Finding homes for orphan enzymes

Frank M. Raushel

Department of Chemistry, Texas A&M University, College Station, TX 77843, USA

Received 17 December 2015; accepted 1 February 2016
Available online 31 August 2016

KEYWORDS
Amidohydrolase superfamily; Annotation of function; Deamination reactions; Carbohydrate metabolism; Phosphate hydrolysis

Summary The rate at which new genes are being sequenced greatly exceeds our ability to correctly annotate the functional properties of the corresponding proteins. Annotations based primarily on sequence identity to experimentally characterized proteins are often misleading because closely related sequences may have different functions, while highly divergent sequences may have identical functions. Our understanding of the principles that dictate the catalytic properties of enzymes, based on protein sequence alone, is often insufficient to correctly annotate proteins of unknown function. To address these problems, we are working to develop a comprehensive strategy for the functional annotation of newly sequenced genes using a combination of structural biology, bioinformatics, computational biology, and molecular enzymology. The power of this multidisciplinary approach for discovering new reactions catalyzed by uncharacterized enzymes has been tested using the amidohydrolase superfamily as a model system.
© 2016 Beilstein-Institut. Published by Elsevier GmbH. This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/).

Introduction

The rate at which new genes are being sequenced greatly exceeds our ability to correctly annotate the functional properties of the corresponding proteins. Annotations based primarily on amino acid sequence identity to experimentally characterized proteins are often misleading because closely related sequences can have different catalytic functions, while highly divergent sequences may have identical functions. Our understanding of the underlying physical principles that dictate the catalytic properties of enzymes, based on protein sequence alone, is often insufficient to correctly annotate proteins of unknown function. New methods must therefore be developed to define the sequence boundaries for a given catalytic activity and new approaches must be formulated to identify those proteins that are functionally distinct from their close sequence homologues. To address these problems, we are working to develop a comprehensive strategy for the functional annotation of newly sequenced genes using a combination of structural biology, bioinformatics, computational biology, and molecular enzymology. The power of this multidisciplinary approach for discovering new reactions catalyzed by uncharacterized enzymes has been tested using the amidohydrolase superfamily as a model system.

http://dx.doi.org/10.1016/j.pisc.2016.02.002
2213-0209 © 2016 Beilstein-Institut. Published by Elsevier GmbH. This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/).
approach for discovering new reactions catalyzed by uncharacterized enzymes is being tested and refined using the amidohydrolase superfamily (AHS) as a model system.

The AHS is an ensemble of evolutionarily related enzymes capable of hydrolyzing amide, amine, or ester functional groups at carbon and phosphorus centers (Selbert and Raushel, 2005). More than 100,000 protein sequences have been identified as belonging to this superfamily and the proteins have been segregated into 24 clusters of orthologous groups (COG). The amidohydrolase superfamily was first identified by Sander and Holm when they recognized the structural similarities between adenosine deaminase, phosphorriesterase, and urease (Holm and Sander, 1997). These three proteins fold as (β/α)8-barrels where the active site is located at the C-terminal end of the β-barrel. The active site contains one or two divalent cations and these metal ions are used for the activation of a water molecule for nucleophilic attack on the substrate. Since the original discovery of the amidohydrolase superfamily, other enzymes have been shown to require the binding of three divalent cations for catalytic activity and some other enzymes have been demonstrated to not need divalent cations for catalytic activity (Ghodge et al., 2013a; Hobbs et al., 2012). Most of the structurally and functionally characterized enzymes in this superfamily require either zinc, iron, or nickel in the active site.

In this presentation we describe the efforts that were used to identify the reactions catalyzed by five different enzymes from the amidohydrolase superfamily. Three of these enzymes catalyze the hydrolysis of phosphate esters, with the unusual discovery that two of these are able to catalyze both monoester and diester substrates (Ghodge et al., 2013b; Cummings et al., 2014; Ghodge and Raushel, 2015). Another enzyme is shown for the first time to catalyze the deamination of cAMP (Goble et al., 2013). The last enzyme catalyzes the hydrolysis of a carbohydrate-based lactone in a novel pathway for the metabolism of L-galactose in a bacterium found in the human gut microbiome (Hobbs et al., 2014).

