, Supplementary Figure S3). The much larger conformational space sampled by Arg356 in the latter simulations reflects its considerable flexibility in the apo state, whereas simulations of the ternary complex indicate that binding native substrates has the effect of stabilizing this key active site residue in the preferred conformation. Notably, the apo simulations also reveal Arg356 conformations resembling the internal state (Supplementary Figure S3), suggesting that the swinging motion of this residue constitutes a rapid conformational shift that can be utilized by the enzyme to facilitate its native reaction.

An arginine handle dictates acyl acceptor specificity in HCT

As an indispensable enzyme in plant phenylpropanoid metabolism, HCT is conserved across all extant land plants¹³. To infer the evolutionary history and conserved sequence features of HCTs, we performed multiple sequence alignment (MSA) and phylogenetic analysis using 35 HCT ortholog sequences collected from various sequence databases, representing all major land plant lineages (Figure 2F, Supplementary Figure S1, and Supplementary Data S1). The MSA reveals universal conservation of Arg356 as well as a high degree of conservation of the surrounding residues (354–362), indicating that Arg356 is likely essential for the catalytic function of HCT and the particular combination of its neighboring residues as part of L2 may confer its structural flexibility (Figure 2F). Other universally conserved residues among HCTs include Thr369 and Trp371, both of which play important roles throughout the catalytic cycle of HCT (Figure 1B and 2F). Interestingly, *Coleus blumei* RAS (CbRAS), a neofunctionalized enzyme derived from HCT^{24, 53}, retains neither Arg356 nor Thr369 (Figure 2F). These substitutions likely reflect the shift in RAS specificity toward a new acyl acceptor.

The presence of a specific and highly conserved electrostatic interaction between the structurally flexible Arg356 and shikimate led us to hypothesize that Arg356 may act as a catalytic “handle” that affixes to shikimate by its carboxyl and orients the shikimate 5-hydroxyl toward the catalytic center via a specific arginine swing-in movement during catalysis. Several conformational states of the Arg356 handle were indeed captured by various HCT structures as well as MD simulations (Figure 2). Based on this hypothesis, we also postulated that it might be possible to strategically engineer HCT to favor certain positively charged acyl acceptor substrates if the arginine handle is substituted with a negatively charged residue.

To test this hypothesis, we carried out a series of mutagenesis experiments targeting Arg356 followed by enzyme kinetic assays. CbHCT was chosen for these experiments because it was the most soluble and highly expressed enzyme among several HCT orthologs we produced recombinantly in *E. coli*. Wild-type CbHCT and its R356D, R356E, and R356A mutants (corresponding to residue 350 in the CbHCT sequence) were produced and purified for biochemical assays. Notably, introducing these mutations to CbHCT did not appear to alter the production, solubility, and stability of the recombinant protein when expressed in *E. coli* (Supplementary Figure S4).

Pseudo-first-order kinetic assays of wild-type CbHCT against shikimate reveal a K_m of 388 μM and a k_{cat} of 30.1 s^{-1} , consistent with previously published kinetic parameters on the same enzyme²⁴. The catalytic activity toward shikimate was entirely abolished in the R356D, R356E and R356A mutants, demonstrating that Arg356 plays an essential role in the native reaction catalyzed by HCT (Figure 3A and Table 2). We posited that the role of Arg356 is two-fold. First, the electrostatic attraction between the positively charged Arg356 side chain and negatively charged shikimate molecule facilitates binding of the acyl acceptor substrate into the active site. Second, the nature of the attraction between Arg356 and shikimate is geometrically unique, and thus orients the substrate’s functional groups properly relative to the enzyme’s catalytic machinery, thereby facilitating efficient catalysis. We then investigated the role of Arg356 in catalytic efficiency via MD simulations of wild-

type CbHCT and the CbHCT R356E mutant, both in complex with *p*-coumaroyl CoA and shikimate (Supplementary Table S2). In simulations of wild-type CbHCT, Arg356 is found to anchor shikimate in orientations that facilitate nucleophilic attack by its 3-hydroxyl onto *p*-coumaroyl CoA (Figure 4A). By contrast, in the R356E mutant simulations, shikimate quickly lost such preferred orientations and explored a diversity of positions in the active site (Figure 4B). In all three 100-ns simulations of the R356E mutant, shikimate left the active site toward the end of the trajectories.

