Evolution of a Catalytic Mechanism

Alissa Rauwerdink,¹ Mark Lunzer,¹ Titu Devamani,¹ Bryan Jones,¹ Joanna Mooney,¹ Zhi-Jun Zhang,¹,² Jian-He Xu,¹,² Romas J. Kazlauskas,¹ and Antony M. Dean*,¹,³,⁴

¹Department of Biochemistry, Molecular Biology & Biophysics and the Biotechnology Institute, University of Minnesota
²Department of Biotechnology, East China University of Science and Technology, Shanghai, PR China
³Department of Ecology, Evolution and Behavior, University of Minnesota
⁴College of Ecology and Evolution, Sun Yat-Sen University, Guangzhou, PR China
*Corresponding author: E-mail: dean@mail.sysu.edu.cn.
Associate editor: Miriam Barlow

Abstract

The means by which superfamilies of specialized enzymes arise by gene duplication and functional divergence are poorly understood. The escape from adaptive conflict hypothesis, which posits multiple copies of a gene encoding a primitive inefficient and highly promiscuous generalist ancestor, receives support from experiments showing that resurrected ancestral enzymes are indeed more substrate-promiscuous than their modern descendants. Here, we provide evidence in support of an alternative model, the innovation–amplification–divergence hypothesis, which posits a single-copied ancestor as efficient and specific as any modern enzyme. We argue that the catalytic mechanisms of plant esterases and descendent acetone cyanohydrin lyases are incompatible with each other (e.g., the reactive substrate carbonyl must bind in opposite orientations in the active site). We then show that resurrected ancestral plant esterases are as catalytically specific as modern esterases, that the ancestor of modern acetone cyanohydrin lyases was itself only very weakly promiscuous, and that improvements in lyase activity came at the expense of esterase activity. These observations support the innovation–amplification–divergence hypothesis, in which an ancestor gains a weak promiscuous activity that is improved by selection at the expense of the ancestral activity, and not the escape from adaptive conflict in which an inefficient generalist ancestral enzyme steadily loses promiscuity throughout the transition to a highly active specialized modern enzyme.

Key words: catalysis, transition state stabilization, gene duplication and divergence, escape from adaptive conflict, innovation, amplification, divergence.

Introduction

Enzyme superfamilies are the products of gene duplication and functional divergence. Members of a superfamily share a common protein fold and similar amino acid sequences (indicating descent from a common ancestor) yet have different activities (suggesting selection for novel functions). Members’ activities often overlap. Substrate promiscuous enzymes favor one substrate, yet also catalyze the same chemical transformation, to a greater or lesser extent, with other substrates (Humble and Berglund 2011). Enzymes that catalyze different reactions (often similar to those of other superfamily members) are said to display catalytic promiscuity (O’Brien and Herschlag 1999; Baas et al. 2013; Baier and Tokuriki 2014). For example, o-succinylbenzoate synthase catalyzes both the normal dehydration of its substrate and the racemization of N-acyl amino acids through the same metal ion stabilized enediolate characteristic of members of the enolase superfamily (Ringjä et al. 2004).

Functional promiscuity is key to several models of superfamily evolution. The escape from adaptive conflict hypothesis (Hughes 1994) posits multiple copies of a gene that encode a primitive inefficient and highly promiscuous generalist enzyme. The modern superfamily emerges as selection for catalytic efficiency forces functional specialization through trade-offs (the idea that a jack-of-all-trades is a master at none). An alternative model, the innovation–amplification–divergence hypothesis (Bergthorsson et al. 2007; Näsvall et al. 2012), posits a single-copied ancestor as efficient and specialized as any modern enzyme. Once mutation has created a weakly beneficial promiscuous function, selection acts to increase expression through gene amplification until the costs associated with excess protein expression become prohibitive. Once again, trade-offs ensure that increases in catalytic efficiency force increases in functional specialization, whereas selection against excess protein expression eliminates unneeded copies.

The key difference between these models lies not in their outcomes—both predict superfamilies of enzymes with related, though distinct (if overlapping) functions—but in their origins. The escape from adaptive conflict hypothesis posits an inefficient but highly promiscuous ancestor. The innovation–amplification–divergence hypothesis posits an ancestor as efficient and as specialized as any modern enzyme.