cAMP deaminase

Enzymes within three different COGs from the AHS have been shown to catalyze the deamination of nucleic acid bases including adenosine (cog1816), adenine (cog1001), guanine (cog0402), and cytosine (cog0402). When we initiated our investigation, all of the enzymes contained within cog1816 were annotated as catalyzing the deamination of either adenosine or AMP. As a graduate student, Alissa Goble elected to further interrogate the reactions catalyzed by member of cog1816. The first of these proteins, with the locus tag Lic10459, was isolated from the human pathogen Leptospira interrogans. At the time Lic10459 was annotated as an adenosine deaminase. Examination of the genomic context for the gene that encodes the protein for this enzyme did not identify any other related genes that would suggest that Lic10459 was part of an obvious metabolic pathway. A protein sequence alignment with other enzymes known to be adenosine deaminases indicated that there was a high degree of sequence identity with residues known to interact directly with adenosine in the active site. These conserved residues included an aspartate and glutamate that function as general acid/base residues for the transfer of protons from the nucleophilic water molecule and the ammonia product. The purified enzyme was tested as a catalyst for the deamination of adenosine but no activity could be found.

The lack of substrate activity with adenosine prompted the collection of compounds that were structurally similar to adenosine. A group of approximately 60 compounds was ultimately tested as substrates for Lic10459 and the only compound that was found to be a substrate was cAMP (Goble et al., 2013). This finding was quite surprising since cAMP had not previously been shown to be deaminated to cIMP. In bacterial cells cAMP can function as a regulator of gene expression where it has been shown to bind to a cAMP receptor protein that subsequently enables the transcription of DNA by RNA polymerase. In many bacteria the concentration of cAMP is dictated by the rate of synthesis from ATP and the rate of degradation by cAMP-specific phosphodiesterases. Lic10459 was shown to catalyze the deamination of cAMP with a $k_{cat}$ of 46 s$^{-1}$ and a $k_{cat}/K_m$ of $3 \times 10^5$ M$^{-1}$ s$^{-1}$. The reaction is presented in Scheme 1.

In collaboration with the laboratory of John Cronan from the University of Illinois we were able to express the enzyme in the bacterium Escherichia coli. The Cronan laboratory was able to demonstrate that expression of Lic10459 in E. coli resulted in a significant reduction in the steady state concentration of cAMP in the cell (Goble et al., 2013). Moreover, E. coli was no longer able to grow in the presence of those sugars that require the binding of cAMP to the receptor protein before the genes needed for the metabolism of these sugars are expressed. These low-status sugars included

![Scheme 1](image)

Scheme 1  Reaction catalyzed by Lic10459 from Leptospira interrogans.
lactose, maltose, and arabinose. We were also able to demonstrate that cIMP was unable to productively bind to the cAMP receptor protein from E. coli. We believe that these results are consistent with the conclusion that the steady state concentration of cAMP in L. interrogans is regulated in part by the deamination of cAMP to cIMP.

**Metabolism of L-galactose**

Enzymes in cog3618 have previously been annotated as catalyzing the hydrolysis of lactones (Hobbs et al., 2013). One of these enzymes has been shown to catalyze the hydrolysis of 2-pyrene-4,6-dicarboxylate in a pathway for the degradation of lignin (Hobbs et al., 2012). Another enzyme has been shown to catalyze the hydrolysis of L-rhamnonolactone in a pathway for the metabolism of L-rhamnose. This particular cluster of enzymes is unusual in that a divalent metal ion is apparently not required for catalysis. Eric Hobbs elected to explore other unannotated enzymes from cog3618 and chose Bvu0220 from the human gut bacterium Bacteroides vulgatus. The gene for Bvu0220 resides in an operon that appears to be responsible for the metabolism of a sugar since it is adjacent to an NADP-dependent oxidoreductase (Bvu0219), a sugar permease (Bvu0221), a putative glycoside hydrolase (Bvu0217) and a zinc-dependent dehydrogenase (Bvu0222).

The protein Bvu0220 was purified to homogeneity and tested again a small library of sugar lactones. Of the 25 lactones that were investigated only L-glucono-1,5-lactone, L-galactono-1,4-lactone, and L-fucono-1,4-lactone were shown to be substrates with values of $k_{cat}/K_m$ that exceeded $10^4 \text{M}^{-1}\text{s}^{-1}$ (Hobbs et al., 2014). These results confirmed the earlier conclusion that this operon encoded enzymes for the metabolism of an unidentified sugar. Since the values of $k_{cat}/K_m$ for the sugar lactones were unable to differentiate which sugar was going to be the best substrate for this particular metabolic pathway, we continued with the purification of the two dehydrogenases, in an attempt to identify the enzyme responsible for the synthesis of the lactone that would be the physiological substrate for Bvu0220. Bvu0219 was subsequently shown to catalyze the oxidation of L-galactose with $k_{cat}/K_m$ of $2 \times 10^3 \text{M}^{-1}\text{s}^{-1}$ (Hobbs et al., 2014). This value is more than 3 orders of magnitude greater than the catalytic activity for the next best substrate, L-glucose.

