

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

An investigation into catalytic promiscuity in the Enoyl CoA Hydratase/Isomerase superfamily

Permalink

<https://escholarship.org/uc/item/05z287tj>

Author

Shackelford, Grant S

Publication Date

2007

Peer reviewed|Thesis/dissertation

An investigation into catalytic promiscuity in the Enoyl CoA Hydratase/Isomerase
Superfamily

by

Grant S. Shackelford

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Abstract

**An investigation into catalytic promiscuity in the Enoyl CoA Hydratase/Isomerase
Superfamily**

Grant S. Shackelford

When subjected to environmental pressures, organisms are able to sample extensive repertoires of proteins to identify the most suitable scaffold to evolve the desired functionality. In the laboratory we cannot sample thousands of proteins to identify the most evolvable one. We must predict the best starting point. A better understanding of the avenues taken in nature to evolve new activity may facilitate the selection of promising candidates for the design of novel biocatalysts. To this end the reported work examines enzymes from the enoyl-CoA hydratase/isomerase superfamily for promiscuous activities. This line of research aims to provide insight into the prevalence of promiscuous activities amongst enzymes in nature, how such reactions are catalyzed, and the role of catalytic promiscuity in the evolution of new function. Results from this work found enoyl CoA hydratase/isomerase superfamily enzymes to be chemically specific, though occasionally exhibiting substrate promiscuity. Implications of these results are discussed.

Table of Contents

Introduction : Enzyme Specificity	p. 1
Introduction : Enzyme Promiscuity	p. 2
Introduction : Enoyl CoA Hydratase/Isomerase Superfamily	p. 5
Introduction : Identifying Promiscuity in the Enoyl CoA Hydratase/Isomerase Superfamily	p. 17
Materials and Methods	p. 19
Results	p. 24
Technical Discussion	p. 53
Discussion	p. 53
Works Cited	p. 58
Appendix	p. 62

List of Tables

Table 1: Summary of results from promiscuity analysis	p. 24
Table 2: Superfamily clones	p. 62

List of Figures

Figure 1: O-succinylbenzoate synthase promiscuity	p. 3
Figure 2: Characterized reactions in the Enoyl CoA hydratase/isomerase superfamily	p. 6
Figure 3: Structures of Enoyl CoA hydratase/isomerase superfamily enzymes	p. 7
Figure 4: Network analysis of the Enoyl CoA hydratase/isomerase superfamily	p. 18
Figure 5: Beer's law equation	p. 22
Figure 6: Enoyl CoA hydratase spectrophotometric assay	p. 23
Figure 7: Enzymatic activity formulas	p. 23
Figure 8: Enoyl CoA hydratase amplified sequenced	p. 25
Figure 9: Enoyl CoA hydratase overexpression analysis	p. 25
Figure 10: Enoyl CoA hydratase purification	p. 26
Figure 11: Enoyl CoA hydratase activity assay	p. 27
Figure 12: 4-Chlorobenzoyl CoA dehalogenase amplified and sequenced	p. 28
Figure 13: 4-Chlorobenzoyl CoA dehalogenase overexpression analysis	p. 28
Figure 14: 4-Chlorobenzoyl CoA dehalogenase purification	p. 29
Figure 15: $\Delta^{3,5}$ - Dodecenoyl-CoA Isomerase amplified and sequenced	p. 30
Figure 16: $\Delta^{3,5}$ - Dodecenoyl-CoA Isomerase overexpression analysis	p. 30
Figure 17: $\Delta^{3,5}$ - Dodecenoyl-CoA Isomerase purification	p. 31
Figure 18: Histone acetyltransferase amplified and sequenced	p. 32
Figure 19: Histone acetyltransferase overexpression analysis	p. 32
Figure 20: Histone acetyltransferase purification	p. 33
Figure 21: Crotonobetainyl CoA hydratase amplified and sequenced	p. 34
Figure 22: Crotonobetainyl CoA hydratase overexpression analysis	p. 34
Figure 23: Crotonobetainyl CoA hydratase purification	p. 35