Next, we characterized the kinetics of each CbHCT mutant toward non-native acyl acceptor substrates. We chose non-native acyl acceptors that resemble shikimate but contain a positively charged amine group in place of the carboxylic acid moiety. Commercial availability limited our substrate selection to 3,4-dihydroxybenzylamine and dopamine, which contain an aromatic ring instead of the non-aromatic six-membered ring as in shikimate (Figure 3B-C). Our hypothesis predicts that R356D and R356E mutants should attract the positively charged non-native substrates 3,4-dihydroxybenzylamine and dopamine, while wild-type CbHCT should repel these substrates. As R356A is unable to engage any charge-based interaction, it may be expected to display an intermediate catalytic activity. The kinetics of the three CbHCT mutants challenged with 3,4-dihydroxybenzylamine revealed precisely this pattern (Figure 3B and Table 2). Whereas the mutant enzymes are incapable of catalyzing the native HCT reaction, their catalytic performance in utilizing 3,4-dihydroxybenzylamine surpasses that of wild-type CbHCT. R356D is nearly an order of magnitude more efficient than wild-type CbHCT in terms of k_{cat}/K_m . Notably, this is primarily due to a significantly lower K_m in R356D compared to wild-type CbHCT. R356E has a compromised catalytic rate, but demonstrates an improvement in K_m to the extent that it remains five-fold more efficient than wild-type CbHCT. Dopamine is less sterically similar to shikimate than 3,4-dihydroxybenzylamine is, and this property is reflected in the generally unfavorable kinetic parameters measured for this substrate. Neither V_{max} nor K_m could be determined for wild-type CbHCT toward this acyl acceptor, as the reaction velocity increased linearly even at 100 mM substrate. Nevertheless, the CbHCT mutants outperformed the wild-type enzyme in catalyzing this non-native reaction (Figure 3C and Table 2). The differences in magnitude for catalytic efficiency (k_{cat}/K_m) among the three mutants are less pronounced for dopamine than for 3,4-dihydroxybenzylamine, but R356D and R356E are slightly more efficient than R356A. Consistent with the 3,4-dihydroxybenzylamine results, this improvement in catalytic efficiency is contributed entirely from K_m , as R356D and R356E indeed display lower k_{cat} toward dopamine compared to R356A.

A neutral non-native substrate does not induce conformational change in HCT active site

Our kinetic experiments and MD simulations informed us of the necessity of the Arg356 handle in the native reaction catalyzed by HCT involving shikimate, and the sufficiency of the handle's charge identity to dictate substrate specificity toward charged acyl acceptor molecules. Because electrostatic interactions appear to have a large role in substrate recognition in HCT, we sought to characterize the nature of the interaction between HCT and neutral non-native substrates. We chose 3-hydroxyacetophenone, as it resembles shikimate but lacks an ionizable functional group (Figure 5A). Moreover, we identified 3-

hydroxyacetophenone as an HCT substrate in a preliminary screen designed to reveal HCT-mediated catalysis of non-native acyl acceptors (Supplementary Figure S5). Notably, the same assay also identified the previously reported 2,3-dihydroxybenzoic acid as a non-native acyl acceptor (Supplementary Figure S5)²⁴. However, all of the CbHCT variants displayed such low activity toward 3-hydroxyacetophenone that kinetic characterization was infeasible.

To shed light on the molecular basis underlying HCT-mediated catalysis of a neutral non-native substrate, we obtained crystal structures of wild-type CbHCT in apo form, as well as the ternary complex with *p*-coumaroyl CoA and 3-hydroxyacetophenone – though in the latter, electron density supporting *p*-coumaroyl CoA is scarce. The ternary structure reveals that 3-hydroxyacetophenone is bound in a hydrophobic pocket of the enzyme active site (Figure 5A). An overlay of this structure with the *p*-coumaroylshikimate-bound AtHCT structure further reveals that 3-hydroxyacetophenone is shifted significantly away from the catalytic center relative to the shikimate moiety of *p*-coumaroylshikimate (Figure 5B). Furthermore, 3-hydroxyacetophenone adopts a nonproductive orientation, in which its 3-hydroxyl points away from the catalytic center. The combination of these binding features cause the catalytic group of 3-hydroxyacetophenone to be 7.5 Å away from the carbonyl carbon of *p*-coumaroyl CoA, and 8.1 Å away from the τ -nitrogen of the catalytic histidine, too far to elicit efficient catalysis (Figure 5B). In the ternary structure, electron density supports two conformations of the catalytic histidine, similar to that of the *p*-coumaroyl CoA-bound AtHCT structure (Supplementary Figure S6). However, no other significant conformational changes are induced. The conformations of Arg356 and the active site loops resemble those of the apo HCTs more closely than those of the *p*-coumaroyl CoA or *p*-coumaroylshikimate-bound HCTs. Consequently, there is little active site shrinkage upon 3-hydroxyacetophenone binding (Figure 5C, Table 1).

To further assess the mode of interaction between CbHCT and 3-hydroxyacetophenone, we performed five 100-ns MD simulations of CbHCT in complex with *p*-coumaroyl CoA and 3-hydroxyacetophenone. Five simulations for apo CbHCT as well as *p*-coumaroyl CoA and shikimate-bound CbHCT were also carried out as controls (Supplementary Figure S2). Similar to apo AtHCT, Arg356 in apo CbHCT adopts both external and internal conformations, with the former dominating the distribution (Supplementary Figure S7). Overall, active site dynamics in simulations of apo versus *p*-coumaroyl CoA and shikimate-bound CbHCT are consistent with those observed in AtHCT (Supplementary Figure S7, Figure 5D-E). For instance, in the presence of shikimate, Arg356 predominantly adopts the internal conformation. In contrast, due to an inability to form a stable salt bridge with Arg356, 3-hydroxyacetophenone explores many conformations inside the active site and is unable to sustain the preferred orientation (Figure 4C). To comprehend whether the loss in productive binding is due to 3-hydroxyacetophenone's aromaticity, its absence of 4- and 5-hydroxyls, or its absence of the carboxylic acid moiety, we performed three 100-ns simulations of wild-type CbHCT in complex with two shikimate “mutants”. Whereas the mutant lacking the 4- and 5-hydroxyl groups sustains a similar location and orientation to shikimate, the mutant lacking the carboxyl group explores many conformations inside the active site, reflecting the necessary role of the shikimate carboxyl in facilitating productive binding of the native substrate (Supplementary Figure S8). Simulations of 3-

hydroxyacetophenone-bound CbHCT reveal that Arg356 primarily adopts the external conformation, consistent with the apo and 3-hydroxyacetophenone-bound CbHCT crystal structures (Figure 5F). Furthermore, compared with shikimate-bound simulations, where Arg356 is more likely to be located close to the substrate, the 3-hydroxyacetophenone-bound CbHCT simulations resemble apo-CbHCT simulations more closely due to the broader conformational distribution of this residue (Figure 5D). Altogether, these results demonstrate that a neutral non-native substrate is unable to induce active site conformational changes associated with catalysis.