The reaction catalyzed by Bvu0219 was monitored by $^{13}$C NMR spectroscopy. Using $[1,^{13}$C$]$-lactose we were able to demonstrate that Bvu0219 catalyzes the formation of L-galactose to L-galactono-1,5-lactone. However, this product is thermodynamically unstable and rapidly converts non-enzymatically to L-galactono-1,4-lactone. We were subsequently able to demonstrate via NMR that the kinetic product, L-galactono-1,5-lactone was hydrolyzed by Bvu0220 with a value of $k_{cat}/K_m$ of $6 \times 10^0 \text{M}^{-1}\text{s}^{-1}$, a value that is 100-fold greater than for the hydrolysis of the thermodynamic L-galactono-1,4-lactone product.

The second dehydrogenase, Bvu0222, was tested for the oxidation of various acid sugars since L-galactonate is the product for the hydrolysis of the L-galactono-1,5-lactone. Of the acid sugars tested, L-galactonate was oxidized with a $k_{cat}/K_m$ of $2 \times 104 \text{M}^{-1}\text{s}^{-1}$, a value that is two orders of magnitude greater than the oxidation of L-glucurate or L-fuxonate (Hobbs et al., 2014). Moreover, the product of the reaction was shown to be $\beta$-tagaturonate and thus C5 of L-galactonate is oxidized. Collectively, these results identified a novel metabolic pathway for the metabolism of L-galactose where this substrate is oxidized to the 1,5-lactone by Bvu0219. This lactone is subsequently hydrolyzed to L-galactonate by Bvu0220, and then Bvu0220 oxidizes this product to $\beta$-tagaturonate. Additional enzymes can be confidently identified in B. vulgatus that are likely to be responsible for the metabolism of $\beta$-tagaturonate to pyruvate and $\beta$-glyceraldehyde-3-phosphate. This pathway is summarized in Scheme 2.

**Hydrolysis of phosphate esters**

Enzymes from cog0613 were the most recent enzymes to be interrogated from the amidohydrolase superfamily. Enzymes from the structurally related cog1387 have been shown to catalyze the hydrolysis of histidinol phosphate (Ghodge et al., 2013b). Our initial interest was Cv1693 from Chromobacterium violaceum, Elen0235 from Eggertella lenta, and TrpH from E. coli. In collaboration with the group of Steve Almo from the Einstein College of Medicine we were able to crystallize Cv1693 in the presence of AMP. Enzymes from cog0613 are one of the few cases where three divalent cations are bound in the active site (Cummings et al., 2014). The orientation of 5′-AMP in the active site initially suggested three different compounds as potential substrates: 5′-adenosine monophosphate (AMP), 3′-5′-cyclic adenosine monophosphate (cAMP), and 3′, 5′-adenosine bisphosphate (pAp). These and other compounds were tested by Jennifer Cummings for catalytic activity and only pAp was shown to be hydrolyzed by Cv1693 (Cummings et al., 2014). The reaction products were AMP and inorganic phosphate. The enzyme was not specific for the aromatic base but the
5'-phosphate group was absolutely required for catalytic activity. In the X-ray structure of Cv1693, the 5'-phosphate group was interacting with a cluster of highly conserved arginine residues in the active site. The reaction catalyzed by Cv1693 is presented in Scheme 3.

The next enzyme to be interrogated from cog0613 was Elen0235. The genomic context of the gene for this enzyme indicated that this enzyme was involved in some fashion with the metabolism of methylphosphonates. In E. coli the metabolism of phosphonates has been shown to be catalyzed by enzymes contained within the phn operon (Metcalf and Wanner, 1993). This operon consists of 14 genes extending from phnC to phnP. In E. coli the enzymes that are encoded from phnG through phnM are absolutely required for the metabolism of methylphosphonate, resulting in the synthesis of ribose-1,2-cyclic phosphate 5-phosphate (He et al., 2012). In E. coli PhnP catalyzes the hydrolysis of this compound to ribose-1,5-bisphosphate and PhnN catalyzes the phosphorylation of this compound to ribose-1-pyrophosphate-5-phosphate (Hove-Jensen et al., 2011). In E. lenta the genes phnN, phnO, and phnP are missing and replaced with the gene of unknown function from cog0613. Since Cv1693 was shown previously to catalyze the hydrolysis of pAp, we predicted that in E. lenta Elen0235 would catalyze the same reaction as that catalyzed by PhnP. This prediction was correct but not entirely accurate.