Figure 24: Crotonobetainyl CoA hydratase activity assay	p. 36
Figure 25: 3-Hydroxyisobutyryl CoA hydrolase amplified and sequenced	p. 37
Figure 26: 3-Hydroxyisobutyryl CoA hydrolase overexpression analysis	p. 37
Figure 27: 3-Hydroxyisobutyryl CoA hydrolase purification	p. 38
Figure 28: Methylmalonyl CoA decarboxylase amplified and sequenced	p. 39
Figure 29: Methylmalonyl CoA decarboxylase overexpression analysis	p. 39
Figure 30: Methylmalonyl CoA decarboxylase purification	p. 40
Figure 31: Methylglutaconyl CoA hydratase amplified and sequenced	p. 41
Figure 32: Methylglutaconyl CoA hydratase overexpression analysis	p. 41
Figure 33: Methylglutaconyl CoA hydratase purification	p. 42
Figure 34: Feruloyl-CoA Hydratase/Lyase amplified and sequenced	p. 43
Figure 35: Feruloyl-CoA Hydratase/Lyase overexpression analysis	p. 43
Figure 36: Feruloyl-CoA Hydratase/Lyase purification	p. 44
Figure 37: Cyclohex-1-enecarboxyl-CoA hydratase amplified and sequenced	p. 45
Figure 38: Cyclohex-1-enecarboxyl-CoA hydratase overexpression analysis	p. 45
Figure 39: Cyclohex-1-enecarboxyl-CoA hydratase purification	p. 46
Figure 40: 2-Ketocyclohexanecarboxyl-CoA hydrolase amplified and sequenced	p. 47
Figure 41: 2-Ketocyclohexanecarboxyl-CoA hydrolase overexpression analysis	p. 47
Figure 42: 2-Ketocyclohexanecarboxyl-CoA hydrolase purification	p. 48
Figure 43: Cyclohexa-1,5-dienecarbonyl-CoA hydratase amplified and sequenced	p. 49
Figure 44: Cyclohexa-1,5-dienecarbonyl-CoA hydratase overexpression analysis	p. 49
Figure 45: Cyclohexa-1,5-dienecarbonyl-CoA hydratase purification	p. 50
Figure 46: 1,4-dihydroxy-2-napthoyl CoA synthase amplified and sequenced	p. 51
Figure 47: 1,4-dihydroxy-2-napthoyl CoA synthase overexpression analysis	p. 51
Figure 48: 1,4-dihydroxy-2-napthoyl CoA synthase purification	p. 52

Introduction

Enzymatic Specificity

Enzymes are expected to be specific, though cases have been identified of enzymes promiscuously catalyzing reactions in addition to their native one. Phosphotriesterase, for example, catalyzes the hydrolysis of phosphate bonds through stabilization of a trigonal bipyramidal intermediate, though it also catalyzes the hydrolysis of amide bonds through a tetrahedral intermediate.¹

Biological systems require enzymes to convert available substrates into usable products for processes such as metabolism, cellular signaling, and biosynthesis. Contingent within this notion is the ability of an enzyme to select the appropriate substrate to yield the necessary product. If enzymes indiscriminately catalyze reactions, biological systems would become unbalanced and collapse.

One striking example of catalytic specificity comes from t-RNA synthases. These enzymes are involved in protein biosynthesis and must differentiate between twenty amino acids, some of which differ by one methylene group. For example, valine differs from isoleucine by one methyl group, though binds to isoleucine t-RNA synthase 150-fold weaker than isoleucine. Alanyl-tRNA synthase discriminates against glycine by a factor of 250. Threonine binds to valyl-tRNA synthase 100-200 times more weakly than valine, and isoleucine has $2E^5$ less affinity.² This ability to selectively bind a particular substrate is common throughout biology where many enzymes exhibit regiospecificity, chemospecificity, and stereospecificity. Beyond the necessity for specificity in a biological context, specific interactions between a substrate and an enzyme are also necessary for catalysis.