At first glance it might seem impossible to distinguish between the models; after all, the ancestor that existed millions of years ago is extinct. However its genes were passed on,
generation upon generation, albeit modified by mutation, to its living descendants. By analyzing the gene sequences of extant species, researchers can infer the sequences of ancestral genes (Thornton 2004; Harms and Thornton 2013). Such an analysis uses a set of aligned sequences from extant species, a phylogenetic tree, and a probabilistic model of sequence evolution that incorporates differences in the rates of evolution and the likelihood that one amino acid will be exchanged for another. Then, the ancient gene is chemically synthesized and the protein expressed in a suitable host. Where analysis suggests several possible amino acids at a location, several plausible ancestors might be constructed and characterized. In most cases, such ancestors have similar properties. For example, reconstructed ancestral β-lactamases and plausible alternative protein sequences have similar phenotypes (Risso et al. 2013).

Reconstructed ancestral enzymes support the escape from adaptive conflict model because they have broader substrate ranges than their modern descendants. Voordecker et al. (2012) showed that reconstructed ancestral α-glycosidases catalyzed hydrolysis of a wide range of substrates (maltose and isomaltose analogs), whereas modern glycosidases are comparatively specific. Likewise, reconstructed 600-My-old β-lactamases catalyze hydrolysis of a wider range of β-lactams, including third-generation semisynthetic antibiotics that never existed in nature (Risso et al. 2013). However, the hypothesis that ancestral enzymes were more catalytically promiscuous than modern enzymes has not been tested.

Acetone cyanohydrin lyases (ACLs) are hydroxynitrile lyases, a heterogeneous class of enzymes that produce cyanide as a defense against insects (Poulant 1990) and that independently have evolved in at least five distinct protein folds. We tentatively have evolved in at least five distinct protein folds. We present data suggesting several possible amino acids at a location, several plausible ancestors might be constructed and characterized. In most cases, such ancestors have similar properties. For example, reconstructed ancestral β-lactamases and plausible alternative protein sequences have similar phenotypes (Risso et al. 2013).

Reconstructed ancestral enzymes support the escape from adaptive conflict model because they have broader substrate ranges than their modern descendants. Voordecker et al. (2012) showed that reconstructed ancestral α-glycosidases catalyzed hydrolysis of a wide range of substrates (maltose and isomaltose analogs), whereas modern glycosidases are comparatively specific. Likewise, reconstructed 600-My-old β-lactamases catalyze hydrolysis of a wider range of β-lactams, including third-generation semisynthetic antibiotics that never existed in nature (Risso et al. 2013). However, the hypothesis that ancestral enzymes were more catalytically promiscuous than modern enzymes has not been tested.

Acetone cyanohydrin lyases (ACLs) are hydroxynitrile lyases, a heterogeneous class of enzymes that produce cyanide as a defense against insects (Poulant 1990) and that independently have evolved in at least five distinct protein folds. We predict that ancestral plant esterases in the α/β-hydrolase fold superfamily, which gave rise to modern ACLs, were as catalytically specialized as modern esterases. Unlike substrate promiscuity and catalytic promiscuity among mechanistically related reactions, the reaction mechanisms of esterases and ACLs differ dramatically (fig. 1; Gruber et al. 2004), even though both contain the same Ser–His–Asp catalytic triad (a chemical example of the Darwinian principle of finding new uses for old parts). First, the binding orientations of the substrates differ. In esterases, the ester carbonyl binds in an oxyanion hole with its carbon near the Ser. In ACLs, the oxyanion hole is blocked and the substrate aldehyde carbonyl binds with its oxygen near the Ser. Second, ester hydrolysis forms and later releases an acyl enzyme intermediate in which the acyl part of the ester is linked covalently to the active site Ser. In contrast, acetone cyanohydrin cleavage occurs in a single step without covalent links between substrate and enzyme. Third, ester hydrolysis releases an uncharged hydrophobic alcohol whereas acetone cyanohydrin cleavage releases a small charged cyanide group. These mechanistic differences are so marked that stabilizing an esterase transition state must necessarily impede stabilizing the ACL transition state. For this reason we predict that the ACLs evolved, not by escape from adaptive conflict, but by innovation, amplification and divergence, from a weakly promiscuous ancestor in a lineage of otherwise highly specific esterases.