DISCUSSION

Distinct dynamic conformational states are prevalent throughout HCT substrate binding and catalysis. Though the conformational differences described here are comparatively small in magnitude, they play exquisite roles in various steps within the HCT catalytic cycle. For example, the flipping motion of the imidazole ring of catalytic His153 places the imidazole τ -nitrogen 1 Å closer to the 5-hydroxyl of the acyl acceptor substrate shikimate, establishing the proximity necessary to trigger the general base-mediated catalysis. Additionally, the active site surrounding loops, which display root-mean-square deviations (RMSDs) between 1.1 Å and 3.2 Å transitioning from the apo to various native substrate and product-bound states in both AtHCT and SbHCT structures, collectively result in substantial active site shrinkage. Finally, the Arg356 handle undergoes a swing-in movement upon shikimate binding that corresponds to a shift of 1.7 Å in the alpha carbon and 3.5 Å in the guanidinium group. Consequently, a neutral ligand that is unable to induce these active site conformational changes is a poor substrate.

Our results suggest that the specific ionic interaction between the Arg356 handle and shikimate is a key structural feature that allows HCT to distinguish between native and non-native acyl acceptor substrates. The Arg356 handle contributes to multiple facets of HCT's catalytic mechanism. First, salt bridge formation between the arginine side chain and shikimate carboxyl confers acyl acceptor substrate binding affinity. Further, the specific movement of the Arg356 handle orients shikimate in a catalytically productive conformation in HCT's rather large active site, thereby facilitating efficient catalysis by increasing the fraction of productive encounters that occur between the enzyme and substrate⁵⁴. MD simulations exploring interaction of wild-type CbHCT with a shikimate "mutant" lacking the carboxyl group further corroborate the necessary role of the ionic interaction between Arg356 and the native substrate in catalytically productive binding.

The Arg356 handle is an apparently important specificity feature – present in HCT and the closely related BAHD acyltransferase hydroxycinnamoyl CoA: quinate hydroxycinnamoyl transferase (HQT), which catalyzes a reaction with quinate, an acyl acceptor that strongly resembles shikimate – but not conserved in the closely related RAS, which utilizes a structurally different acyl acceptor. It is difficult to assess whether more distant BAHD homologs that catalyze reactions with positively charged acyl acceptors use a negatively charged handle as part of the catalytic mechanism, as low sequence identity and divergent active site architecture makes it difficult to predict enzyme-ligand interactions. While mutating Arg356 resulted in a total loss of HCT's native enzymatic activity, mutating

Arg356 to negatively charged residues actually increased specificity toward certain positively charged non-native substrates, illustrating that Arg356 is both necessary and sufficient to confer HCT substrate selectivity on the level of electrostatic identity. Interestingly, our kinetic analysis revealed that R356A is unable to utilize shikimate, but accepts 3,4-dihydroxybenzylamine and dopamine as substrates. A possible explanation is that the greater hydrophobicity of the aromatic substrates makes them more likely than shikimate to diffuse into areas of the active site that are proximal to the catalytic center, leading to higher turnover in Arg356 mutants. Overall, the observation implies that the Arg356 handle is more essential in the context of shikimate catalysis than it is in the context of non-native substrate catalysis.

HCT is an essential enzyme that directs metabolic flux into a crucial branch of the plant phenylpropanoid pathway. But unlike some other essential metabolic enzymes, such as acetylcholinesterase or superoxide dismutase, which are characterized as diffusion-limited and evolutionarily perfected enzymes^{55, 56}, HCT has a few-hundred micromolar K_m , and exists in a cellular environment containing many endogenous plant metabolites with suitable functional groups to qualify as competing substrates. As an enzyme with non-ideal catalytic efficiency and low substrate specificity, how does HCT maintain specificity *in vivo* in order to maintain flux into the important downstream phenylpropanoid pathway?

An important consideration is that improved catalytic efficiency of an enzyme in isolation does not necessarily translate to improved organismal fitness. For example, if a metabolite is a common precursor for multiple metabolic pathways, each of the enzymes utilizing this metabolite would have to evolve toward optimal catalytic efficiency in order to collectively achieve balanced metabolic fluxes into the downstream pathways. The acyl acceptor substrate of HCT, shikimate, is a common precursor of aromatic amino acid biosynthesis. As many characterized BAHD acyltransferases exhibit K_m values lower than that of HCT, it may be that the seemingly low substrate binding affinity of HCT toward shikimate has evolved to be in concordance with the concentration of this metabolite in the cell, establishing a control mechanism wherein metabolic flux through HCT into the downstream phenylpropanoid pathway is sensitive to the availability of cytosolic shikimate¹⁹. For instance, HCT can act as a shikimate “sensor” by operating at a low turnover rate in conditions where shikimate is present in limiting concentrations, thereby limiting flux into the phenylpropanoid pathway¹⁹. This suggests that HCT’s kinetic parameters have evolved under systems-level constraints, reflecting the optimization of flux into different branches of organismal metabolism^{22, 57}.