Chemical transformations require energy to break and form bonds. A primary reason enzymes catalyze reactions is because they provide a fraction of the energy necessary for chemical reactions. This energy is derived from enthalpic binding interactions and entropic gain that occurs when a substrate is removed from bulk solvent. Additionally, enzymes have precisely placed amino acids that promote chemical transformations by attacking the substrate directly or activating a molecule to do so. Thus, a second requirement for specificity is to facilitate catalysis of chemical reactions. Between the biological and thermodynamic needs for specificity, it is counterintuitive for an enzyme to catalyze multiple reactions.

Regardless of these necessities, enzymes are not always entirely specific; they sometimes exhibit promiscuity. This raises questions concerning the frequency of these activities, how they are catalyzed, and how they are managed *in vivo*. To begin understanding these reactions, the identified cases have been divided into two general groups: enzymes that perform the same chemistry on different substrates and enzymes that catalyze multiple chemical reactions within one active site.

Enzymatic Promiscuity

Substrate promiscuity is the most identified source of promiscuity. One extreme case is methane mono-oxygenase. This enzyme catalyzes the hydrolysis of 150 substrates in addition to its primary substrate methane.³ Another form of substrate promiscuity arises when a ligand is held in different orientations, causing chemical reactions to occur at different sites. An example of this type of promiscuity is γ -humelene synthase, which catalyzes different cyclization mechanisms on farnesyl diphosphate to produce 52 different sesquiterpenes.⁴

Cases of catalytic promiscuity have been identified that extend beyond changes in substrate to affect chemical promiscuity. One type of chemical promiscuity entails enzymes that catalyze different chemical reactions while using the same catalytic residues. For example, O-succinyl benzoate synthase has been shown to also catalyze the racemization of N-acyl amino acids at rates approaching the native reaction. These two reactions involve different substrates and chemical mechanisms. (Figure 1) The native reaction involves the cyclization of 2-succinyl-6R-hydroxy-2,4-cyclohexadiene-1R-carboxylate to form o-succinylbenzoate, and the promiscuous reaction involves the racemization of N-acyl amino acids. Both of these chemical reactions are accomplished utilizing the same set of catalytic residues.⁵

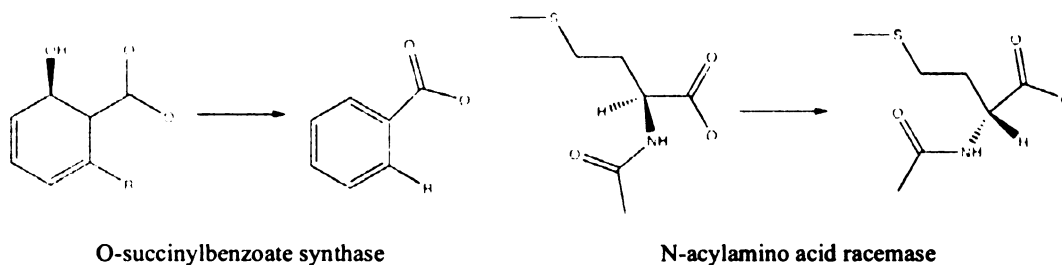


Figure 1: O-succinylbenzoate synthase catalyzes its native reaction with a k_{cat}/K_M value of $2.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and the promiscuous racemization of N-acylamino acids with a k_{cat}/K_M value of $3.7 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$.

Catalytic promiscuity has also been identified where an enzyme catalyzes two chemical reactions using unique residues. Within this class two sources of promiscuity have been identified. Some enzymes catalyze different reactions using a common active site feature for part of the reaction and unique residues for the remainder. Lipase B utilizes a serine-histidine-aspartate catalytic triad and an oxyanion hole to catalyze the hydrolysis of lipid molecules. This enzyme also promiscuously catalyzes the synthesis of carbon-carbon bonds with the aid of the oxyanion hole, but without the catalytic triad.⁶ Additionally,

changes in enzyme conformation have been proposed as a possible source of promiscuity. In this scenario an enzyme's active site adopts different orientations to catalyze different reactions. For example, isopropylmalate isomerase has a flexible loop that in one configuration enables the enzyme to bind isopropylmalate and catalyze the shifting of a double bond therein and in another orientation bind homocitrate and catalyze the homoaconitase reaction.⁷