**Results**

**Phylogenetic Analysis**

The α/β-hydrolase superfamily contains greater than 60,000 proteins (Kourist et al. 2010; Lenfant et al. 2013) with wide-ranging catalytic activities (Ollis et al. 1992; Holmquist 2000; Nardini and Dijkstra 1999). Most are hydrolases (Enzyme Classification group 3), although the superfamily also includes ACLs (Enzyme Classification group 4). Phylogenetic analysis shows the ACLs cluster deep within a larger group of plant esterases (fig. 2). The maximum-likelihood tree is well resolved with most clusters receiving strong (> 80%) bootstrap support. The topology of the neighbor-joining tree is similar (differences are restricted to short branches with low bootstrap support).

ACLs, from the rubber tree *Hevea brasiliensis* (HbACL; Gruber et al. 1999), cassava *Manihot esculenta* (MeACL; Selmar et al. 1989; Hughes et al. 1994), and wild castor *Baliospermum montanum* (Nakano et al. 2014), share approximately 40% identity with related plant esterases. The ACL cluster is allied to a cluster that includes an esterase from *Arabidopsis thaliana* (AtEST5; Andexer et al. 2007) and other esterases of unknown function. Next is a cluster of 174 uncharacterized sequences, and then a cluster that includes two well-characterized esterases, salicylic acid binding protein 2 (SABP2) from the tobacco, *Nicotiana tabacum* (NtEST; Forouhar 2005), and polyneuridine-aldehyde esterase from snakeroot, *Rauwolfia serpentina* (RsEST; Mattern-Dogru 2002).

More distantly related (with only 20–40% amino acid identity to the ACLs) are the decarboxylase methylketone synthase I (ShMKS) from *Solamun habrochaites* (Yu 2010; Auldridge 2012) and in a separate cluster of 41 uncharacterized sequences a previously uncharacterized esterase from *Ricinis communis* (ReEST). Other enzymes in the α/β-hydrolase superfamily from bacteria and animals, including C–C hydrolyses (e.g., BphD from *Burkholderia xenovorans*; Horsman et al. 2007), epoxide ring hydrolyses (e.g., human soluble epoxide hydrolase, *Beetham et al. 1993;* Gomez et al. 2006; Thalji et al. 2013) and haloalkane hydrolases (e.g., LinB from *Sphingobium* sp.; PDB: 4H77_A; Okai 2013), share the same fold but are too distantly related (only 10–20% amino acid identity) for reliable phylogenetic analysis.

**Structural Evolution**

Different sites in these plant esterases and ACLs evolve at different rates. The broad pattern is typical of many globular enzymes. Sites critical to function are invariant. For example, the Ser–His–Asp catalytic triad is invariant and only in distantly related α/β hydrolyses, such as dehalogenases and epoxide hydrolyses, are variations found. Conservative replacements (e.g., Ile for Leu) predominate in the slowly evolving hydrophobic core. Surface sites far from the active site evolve rapidly, as do the amino and carboxy termini which have been subject to numerous insertions and deletions. The catalytic domain is more conserved than the lid
domain that forms part of the substrate-binding site. Several residues contacting ligands bound in the active site are not conserved. These presumably contribute to differences in substrate specificity.

Ancestral Enzyme Reconstruction

Plant esterases and ACLs share greater than 40% amino acid identity making it feasible to robustly infer ancestral enzyme sequences. Ancestral sequences at focal nodes ACL1, and EST1, EST2 and EST3 (fig. 2) were estimated using maximum likelihood in RaxML (Stamatakis 2006) for two trees, one constructed by maximum likelihood (Felsenstein 1981) and the other using neighbor joining (Saitou and Nei 1987). As expected, all predicted ancestral enzyme sequences contain conserved residues common to most enzymes in the family, including the Ser–His–Asp catalytic triad. Nevertheless, the ancestral enzyme sequences are different from the sequences of modern enzymes. Most similar are those at node ACL1 and $Hb$ACL ($\sim81$% identical), then those at nodes EST1 and EST2 to $Nt$EST ($\sim73$% identical) and least similar are those at node EST3 to $At$EST5 and $Hb$ACL ($\sim67$% identical). Note that ancestral sequences at EST3 occupy positions in raw sequence space intermediate between modern ACLs and modern plant esterases. The longer branch lengths leading to the ACLs reflect multiple mutations at each evolving site.