HCT’s other challenge is avoiding competitive inhibition *in vivo*. At first glance, it may seem that any endogenous metabolite containing hydroxyl or amine functional groups can be a potential substrate, readily entering HCT’s rather large active site and competing with shikimate for catalysis. However, our structural and functional analyses shed light on the molecular strategies that HCT employs to maintain specificity toward shikimate over other non-native substrates. For instance, Thr369, which forms a stabilizing hydrogen bond with the shikimate 3-hydroxyl group, is adapted to accommodate the native substrate’s unique non-aromatic 6-membered ring geometry. This suggests that aromatic compounds with planar aromatic ring hydroxyls are worse substrates because they cannot simultaneously

engage Thr369 and the τ -nitrogen of the catalytic His153. Accordingly, an effective substrate should be able to engage these active site elements in a way that stabilizes catalytically productive conformations of HCT.

Although dynamic exchange between conformational states is crucial to substrate binding and catalysis in HCT, recent literature suggests that too much conformational freedom can be detrimental to both native and non-native functions. For instance, a study of negative epistasis in TEM-1 β lactamase suggests that increased structural flexibility can lead to a greater conformational diversity, potentially dominated by states that are counterproductive to substrate binding and/or catalysis⁵⁸. Additionally, a study of intermediates in the laboratory evolution of a phosphotriesterase to an arylesterase suggests that the emergence of new, productive catalytic states can occur via stabilization of pre-existing, though perhaps minor, conformational states⁵⁹. Altogether, these findings suggest that the coordination of conformational ensembles is a highly complex and specialized event unique to each enzyme, and that the impact of conformational diversity on the evolution of new functions must be weighed in the context of an enzyme's catalytic cycle. Pre-existing, catalytically competent states in an enzyme's conformational ensemble can be stabilized by non-native ligands and facilitate neofunctionalization. On the other hand, excessive dynamic freedom may translate into the prevalence of an unproductive ensemble of conformational states. Our structural and molecular dynamic observations indicate that non-native catalysis in HCT may not always proceed through the same catalytically productive conformation as shikimate catalysis, and is marked by excessive conformational freedom. This implies that such catalytic modes are feasible and inefficient because they leverage HCT's dynamic conformational diversity.

Interestingly, our analysis of MD simulation trajectories reveals that a subset of the apo AtHCT and CbHCT populations exhibit dynamic states resembling catalysis, featuring Arg356 in the internal conformation (Figure 2C and 5D). This result is in accordance with a previous study showing that dynamic states relevant to catalysis are present in the intrinsic dynamics of the prolyl cis-trans isomerase cyclophilin A in its substrate-free state⁶⁰. HCT's native substrate, shikimate, is catalyzed efficiently because it can leverage the specificity of electrostatic interactions and hydrogen bonds to achieve favorable orientation in the active site while simultaneously shifting the population equilibrium of enzyme molecules into the catalytic state. On the other hand, a plethora of compounds lacking these features will be catalyzed much less efficiently, as they must enter the active site in a proper conformation with little guidance from key surrounding residues at a time when the enzyme is in a catalytically-favorable substrate.

In summary, our study comprises the first in-depth investigation of the molecular mechanisms dictating native versus non-native catalysis in the substrate-permissive BAHD acyltransferases. As enzyme promiscuity is a framework for the evolution of new enzymatic functions, it is likely that the extensive divergence of the BAHD acyltransferase family throughout land plants was facilitated by the substrate flexibility of an ancestral enzyme or enzymes. Furthermore, as evidenced in the case of RAS neofunctionalization from its progenitor enzyme, HCT's messy biochemistry provides a starting point for the evolution of new enzymatic activities^{24, 26}. In general, the complex dynamic interplay between enzymes and their substrates, and the mutations that perturb these intricate networks, are vital

mechanisms leveraged by nature to give rise to new enzymatic functions and subsequently increase metabolic diversity^{61, 62}.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank John Ralph for providing *p*-coumaroylshikimate.

FUNDING

The X-ray crystallography was partly conducted at beamlines 8.2.1 and 8.2.2 at Advanced Light Source supported by Howard Hughes Medical Institute, and partly conducted at the Northeastern Collaborative Access Team (NE-CAT) beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector on NE-CAT 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. J.P.N. is an investigator with the Howard Hughes Medical Institute. This work was supported by the Project 409813 from University Grants Committee of Hong Kong (Y.W.), the Pew Scholar Program in the Biomedical Sciences (J.K.W.), and the Searle Scholars Program (J.K.W.).