What benefit exists for an organism that houses promiscuous enzymes? One hypothesis proposes that new enzymes may emerge from the promiscuous activity of another.⁸ Because the number of protein folds are limited (one estimate is less than two-thousand)⁹ folds may have been reused through divergent evolution to provide the catalytic range found in nature. Analysis of divergently evolved enzymes indicates that many homologues exhibit a common chemical capability and differing substrate specificity and overall reactions.¹⁰ Such proteins retain a constellation of amino acids that catalyze a partial reaction, but vary in residues to determine substrate specificity and mediate additional catalytic steps with reactive intermediates to produce products. Although chemistry-constrained divergent evolution may be a strategy by which nature can minimize the number of mutations necessary to alter activity, the likelihood of several beneficial mutations accumulating in the absence of positive selection is low. If instead an organism housed an enzyme that promiscuously catalyzed the desired reaction, then the likelihood of that enzyme developing specificity for said activity is high. Promiscuous activities have been found to provide suitable starting points for engineering, often showing increases in activity over several orders of magnitude with a few mutations.¹¹

A notable study published by Aharoni *et al.* enhanced several promiscuous activities to examine the initial stage of evolution of a promiscuous activity. They found the mutations that enhanced the rate of the promiscuous functions occurred at the perimeter of the active site, refining substrate positioning without altering catalytic machinery. Accordingly, the native functions were largely unaffected by the mutations, dropping by 1-2 fold in six experiments, while the promiscuous activities were enhanced by up to 500-fold.¹² These results indicate that promiscuous activities may be enhanced in biological systems prior to gene duplication, allowing the enzyme to continue performing its primary function while adapting to an external pressure.

Studying catalytic promiscuity has potential for practical applications and academic impact. The promiscuous capabilities of a relatively small set of proteins may have led to the diversity of enzymatic function. Investigation into the prevalence of promiscuous activities amongst evolutionarily related enzymes might provide insight into how life has evolved into its current diversity. A deeper understanding of the properties of promiscuous activities may help to identify or create better scaffolds for engineering new macromolecules for applications such as biosensors, therapeutics, and bioremediation. Additionally, the promiscuous activities of some enzymes have been shown to convert man-made chemicals into toxic byproducts.¹³ A more thorough catalog of promiscuous activities and appreciation for how these reactions are catalyzed may help to avoid such toxicity. The reported work aims to investigate catalytic promiscuity by identifying cases in the enoyl-CoA hydratase/isomerase superfamily and examining their mechanism of catalysis.