Reconstruction of an exact ancestral sequence is highly unlikely because of uncertainties at some sites. Marginal ancestral state reconstructions suggest that the most likely amino acids were identified with greater than 90% confidence at greater than 85% of sites. Only at node EST3 was the maximum-likelihood sequence less robustly estimated ($>90$% confidence at $>75$% of sites), a consequence of its position at the intersection of three relatively long branches. To accommodate these uncertainties, we constructed enzymes for each focal node that contained the most likely alternative amino acids at ALL uncertain sites. Although these alternative sequences are highly improbable ancestors, their activities can be expected to be similar to their respective maximum-likelihood-estimated ancestors because most differences are physically conservative (e.g., between Asp and Glu) at rapidly evolving and functionally unconstrained sites at the protein surface. Also reconstructed by maximum likelihood were ancestral sequences using the neighbor-joining tree. Based on their positions in the trees we expect reconstructed enzymes at node ACL1 to be efficient ACLs, at EST3 to be esterases with weak promiscuous acetone cyano hydridin cleavage activities, and at EST1 and EST2 to be specialist esterases.

Activities

Six-His-tagged enzymes, both ancient and modern, were expressed and purified to homogeneity and their activities determined toward acetone cyano hydridin and four esters. Acetone cyano hydridin is the natural substrate for this cluster of ACLs (Butler 1965). The natural substrates for the esterases are not known, except for $Rs$EST which is a polynereurdine esterase (Mattern-Dogru 2002) and $Nt$EST which hydrolyzes methyl salicylate in a plant-defense-signaling pathway (Forouhar 2005). In order to focus on esterase activity rather than substrate specificity, we surveyed hydrolysis of four esters that differ in the sizes of their acyl and alcohol moieties.

As expected, $Rc$EST and $Nt$EST catalyze hydrolysis of all four esters and do not catalyze acetone cyano hydridin cleavage. From its position in the phylogenetic tree, we anticipated that the previously uncharacterized $Rs$EST would be an esterase. Indeed, $Rs$EST is an efficient esterase that lacks acetone cyano hydridin cleavage activity (table 1).

$At$EST5 is an esterase. $At$EST5 is unusual because it cleaves mandelonitrile (Andexer et al. 2007) in addition to ester hydrolysis (Koo et al. 2013). Its natural function is not cleaving hydroxynitriles, however, because 1) the family Brassicaceae, of which A. thaliana is a member, contains no cyanogenic glucosides (Olson-Manning et al. 2013); 2) acetone cyano hydridin, the natural substrate for this class of ACLs (Butler 1965), is not a substrate for $At$EST5; and 3) $At$EST5 hydrolyzes esters at rates comparable to known the esterases (table 1). $At$EST5 and other sequences in this cluster are, like their ancestors, esterases, although their natural substrates have yet to be identified.

Nestled deep within the esterases, ancient enzymes at EST1 and EST2 hydrolyze all four esters efficiently and lack acetone cyano hydridin cleavage activity. The maximum-likelihood ancestor at EST3 is an esterase with weak acetone cyano hydridin cleavage activity. The alternative enzymes at this node are esterases, although they vary somewhat in substrate specificity. All three reconstructed enzymes at ACL1 have good acetone cyano hydridin cleavage activities and retain relatively high esterase activities. These enzymes are catalytically promiscuous. $Hb$ACL and $Me$ACL are highly specific ACLs, although $Hb$ACL can hydrolyze methyl salicylate, albeit slowly.

That esterase activities increase from EST3 to ACL1 is not surprising. Most amino acid substitutions compromise protein structures. Those that cause functional changes are no exception. Compensatory mutations following the acquisition of acetone cyano hydridin cleavage activity evidently raised the ancestral esterase activity. Importantly, the specificity of EST3-ML toward acetone cyano hydridin versus 1-naphthyl acetate is 0.078/0.0036 = 21.7, whereas that for ACL1-ML is 720/0.054 = 13,333.3. In other words, the specificity toward acetone cyano hydridin cleavage has increased more than 600-fold. This represents a decisive shift away from esterase activity toward acetone cyano hydridin cleavage specialization.