Purified Elen0235 was used to catalyze the hydrolysis of ribose-1,2-cyclic phosphate-5-phosphate and the reaction was monitored by 31P NMR spectroscopy. This compound was shown to be a substrate but the reaction products proved to be ribose-5-phosphate and inorganic phosphate (Ghodge et al., 2013a). Therefore, Elen0235 catalyzes the hydrolysis of both ester bonds to the ribose moiety instead of only catalyzing the bond to the hydroxyl group at C2 of the ribose moiety. This is the first reported instance of an enzyme being capable of hydrolyzing both bonds to a cyclic phosphate. We next addressed whether or not these bonds were catalyzed randomly or sequentially. To determine the order of hydrolysis we tested ribose-1,5-bisphosphate and ribose 2,5-bisphosphate as potential reaction intermediates. Surprisingly, ribose-1,5-bisphosphate was not detectably hydrolyzed over a period of 24 hours but ribose 2,5-bisphosphate was quantitatively hydrolyzed to ribose-5-phosphate and phosphate (Ghodge et al., 2013a). Therefore, Elen0235 catalyzes the hydrolysis of the 1,2-cyclic phosphate substrate to a 2-phosphate intermediate and then hydrolyzes this compound to ribose-5-phosphate and phosphate. The diester substrate is hydrolyzed at approximately the same rate as the monoester intermediate. We have also shown that the ribose-2,5-bisphosphate intermediate is released into solution after the first hydrolysis reaction. The overall reaction is presented in Scheme 4.

The last enzyme to be investigated was TrpH from E. coli. The gene for this enzyme is localized near the operon for the biosynthesis of tryptophan but this enzyme has apparently nothing to do with the metabolism of tryptophan. The amino acid sequence of this enzyme was reasonably similar to that of Cv1693 and thus, we initially predicted that the substrate profile for TrpH would be nearly identical to that of Cv1693. This prediction was also borne out by experiments designed to test for the hydrolysis of pAp and other bisphosphorylated substrates such as pGp (Ghodge and Raushel, 2015). However, we were intrigued by the possibility that perhaps TrpH would also be able to efficiently catalyze the hydrolysis of phosphate monoesters (pAp) but also phosphate diesters related to pAp. The appropriate compounds included short pieces of RNA that are terminated at the 5'-end with phosphate. This prediction was also borne out by experiment. TrpH was shown to catalyze the hydrolysis of short pieces of RNA that are capped with a 5'-phosphate.
Finding homes for orphan enzymes

(Ghodge and Raushel, 2015). The reaction proceeds in the 5’ to 3’-direction and thus, TrpH is an exonuclease for short single stranded pieces of RNA. DNA can also be hydrolyzed at slower rates. We have dubbed this enzyme RNase AM. Current experiments are being designed to more clearly understand how RNase AM is able to hydrolyze pAp and longer piece of RNA with nearly equal efficiency, but the closely related enzyme Cv1693 is only able to hydrolyze pAp at comparable rates. The reaction catalyzed by RNase AM is presented in Scheme 5.

Summary

The explosion in the number of DNA sequences now available in public databases has highlighted the fact that a significant number of enzymes and associated metabolic pathways remain to be discovered. We have embarked on a research program in an attempt to develop more efficient strategies for the determination of substrate profiles for enzymes of unknown function. Toward this end we have discovered the first enzyme capable of deaminating cAMP to cIMP. This activity may be important for the regulation of the concentration of cAMP in bacterial cells or it may be required for the biosynthesis of cIMP. We have also discovered a novel pathway for the metabolism of L-galactose in the human gut bacterium B. vulgatus. This discovery highlights the fact that the metabolic capabilities of the many bacterial species that inhabit to the human gut are largely unknown. Finally, we have discovered the first enzyme shown to hydrolyze both bonds to a cyclic phosphate substrate (Elen0235) and thus, this enzyme can function nearly equally well with phosphomonoester and phosphodiester substrates. This discovery led to the identification of a new previously unrecognized RNase in the well-studied bacterium E. coli.

References


Scheme 5 Reaction catalyzed by RNase AM from Escherichia coli.