The Enoyl-CoA hydratase/isomerase superfamily

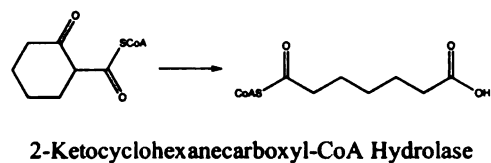
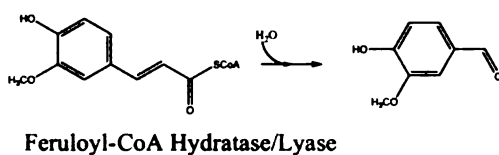
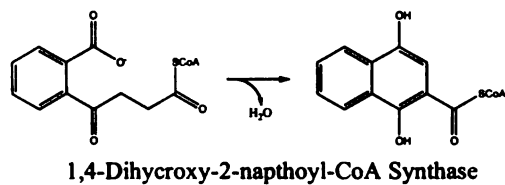
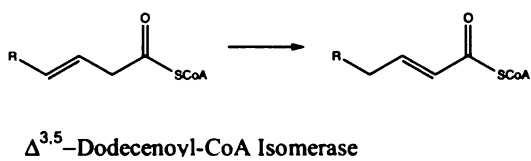
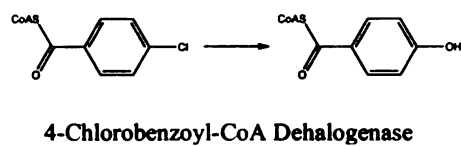
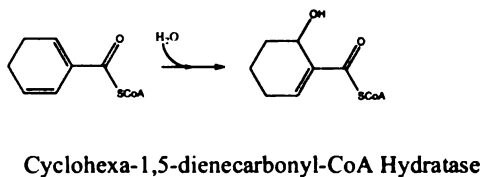
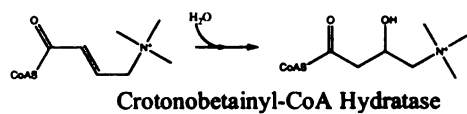
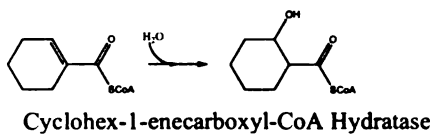
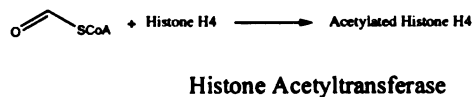
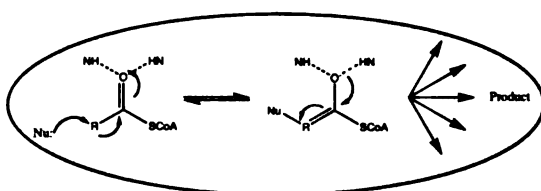
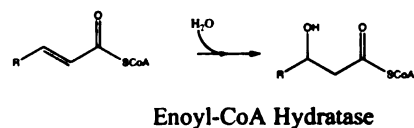
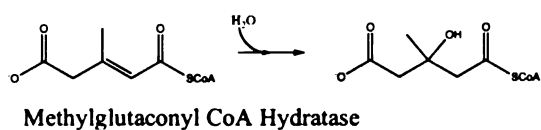
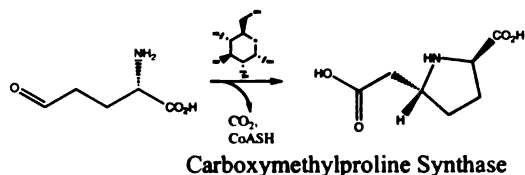
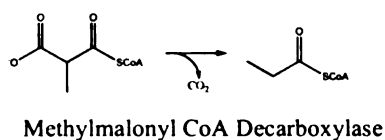
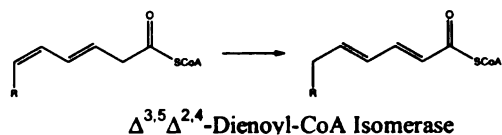
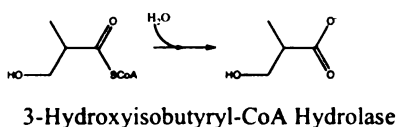


Figure 2: Enzymes in the enoyl CoA hydratase/isomerase superfamily catalyze a broad range of chemical reactions on diverse substrates that all proceed through an anionic intermediate that is stabilized by a structurally conserved oxyanion hole.