ABBREVIATIONS

HCT	hydroxycinnamoyl
CoA	shikimate hydroxycinnamoyl transferase
RAS	rosmarinic acid synthase
MD	molecular dynamics
ORF	open reading frame
SAD	single-wavelength anomalous diffraction
LC-MS	liquid chromatography-mass spectrometry
MSA	multiple sequence alignment
RMSD	root-mean-square deviations

REFERENCES

- [1]. Weng JK, and Noel JP (2012) The remarkable pliability and promiscuity of specialized metabolism, *Cold Spring Harb. Symp. Quant. Biol.* 77, 309–320. [PubMed: 23269558]
- [2]. Jensen RA (1976) Enzyme Recruitment in Evolution of New Function, *Annu Rev Microbiol* 30, 409–425. [PubMed: 791073]
- [3]. Peisajovich SG, and Tawfik DS (2007) Protein engineers turned evolutionists, *Nat Methods* 4, 991–994. [PubMed: 18049465]
- [4]. Campbell JH, Lengyel JA, and Langridge J (1973) Evolution of a second gene for beta-galactosidase in *Escherichia coli*, *Proc Natl Acad Sci U S A* 70, 1841–1845. [PubMed: 4124306]

- [5]. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, and Tawfik DS (2005) The 'evolvability' of promiscuous protein functions, *Nat Genet* *37*, 73–76. [PubMed: 15568024]
- [6]. Harms MJ, and Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution, *Nature* *512*, 203–207.
- [7]. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, and Milo R (2011) The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters, *Biochemistry-US* *50*, 4402–4410.
- [8]. Nam H, Lewis NE, Lerman JA, Lee DH, Chang RL, Kim D, and Palsson BO (2012) Network context and selection in the evolution to enzyme specificity, *Science* *337*, 1101–1104. [PubMed: 22936779]
- [9]. Romero PA, and Arnold FH (2009) Exploring protein fitness landscapes by directed evolution, *Nat Rev Mol Cell Bio* *10*, 866–876. [PubMed: 19935669]
- [10]. Glieder A, Farinas ET, and Arnold FH (2002) Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase, *Nat Biotechnol* *20*, 1135–1139. [PubMed: 12368811]
- [11]. Aharoni A, Gaidukov L, Yagur S, Toker L, Silman I, and Tawfik DS (2004) Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization, *P Natl Acad Sci USA* *101*, 482–487.
- [12]. Weng JK, Philippe RN, and Noel JP (2012) The Rise of Chemodiversity in Plants, *Science* *336*, 1667–1670. [PubMed: 22745420]
- [13]. D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD, *Current Opinion in Plant Biology*, 331–340. [PubMed: 16616872]
- [14]. Banks JA, and Nishiyama T, and Hasebe M, and Bowman JL, and Gribskov M, and dePamphilis C, and Albert VA, and Aono N, and Aoyama T, and Ambrose BA, and Ashton NW, and Axtell MJ, and Barker E, and Barker MS, and Bennetzen JL, and Bonawitz ND, and Chapple C, and Cheng C, and Correa LG, and Dacre M, and DeBarry J, and Dreyer I, and Elias M, and Engstrom EM, and Estelle M, and Feng L, and Finet C, and Floyd SK, and Frommer WB, and Fujita T, and Gramzow L, and Gutensohn M, and Harholt J, and Hattori M, and Heyl A, and Hirai T, and Hiwatashi Y, and Ishikawa M, and Iwata M, and Karol KG, and Koehler B, and Kolukisaoglu U, and Kubo M, and Kurata T, and Lalonde S, and Li K, and Li Y, and Litt A, and Lyons E, and Manning G, and Maruyama T, and Michael TP, and Mikami K, and Miyazaki S, and Morinaga S, and Murata T, and Mueller-Roeber B, and Nelson DR, and Obara M, and Oguri Y, and Olmstead RG, and Onodera N, and Petersen BL, and Pils B, and Prigge M, and Rensing SA, and Riano-Pachon DM, and Roberts AW, and Sato Y, and Scheller HV, and Schulz B, and Schulz C, and Shakhov R, and Shibagaki N, and Shinohara N, and Shippen DE, and Sorensen I, and Sotooka R, and Sugimoto N, and Sugita M, and Sumikawa N, and Tanurdzic M, and Theissen G, and Ulvskov P, and Wakazuki S, and Weng JK, and Willats WW, and Wipf D, and Wolf PG, and Yang L, and Zimmer AD, and Zhu Q, and Mitros T, and Hellsten U, and Loque D, and Otiillar R, and Salamov A, and Schmutz J, and Shapiro H, and Lindquist E, and Lucas S, and Rokhsar D, and Grigoriev IV. (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants, *Science* *332*, 960–963. [PubMed: 21551031]
- [15]. Weng JK, and Noel JP (2013) Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants, *Front Plant Sci* *4*, 119. [PubMed: 23717312]
- [16]. St-Pierre B, and De Luca V (2000) Evolution of acyltransferase genes: Origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism, *Recent Adv Phytochem* *34*, 285–315.
- [17]. Weng JK, and Chapple C (2010) The origin and evolution of lignin biosynthesis, *New Phytol* *187*, 273–285.
- [18]. Vogt T (2010) Phenylpropanoid biosynthesis, *Mol Plant* *3*, 2–20. [PubMed: 20035037]
- [19]. Schoch GA, Morant M, Abdulrazzak N, Asnaghi C, Goepfert S, Petersen M, Ullmann P, and Werck-Reichhart D (2006) The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation, *Environ Chem Lett* *4*, 127–136.
- [20]. Chedgy RJ, Kollner TG, and Constabel CP (2015) Functional characterization of two acyltransferases from *Populus trichocarpa* capable of synthesizing benzyl benzoate and salicyl benzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis, *Phytochemistry* *113*, 149–159.