Discussion

For two reasons, our results support the innovation–amplification–divergence hypothesis (Bergthorsson et al. 2007; Näsval et al. 2012) rather than the escape from adaptive conflict hypothesis (Hughes 1994). First, modern specialist ACLs evolved from an equally specialized ancestor (EST2; table 1 and fig. 2). Second, catalytic promiscuity increased during the transition, from EST2 through EST3 to ACL1, before it decreased in the modern ACLs. These observations are incompatible with escape from adaptive conflict where an inefficient generalist ancestral enzyme steadily loses
Fig. 1. The catalytic mechanisms of esterases (left) and hydroxynitrile lyases (right) are quite different. Esterases use a two-step mechanism. First, the catalytic Ser acts as a nucleophile to attack the ester's carbonyl carbon with the resulting formation of an acyl enzyme intermediate and release of the alcohol. Second, hydrolysis of the carboxylic acid from the Ser is initiated when the catalytic His abstracts a proton from water, the resulting hydroxide attacks the carbonyl carbon and the same mechanism effectively proceeds in reverse. Hydroxynitrile cleavage occurs in a single concerted step involving simple acid-base chemistry. Note that the orientation of the carbonyl differs in the two reactions. In esterases, the carbonyl carbon points toward the Ser and the oxygen toward the oxanion hole. In hydroxynitrile lyases, the carbonyl oxygen points toward the catalytic Ser.
promiscuity throughout the transition to a highly active specialized modern enzyme.

Several studies have shown that, like their modern descendants, reconstructed ancestral enzymes and proteins display substrate and ligand promiscuity (Eick et al. 2012; Voordekers et al. 2012). Both modern and ancestral plant esterases and ACLs also show considerable substrate promiscuity (Guterl et al. 2009; Andexer et al. 2012), yet only rarely are they catalytically promiscuous. This is because the two catalytic mechanisms must stabilize very different transition states (fig. 1; Gruber et al. 2004). The esterases follow the canonical two-step serine hydrolase mechanism, with its

![Fig. 2. Maximum-likelihood phylogenetic tree of $\alpha/\beta$ hydrolases showing extant ACLs (red text) nested within a larger group of esterases (EST, blue text), the neighboring methylketone synthase (MKS, a decarboxylase, green text), and distantly related meta-cleavage product hydrolases (connected by dotted lines). Gray triangles represent clusters of proteins (numbers shown at the tips of the triangles). Resurrected ancestors EST1 and EST2 (blue) are dedicated esterases, EST3 (lavender) shows weak acetone cyanohydrin activity, ACL1 (purple) is an effective ACL with some esterase activity, whereas extant ACLs (red) are dedicated ACLs. Numbers at bifurcations and triangle bases represent bootstrap support.](image-url)
acyl enzyme intermediate, whereas the lyase mechanisms use a concerted mechanism involving only binding and general acid-base chemistry (fig. 1). Moreover, stabilizing two different transition states is made all the more difficult as the enzyme must hold the substrates' reactive groups in opposing positions. Hence, chemistry predicts that 1) modern plant esterases and ACLs, as well as their ancestors, are rarely promiscuous for both reactions; 2) promiscuous enzymes heavily favor one reaction over the other; 3) the mutational origin of acetone cyanohydrin cleavage is a rare event with all modern ACLs descended from an ancestral esterase with an inefficient promiscuous ACL activity; and 4) esterase activity was lost rapidly during divergence to modern ACLs, not only because the transition states of the reactions are so very different but also because the enzyme needs to hold the reactive carbonyl in opposite orientations in the two reactions.

All four predictions are confirmed (table 1). First, neither modern esterases nor the ancestral esterases at nodes EST1 and EST2 have detectable acetone cyanohydrin cleavage activities. By extrapolation, most ancestral esterases in this family never had them either. Modern ACLs are also highly specific enzymes, although HbACL has a weak methyl salicylate esterase activity. This exception suggests the intriguing possibility that HbACL moonlights in the plant-defense--signaling pathway involving methyl salicylate. Second, in esterases a hydrophobic residue (methionine in SABP2) usually forms part of the alcohol binding site, whereas in ACLs a lysine in this location creates a polar site for the cyanide. These two amino acid replacements in SABP2 (Gly12Thr, Met239Lys) switched catalysis from esterase to ACL cleavage. However, the corresponding substitutions in a more distantly related esterase did not create hydroxynitrile activity indicating that other unknown features also contribute to this activity. Ancestral enzyme ACL1 contains both key residues—a threonine to block the oxyanion hole and a lysine in the alcohol binding site.