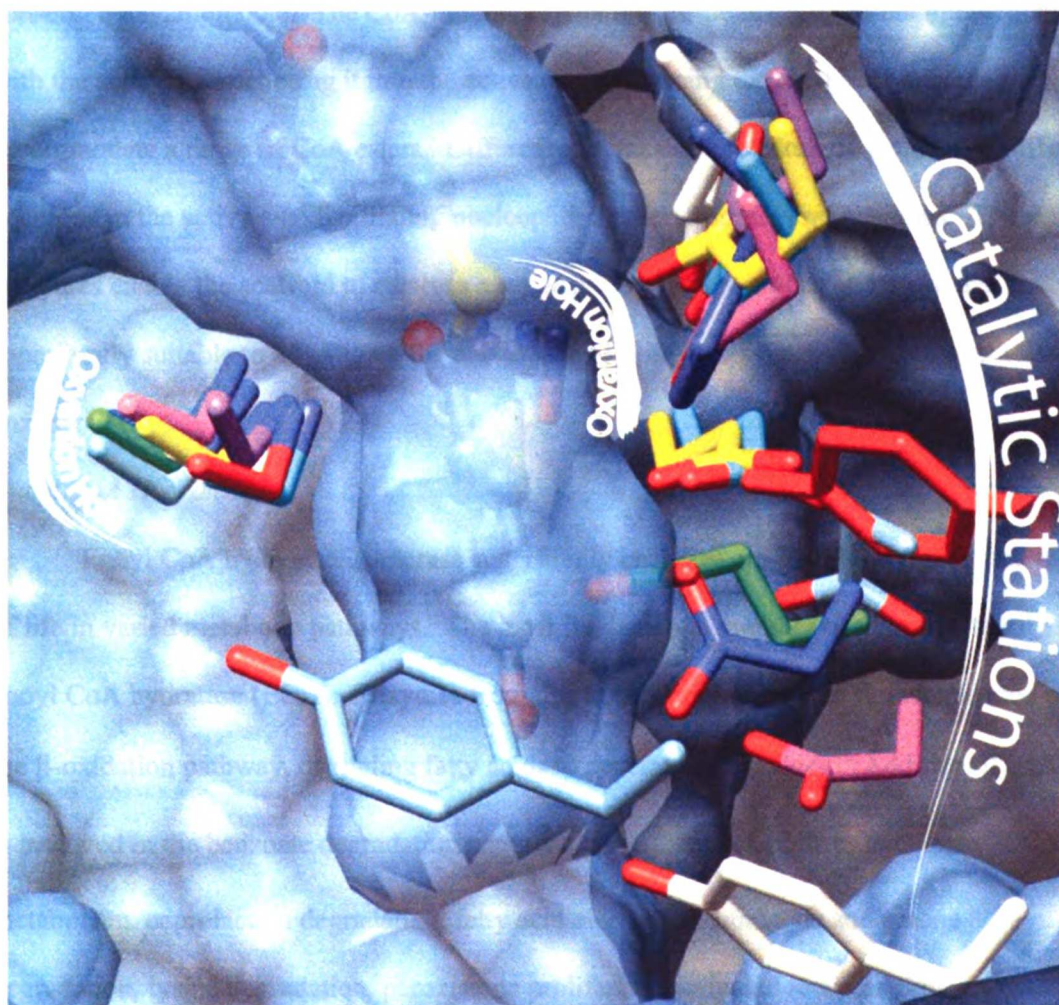
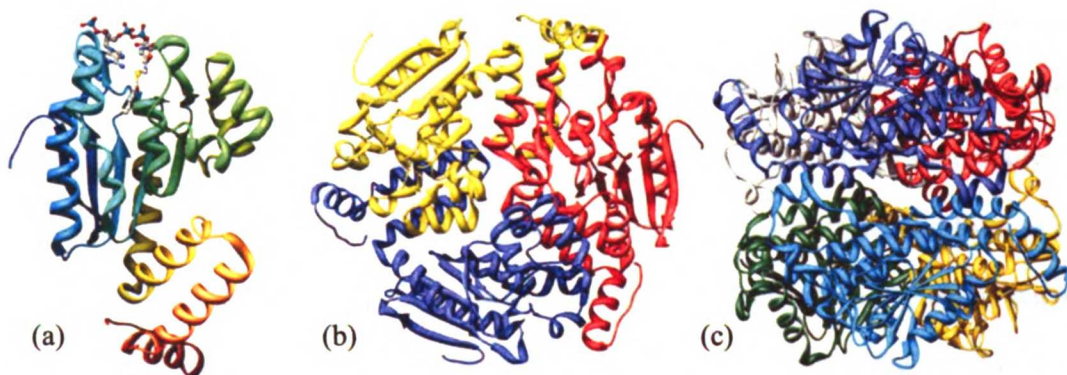


Figure 3: (a) Monomeric side view of ECH. (b) Quarternary top view of ECH. (c) Quarternary side view of ECH. (d) Catalytic residues of all structurally characterized superfamily members superimposed on the active site background of 4CBD. Superimposed structures are distinguished by the following colors: FHL in white, DCI in magenta, ECH in cyan, MGCH in yellow, MMCD in red, 4CBD in blue, DDI in green, CarB in purple and MenB in sky blue.