- [21]. Grothe T, Lenz R, and Kutchan TM (2001) Molecular characterization of the salutaridinol 7-O-acetyltransferase involved in morphine biosynthesis in opium poppy *Papaver somniferum*, *J Biol Chem* *276*, 30717–30723. [PubMed: 11404355]
- [22]. Weitzel C, and Petersen M (2011) Cloning and characterisation of rosmarinic acid synthase from *Melissa officinalis* L, *Phytochemistry* *72*, 572–578. [PubMed: 21354582]
- [23]. Bayer A, Ma XY, and Stockigt J (2004) Acetyltransfer in natural product biosynthesis - functional cloning and molecular analysis of vinorine synthase, *Bioorgan Med Chem* *12*, 2787–2795.
- [24]. Sander M, and Petersen M (2011) Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl-transferase, from *Coleus blumei* Benth., *Planta* *233*, 1157–1171.
- [25]. Khersonsky O, Roodveldt C, and Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects, *Curr Opin Chem Biol* *10*, 498–508. [PubMed: 16939713]
- [26]. Petersen M, and Simmonds MSJ (2003) Molecules of interest - Rosmarinic acid, *Phytochemistry* *62*, 121–125. [PubMed: 12482446]
- [27]. Weng JK, Akiyama T, Ralph J, and Chapple C (2011) Independent Recruitment of an O-Methyltransferase for Syringyl Lignin Biosynthesis in *Selaginella moellendorffii*, *Plant Cell* *23*, 2708–2724. [PubMed: 21742988]
- [28]. Battye TGG, Kontogiannis L, Johnson O, Powell HR, and Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr D* *67*, 271–281. [PubMed: 21460445]
- [29]. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, and Wilson KS (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr D* *67*, 235–242. [PubMed: 21460441]
- [30]. Evans P (2006) Scaling and assessment of data quality, *Acta Crystallogr D* *62*, 72–82.
- [31]. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, and Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D* *66*, 213–221. [PubMed: 20124702]
- [32]. Emsley P, and Cowtan K (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D* *60*, 2126–2132. [PubMed: 15572765]
- [33]. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Lanese K, Charowhas C, Clark K, DiCuccio M, Dondoshansky I, Federhen S, Feolo M, Funk K, Geer LY, Gorenkov V, Hoepfner M, Holmes B, Johnson M, Khotomlianski V, Kimchi A, Kimelman M, Kitts P, Klimke W, Krasnov S, Kuznetsov A, Landrum MJ, Landsman D, Lee JM, Lipman DJ, Lu ZY, Madden TL, Madec T, Marchler-Bauer A, Karsch-Mizrachi I, Murphy T, Orris R, Ostell J, O'Sullivan C, Panchenko A, Phan L, Preuss D, Pruitt KD, Rodarmer K, Rubinstein W, Sayers EW, Schneider V, Schuler GD, Sherry ST, Sirotkin K, Siyan K, Slotta D, Soboleva A, Soussov V, Starchenko G, Tatusova TA, Todorov K, Trawick BW, Vakarov D, Wang YL, Ward M, Wilbur WJ, Yaschenko E, Zbicz K, and Coordinators NR (2016) Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res* *44*, D7–D19. [PubMed: 26615191]
- [34]. Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin YC, Sjodin A, Van de Peer Y, Jansson S, Hvidsten TR, and Street NR (2015) The Plant Genome Integrative Explorer Resource: PlantGenIE.org, *New Phytol* *208*, 1149–1156.
- [35]. Matasci N, Hung LH, Yan ZX, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner MA, Wafula E, Der JP, dePamphilis CW, Roure B, Philippe H, Ruhfel BR, Miles NW, Graham SW, Mathews S, Surek B, Melkonian M, Soltis DE, Soltis PS, Rothfels C, Pokorny L, Shaw JA, DeGironimo L, Stevenson DW, Villarreal JC, Chen T, Kutchan TM, Rolf M, Baucom RS, Deyholos MK, Samudrala R, Tian ZJ, Wu XL, Sun X, Zhang Y, Wang J, Leebens-Mack J, and Wong GKS (2014) Data access for the 1,000 Plants (1KP) project, *Gigascience* *3*.