In contrast, EST3 appears to be an esterase with a novel ACL activity. As in AtEST5, both EST3-ML and EST3-MIX have replaced the Gly in the oxyanion loop with an Asn. Unlike Thr, Asn is not branched at the Cβ and might still permit the substrate's carbonyl oxygen to access the oxyanion hole. Thus, the three enzymes each retain considerable esterase activity. Alternatively, the Asn δ-amide might act as a hydrogen bond donor to stabilize the oxyanion intermediates during ester

Table 1. Activities of Modern and Resurrected Ancestral Enzymes.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Hydroxynitrile</th>
<th>Methyl Salicylate</th>
<th>1-Naphthyl Acetate</th>
<th>2-Naphthyl Acetate</th>
<th>Methyl Pentanoate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbACL</td>
<td>2,400</td>
<td>0.066</td>
<td>NDA</td>
<td>NDA</td>
<td>NDA</td>
</tr>
<tr>
<td>MeACL</td>
<td>12,600</td>
<td>0.66</td>
<td>0.054</td>
<td>0.13</td>
<td>0.41</td>
</tr>
<tr>
<td>ACL1-ML</td>
<td>720</td>
<td>0.048</td>
<td>0.016</td>
<td>0.024</td>
<td>NDA</td>
</tr>
<tr>
<td>ACL1-ALT</td>
<td>350</td>
<td>0.03</td>
<td>0.0090</td>
<td>0.0054</td>
<td>0.94</td>
</tr>
<tr>
<td>EST3-ML</td>
<td>0.078</td>
<td>NDA</td>
<td>0.0036</td>
<td>0.0084</td>
<td>NDA</td>
</tr>
<tr>
<td>EST3-ALT</td>
<td>NDA</td>
<td>NDA</td>
<td>NDA</td>
<td>0.0024</td>
<td>0.60</td>
</tr>
<tr>
<td>EST3-NJ</td>
<td>NDA</td>
<td>0.13</td>
<td>0.28</td>
<td>0.46</td>
<td>42</td>
</tr>
<tr>
<td>EST1-M</td>
<td>26</td>
<td>0.096</td>
<td>2.2</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>EST3-M</td>
<td>NDA</td>
<td>14</td>
<td>0.084</td>
<td>1.5</td>
<td>310</td>
</tr>
<tr>
<td>RsEST</td>
<td>NDA</td>
<td>7.5</td>
<td>0.12</td>
<td>0.18</td>
<td>2.6</td>
</tr>
<tr>
<td>RsEST</td>
<td>NDA</td>
<td>0.042</td>
<td>0.042</td>
<td>0.012</td>
<td>6.5</td>
</tr>
<tr>
<td>SABP2</td>
<td>NDA</td>
<td>0.52</td>
<td>0.012</td>
<td>0.018</td>
<td>180</td>
</tr>
<tr>
<td>AtEST5</td>
<td>NDA</td>
<td>NDA</td>
<td>1.7</td>
<td>0.038</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Note.—NDA, no detectable activity.

aActivities are $k_{cat}$ (min$^{-1}$).
bActivities are min$^{-1}$. Rates measured at one substrate concentration given in Materials and Methods.

cStandard errors $<10\%$ of estimates.

The substrates themselves require different reactions; an ester cannot undergo an elimination reaction and a cyanohydrin cannot undergo hydrolysis. Enabling cyanohydrin cleavage within an esterase active site requires as few as two amino acid replacements (Padhi et al. 2010). Esterases contain a Gly in the oxyanion loop, whereas ACLs contain a threonine, which physically blocks access to the oxyanion hole. Second, in esterases a hydrophobic residue (methionine in SABP2) usually forms part of the alcohol binding site, whereas in ACLs a lysine in this location creates a polar site for the cyanide. These two amino acid replacements in SABP2 (Gly12Thr, Met239Lys) switched catalysis from esterase to ACL cleavage. However, the corresponding substitutions in a more distantly related esterase did not create hydroxynitrile activity indicating that other unknown features also contribute to this activity. Ancestral enzyme ACL1 contains both key residues—a threonine to block the oxyanion hole and a lysine in the alcohol binding site.
hydrolysis. For the other replacement, EST3-MIX contains a lysine typical of ACLs whereas EST3-ML still retains the methionine typical of esterases. One would expect that of the two ancestral reconstructions, EST3-MIX would have higher acetone cyanohydrin cleavage activity. Yet the opposite is true, EST3-ML is the better lyase. Clearly, the details by which successive amino acid replacements, including the role of epistasis, changed a highly specific esterase into a highly specific ACL remain to be determined.