- [36]. Notredame C, Higgins DG, and Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J Mol Biol* *302*, 205–217.
- [37]. Robert X, and Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res* *42*, W320–W324. [PubMed: 24753421]
- [38]. Walker AM, Hayes RP, Youn B, Vermerris W, Sattler SE, and Kang C (2013) Elucidation of the structure and reaction mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other coenzyme a-dependent transferases and synthases, *Plant Physiol* *162*, 640–651. [PubMed: 23624856]
- [39]. Krissinel E, and Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D* *60*, 2256–2268. [PubMed: 15572779]
- [40]. Oliveira SH, Ferraz FA, Honorato RV, Xavier-Neto J, Sobreira TJ, and de Oliveira PS (2014) KVFinder: steered identification of protein cavities as a PyMOL plugin, *BMC Bioinformatics* *15*, 197. [PubMed: 24938294]
- [41]. Delano WL (2002) The PyMOL Molecular Graphics System.
- [42]. Kumar S, Stecher G, and Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets, *Mol Biol Evol*.
- [43]. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res* *32*, 1792–1797. [PubMed: 15034147]
- [44]. Le SQ, and Gascuel O (2008) An improved general amino acid replacement matrix, *Mol Biol Evol* *25*, 1307–1320. [PubMed: 18367465]
- [45]. Humphrey W, Dalke A, and Schulten K (1996) VMD: visual molecular dynamics, *J Mol Graph* *14*, 33–38, 27–38. [PubMed: 8744570]
- [46]. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, and Schulten K (2005) Scalable molecular dynamics with NAMD, *J Comput Chem* *26*, 1781–1802. [PubMed: 16222654]
- [47]. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, and Mackerell AD, Jr. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles, *J Chem Theory Comput* *8*, 3257–3273. [PubMed: 23341755]
- [48]. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, and MacKerell AD (2010) CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields, *Journal of Computational Chemistry* *31*, 671–690. [PubMed: 19575467]
- [49]. Yu WB, He XB, Vanommeslaeghe K, and MacKerell AD (2012) Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations, *Journal of Computational Chemistry* *33*, 2451–2468. [PubMed: 22821581]
- [50]. Vanommeslaeghe K, and MacKerell AD (2012) Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing, *J Chem Inf Model* *52*, 3144–3154. [PubMed: 23146088]
- [51]. Ma XY, Koepke J, Panjekar S, Fritzsche G, and Stockigt J (2005) Crystal structure of vinorine synthase, the first representative of the BAHD superfamily, *Journal of Biological Chemistry* *280*, 13576–13583. [PubMed: 15665331]
- [52]. Lallemand LA, Zubieta C, Lee SG, Wang YC, Acajjaoui S, Timmins J, McSweeney S, Jez JM, McCarthy JG, and McCarthy AA (2012) A Structural Basis for the Biosynthesis of the Major Chlorogenic Acids Found in Coffee, *Plant Physiology* *160*, 249–260. [PubMed: 22822210]
- [53]. Petersen M, Abdullah Y, Benner J, Eberle D, Gehlen K, Hucherig S, Janiak V, Kim KH, Sander M, Weitzel C, and Wolters S (2009) Evolution of rosmarinic acid biosynthesis, *Phytochemistry* *70*, 1663–1679. [PubMed: 19560175]
- [54]. Bar-Even A, Milo R, Noor E, and Tawfik DS (2015) The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness, *Biochemistry-Us* *54*, 4969–4977.
- [55]. Quinn DM (1987) Acetylcholinesterase - Enzyme Structure, Reaction Dynamics, and Virtual Transition-States, *Chem Rev* *87*, 955–979.

- [56]. Rotilio G, Bray RC, and Fielden EM (1972) A pulse radiolysis study of superoxide dismutase, *Biochim Biophys Acta* 268, 605–609.
- [57]. Suzuki H, Nishino T, and Nakayama T (2007) cDNA cloning of a BAHD acyltransferase from soybean (*Glycine max*): Isoflavone 7-O-glucoside-6''-O-malonyltransferase, *Phytochemistry* 68, 2035–2042. [PubMed: 17602715]
- [58]. Dellus-Gur E, Elias M, Caselli E, Prati F, Salverda ML, de Visser JA, Fraser JS, and Tawfik DS (2015) Negative Epistasis and Evolvability in TEM-1 beta-Lactamase--The Thin Line between an Enzyme's Conformational Freedom and Disorder, *J Mol Biol* 427, 2396–2409. [PubMed: 26004540]
- [59]. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, Tokuriki N, and Jackson CJ (2016) The role of protein dynamics in the evolution of new enzyme function, *Nat Chem Biol*.
- [60]. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, and Kern D (2005) Intrinsic dynamics of an enzyme underlies catalysis, *Nature* 438, 117–121.
- [61]. Clarkson MW, and Lee AL (2004) Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c, *Biochemistry-Us* 43, 12448–12458.
- [62]. Fuentes EJ, Gilmore SA, Mauldin RV, and Lee AL (2006) Evaluation of energetic and dynamic coupling networks in a PDZ domain protein, *Journal of Molecular Biology* 364, 337–351. [PubMed: 17011581]

Figure 1.

Structure and catalytic mechanism of AtHCT. (A) Schematic ribbon diagram of the overall structure of AtHCT in complex with *p*-coumaroylshikimate. The two quasi-symmetric N-terminal (light blue) and C-terminal (dark blue) domains are linked by a long loop (red). The zoom-in insert shows the active site highlighting residues key to the native catalytic activity of HCT. Hydrogen bond interactions are indicated with dotted lines (yellow). The $|2F_o - F_c|$ omit electron density map of *p*-coumaroylshikimate is contoured at 0.8σ (B) The proposed catalytic mechanism of HCT.

Figure 2.

Distinct active-site conformational states and dynamics of HCT revealed by X-ray crystallography and MD simulations. (A) Active-site residues His153 and Arg356 show distinct and consistent conformational shifts upon binding of various ligands. (B) Conformational states of the three loop regions surrounding the active site in various HCT structures. The impact of these motions is to constrict the active site volume significantly. (C) Distribution of the distance change (Δd) between Arg356 and *p*-coumaroylshikimate from simulations of apo (black) and *p*-coumaroyl CoA and shikimate-bound (red) AtHCT.

Calculations were performed by first aligning simulation trajectories to the *p*-coumaroylshikimate-bound AtHCT crystal structure and then measuring the distance change between Arg356 and *p*-coumaroylshikimate. A zero d indicates that Arg356 is at an identical distance from the ligand as in the crystal structure. Details can be found in the Supplementary Experimental Procedures. (D) Internal conformation demonstrated by the first cluster of Arg356 from clustering analysis of *p*-coumaroyl CoA and shikimate-bound AtHCT simulations. (E) External conformations demonstrated by the first two clusters of Arg356 from clustering analysis of apo AtHCT simulations. In (D) and (E), *p*-coumaroylshikimate from the *p*-coumaroylshikimate-bound AtHCT crystal structure is shown as an active site reference. Arg356 from *p*-coumaroylshikimate-bound and apo-AtHCT crystal structures are shown in black in (D) and (E), respectively. Percentage of populations for each cluster is labeled. (F) Multiple sequence alignment highlighting L1, L2, L3, and the catalytic regions in select HCT orthologs representing five major land plant lineages. Corresponding sequence regions from CbRAS are displayed at the bottom for comparison. The residue numbering is according to AtHCT.