Not all catalytic mechanisms differ as dramatically as those of plant esterases and ACLs. Work by Gerlt and collaborators (Gerlt et al. 2005; Glasner et al. 2006) has established the existence of enzyme superfamilies whose members catalyze different reactions by stabilizing a common transition state. For example, members of the enolase superfamily use a base to abstract a proton alpha to a carboxylate to produce an enediolate, followed by protonation either on the opposite face for an isomerization (e.g., mandelate racemase) or of the leaving group for an elimination (e.g., enolase, a dehydratase). We conjecture that enzymes that stabilize similar transition states have a greater potential for catalytic promiscuity and could well have arisen through escape from adaptive conflict hypothesis. Unfortunately, so divergent are the members of the enolase and many other superfamilies that reconstructing ancestors becomes a highly unreliable exercise. Testing our conjecture at the limits of detectable sequence similarity will prove difficult, if not impossible.

In summary, we have shown that the highly specific ancestral plant esterase (EST2) that gave rise to modern ACLs first acquired a weak promiscuous acetone cyanohydrin cleavage activity (EST3), and that the subsequent increase in acetone cyanohydrin cleavage activity (ACL1 and modern ACLs) is associated with a rapid loss of esterase activity. This pattern of functional evolution is both consistent with the innovation–amplification–divergence hypothesis (Bergthorsson et al. 2007; Näsvall et al. 2012), and predictable based on different active site chemistry of the two reactions.

Materials and Methods

Phylogenetics

Five thousand protein sequences, 150–600 amino acids long and sharing a minimum 30% sequence identity with Hevea brasiliensis ACL (G1: 1223884), were obtained from the NCBI protein sequence database. Identical copies, mutant peptides, and all patents were removed and the remaining sequences aligned using Muscle (Edgar 2004) in SeaView (Gouy et al. 2010). A preliminary neighbor-joining (Saitou and Nei 1987) tree was used to identify a cluster of 1,285 sequences between 30% and 99% identical, and that included the ACLs and salicylic acid binding protein 2 from Nicotiana tabacum, for further analysis.

The 1,285-amino acid sequence alignment was adjusted manually guided by superpositioned protein structures (H. brasiliensis 1QH4, Manihot esculenta 1EBD, Arabidopsis thaliana 3DZQ, Nicotiana tabacum 1XKL) obtained with DeepView/Swiss PDB-Viewer (Guex and Peitsch 1997) to adjust insertions and deletions into surface loops. The final alignment of 1,285 sequences is available upon request from the authors.

Bootstrapped (1,000 replicates) maximum-likelihood trees were obtained using RAxML (Stamatakis 2006) using the ML + Bootstrap + GAMMAPROT + LG settings. Maximum likelihood uses a probabilistic model of sequence evolution to construct a tree from a given alignment. The GAMMAPROT setting to allow some sites to evolve faster than others and the LG (Le and Gascuel 2008) providing an empirical amino acid exchange matrix to accommodate differences in the rates of exchange between different amino acids. The most likely tree is one with topology, branch lengths and other parameters that maximize the likelihood of the observed alignment. An advantage of the maximum-likelihood approach is that it provides a natural means to test alternative hypotheses—branch lengths, topologies, and ancestral states. Similar trees were obtained using FastML with GAMMAPROT and a JTT (Jones, Taylor, and Thorton) amino acid exchange matrix (Jones et al. 1992) in RAxML with SH support (Shimodaira and Hasegawa 1999) and bootstrapped neighbor joining with a Poisson correction in SeaView.

Ancestral sequences were inferred using maximum likelihood as implemented in RAxML using the ML + Bootstrap + GAMMAPROT + LG tree obtained above. Maximum likelihood calculates the most likely amino acid at each site at each node using an empirical Bayes approach (Yang et al. 1995). The likelihood of observing a particular amino acid, x, at a particular site at a particular node is given as

\[
P(x | a, m, t, \theta) = \frac{P(x | a, m, t, \theta)p_x}{\sum_{a=1}^{20} P(x | a, m, t, \theta)p_x},
\]

where \( a \) is one of 20 amino acids at the focal site, \( t \) is the topology of the phylogenetic tree, \( m \) is an evolutionary model, \( \theta \) represents the various model parameters (rates of amino acid exchange, variable rates across sites, etc.), and \( p_x \) is the prior probability of observing \( x \). Maximum likelihood resolves ambiguities in favor of the most likely model. For example, in the case of tree (((Leu,Leu),(Met,Met)),Arg) the most likely ancestor of sequences 1–4 had an Met because exchanges between Met and Leu and between Met and Arg are commonplace whereas those between Leu and Arg are rare. Ancestral sequences were also inferred for the neighbor-joining tree using maximum likelihood.

Gene Synthesis and Cloning

Genes for the ancestral enzymes were synthesized by GenScript (Piscataway, NJ) and subcloned into a pET21a(+) vector at NdeI and XhoI restriction sites resulting in an upstream T7 promoter and lac operator and an in frame C-terminal six His-tag. DNA sequencing of the entire gene confirmed the fidelity of cloning.

Protein Expression and Purification

Lysogeny broth media containing 100 μg/ml ampicillin (LB-amp, 5 ml) was inoculated with a single colony
(Escherichia coli) BL21 transformed with the with appropriate plasmid) from an agar plate and incubated in an orbital shaker at 37 °C and 200 rpm for 15 h. This culture was used to inoculate terrific broth-amp media (500 ml), which was incubated at 37 °C and 250 rpm until the absorbance at 600 nm reached 1.0 (approximately 3–4 h). This culture was cooled at 17 °C and 200 rpm for 1 h, then isopropyl β-D-1-thiogalactopyranoside (1 mM) was added to induce the protein expression and the cultivation continued for 20 h. The cells were harvested by centrifugation (8,000 rpm, 10 min at 4 °C) and suspended in buffer A (20 mM imidazole, 50 mM NaH2PO4, 300 mM NaCl, pH 7.2, 10 ml). The cells were disrupted by sonication (400 W, 40% amplitude for 5 min) and centrifuged (4 °C, 12,000 rpm 45 min). The supernatant was loaded onto a column containing Ni-NTA resin (1 ml, Qiagen) pre-equilibrated with buffer A (30 ml). The column was washed buffer A (50 ml) followed by buffer B (50 mM imidazole, 50 mM NaH2PO4, 300 mM NaCl, pH 7.2, 50 ml). The His-tagged protein was eluted with elution buffer (150 mM imidazole, 50 mM NaH2PO4, 300 mM NaCl, pH 7.2, 15 ml) and concentrated to approximately 1 ml with an Amicon ultrafiltration centrifuge tube (10 kD cutoff). The imidazole buffer was exchanged by four successive additions of BES buffer (5 mM, pH 7.0, 10 ml) followed by concentration to approximately 1 ml with the centrifuge tube.

Cleavage of Acetone Cyanohydrin

The release of cyanide from acetone cyanohydrin was measured using a modified König reaction (Selmar 1987; Andexer 2006). Enzyme (up to 8 μM in 5 mM citrate phosphate buffer, pH 5) was added to a solution of acetone cyanohydrin (1.28–10 mM in 0.1 M citric acid). After up to 20 min, the reaction was quenched by addition of aqueous N-chlorosuccinimide (2 mM, 62.5 μl, also containing 20 mM succinimide). After 2 min, barbituric acid (230 mM in 30% pyridine, 12.5 μl) was added and after 10 min, the absorbance at 580 nm was measured and compared with a calibration curve constructed using K2[Zn(CN)4] with concentrations of HCN ranging from 2.5 to 100 μM.

Esterase Activity

Methyl salicylate (500 μM), methyl pentanoate, 1-naphthyl acetate, and 2-naphthyl acetate (2 mM) in 5 mM BES buffer pH 7 with up to 50 μM enzyme. Control reactions were performed by adding equal volumes buffer that had been removed from the enzyme by centrifugal filtration. Reactions were shaken at room temperature for up to 18 h before enzyme was removed by centrifugation using Amicon Ultra 0.5 ml regenerated cellulose 10,000 NMWL centrifugal filters. Reactions with methyl salicylate, 1-naphthyl acetate, and 2-naphthyl acetate were analyzed by HPLC on an Agilent Eclipse Plus C18 column eluted with methanol: Water + 0.1% formic acid (80:20) at 1.0 ml/min using a photodiode array detector. Methyl pentanoate hydrolysis was quantified p-nitrophenol as a pH-indicator as described previously (Janes 1998).

Acknowledgments

This work was supported by the National Institute of Health (NIH) grant R01GM102205 and National Science Foundation (NSF) grant 1152804 to A.M.D. and R.J.K.

References


Evolution of a Catalytic Mechanism