Figure 3.

Pseudo first-order Michaelis-Menten kinetics of wild-type, R356A, R356D, and R356E CbHCTs against the native acyl acceptor substrate, shikimate (A), and two non-native acyl acceptor substrates, 3,4-dihydroxybenzylamine (B) and dopamine (C). *p*-coumaroyl-CoA was used as the acyl donor substrate in all assays at constant and excess concentration. Structures of acyl acceptor substrates are shown above the curves.

Figure 4.

2D scatter plots of substrate position and orientation in CbHCT simulations. “Distance” measures the distance between the substrates and Arg356 (or Glu356). “Angle” measures the change in substrates’ orientations with reference to their initial structures (details in the Supplementary Experimental Procedures). Characterized simulations include (A) wild-type CbHCT with regard to shikimate, (B) CbHCT R356E with regard to shikimate, and (C) wild-type CbHCT with regard to 3-hydroxyacetophenone.

Figure 5.

Structural and dynamic features of CbHCT in complex with *p*-coumaroyl-CoA and 3-hydroxyacetophenone. (A) Chemical structure of 3-hydroxyacetophenone (left) and the |2Fo-Fc| omit electron density map of 3-hydroxyacetophenone in the *p*-coumaroyl-CoA and 3-hydroxyacetophenone-bound CbHCT crystal structure contoured at 0.8 σ (right). (B) The conformation of 3-hydroxyacetophenone relative to the catalytic His153 in CbHCT active site. The AtHCT-*p*-coumaroylshikimate structure is overlaid as a comparison. (C) The conformational states of L1, L2, and L3 and Arg356 in various HCT structures. (D) Distribution of the distance change (Δd) between Arg356 and *p*-coumaroylshikimate from simulations of apo CbHCT (black), CbHCT in complex with *p*-coumaroyl CoA and shikimate (red), and CbHCT in complex with *p*-coumaroyl CoA and 3-hydroxyacetophenone (blue). Details can be found in the Supplementary Experimental

Procedures. (E, F) The first cluster of Arg356 conformation from clustering analysis of shikimate- (E) and 3-hydroxyacetophenone-bound (F) CbHCT simulations. As an active site reference, *p*-coumaroylshikimate from the AtHCT crystal structure is shown in (E) and (F), and 3-hydroxyacetophenone from the CbHCT crystal structure is shown in (F). The conformation of Arg356 from the apo CbHCT crystal structure (E) and the 3-hydroxyacetophenone-bound CbHCT crystal structure (F) is shown in black.

Table 1.

Measurement of acyl acceptor binding site volumes for crystal structures of SbHCT and AtHCT.

HCT Ortholog	Ligand	Volume (Å ³)	Constriction (%)
<i>Sorghum bicolor</i>	none	908	-
	<i>p</i> -coumaroyl CoA + shikimate	554	39.0
<i>Arabidopsis thaliana</i>	none	1030	-
	<i>p</i> -coumaroyl CoA	737	28.4
	<i>p</i> -coumaroylshikimate	489	52.5
<i>Coleus blumei</i>	none	1063	-
	<i>p</i> -coumaroyl CoA + 3-hydroxyacetophenone	1037	0.02

Table 2.

Summary of kinetic parameters for wild-type CbHCT and R356 mutants.

	wild-type	R356A	R356D	R356E
	<i>shikimate</i>			
K_m (mM)	0.3882 ± 0.0303	N/A	N/A	N/A
V_{max} (nkat mg ⁻¹)	639 ± 14	N/A	N/A	N/A
k_{cat} (s ⁻¹)	30.1 ± 0.7	N/A	N/A	N/A
k_{cat}/K_m (s ⁻¹ M ⁻¹)	77537 ± 6289	N/A	N/A	N/A
	<i>3,4-dihydroxybenzylamine</i>			
K_m (mM)	265.3 ± 36.8	141.7 ± 8.2	36.03 ± 3.45	16.32 ± 4.83
V_{max} (nkat mg ⁻¹)	485.9 ± 52.2	623.1 ± 23.8	589.2 ± 22.5	142.5 ± 12.7
k_{cat} (s ⁻¹)	22.9 ± 2.5	29.3 ± 1.1	27.7 ± 1.1	6.7 ± 0.6
k_{cat}/K_m (s ⁻¹ M ⁻¹)	86.3 ± 13.1	206 ± 11	768.8 ± 41.5	410.5 ± 51.8
	<i>dopamine</i>			
K_m (mM)	N/A	162.6 ± 14.7	49.4 ± 3.6	41.6 ± 1.6
V_{max} (nkat mg ⁻¹)	N/A	58.02 ± 3.60	30.52 ± 0.99	21.42 ± 0.34
k_{cat} (s ⁻¹)	N/A	2.73 ± 0.17	1.44 ± 0.05	1.01 ± 0.02
k_{cat}/K_m (s ⁻¹ M ⁻¹)	N/A	16.8 ± 1.5	29.1 ± 1.4	24.3 ± 0.6