The enoyl-CoA hydratase/isomerase superfamily is composed of fifteen characterized families (and additional proteins of unknown function) that catalyze a broad set of reactions involving diverse substrates. These enzymes are all related by the use of two structurally conserved backbone amide groups that stabilize an anionic intermediate derived from a Co-enzyme-A (CoA) ester substrate.¹⁴ The members of this superfamily have analogous active site architectures composed of a CoA-binding site and a deep catalytic pocket containing a conserved oxyanion hole. The CoA-binding region makes the majority of the interactions with the substrate, anchoring it into the active site and freeing the catalytic pocket to accommodate a range of CoA esters. Different enzymes station catalytic residues in different positions in the active site, facilitating nucleophilic attack from several positions. These attributes enable the scaffold to catalyze a range of chemical reactions and make the superfamily suitable to investigate the prevalence of promiscuous activities amongst related enzymes.

Enoyl CoA Hydratase

Enoyl CoA hydratase/isomerase superfamily enzymes are found throughout the tree of life in varied metabolic pathways. The most fundamental member of this superfamily is enoyl CoA hydratase (ECH). Enzymes from this family are involved in the second step of the β -oxidation pathway, degrading fatty acids to generate acetyl CoA.¹⁵ Additionally, ECH is involved in the benzoate degradation pathway, β -Alanine metabolism, butanoate metabolism, caprolactam degradation, fatty acid elongation, limonene and pinene degradation, lysine degradation, peroxisome proliferator-activated receptor signaling pathway, propanoate metabolism, tryptophan metabolism, and branched chain amino acid

degradation. Within these pathways ECH catalyzes the stereospecific hydration of trans-2-enoyl-CoA thioesters to the corresponding 3(S)-hydroxyacyl-CoA compound.¹⁶

Mechanistic and structural studies have elucidated the details of this reaction.

Structural work has shown that ECH folds into a dimer of trimers, where the C-terminal of one monomer encloses the active site of its neighbor within the trimer.¹⁷ Mutagenesis studies identified Glutamate¹⁴⁴ and Glutamate¹⁶⁴ as the operative catalytic residues that activate a water molecule for nucleophilic attack and donate a proton to the intermediate to yield product. While both residues are necessary for full activity, these studies have shown that the reaction is still catalyzed at a low level with only one.¹⁸

4-Chlorobenzoyl CoA Dehalogenase

In addition to fatty acids, several enzymes from the superfamily catalyze reactions upon cyclic compounds. 4-Chlorobenzoyl-CoA Dehalogenase (4CBD) is the most well known of these. This enzyme was isolated from soil bacteria capable of growing on 4-chlorobenzoate as a sole carbon source. 4CBD catalyzes the hydrolysis of para-substituted halogens on benzoyl-CoA substrates to form 4-hydroxybenzoyl-CoA. This compound is hydrolyzed into hydroxybenzoate, which is metabolized by the β -ketoacid pathway to generate energy intermediates for oxidative respiration.¹⁹ Mechanistic and structural studies have identified Aspartate¹⁴² as the catalytic nucleophile that attacks the substrate at the para position, generating a Meisenheimer intermediate that is stabilized by the enzyme's oxyanion hole. This anionic compound then regains aromaticity and expels a chloride ion. The enzyme-aryl intermediate is then attacked by an activated water, freeing the enzyme and producing 4-hydroxybenzoyl-CoA.²⁰

Structural work has shown that 4CBD folds into a trimer in solution, where the C-terminal of one monomer encloses the active site of its neighbor.²¹ Additionally, this work has identified other mechanistically important residues. Tryptophan¹³⁷ is predicted to make hydrogen bond interactions with the aryl-intermediate, activating and positioning it for attack by a water molecule. Histidine⁹⁰ has also been shown to be necessary for catalysis, and is hypothesized to close the bottom of the active site from solvent.²²

Carboxymethylproline synthase

Carboxymethylproline synthase (CarB) is a medically important member of the superfamily. This enzyme is involved in the biosynthesis of carbapenem β -lactam antibiotics such as penicillins and cephalosporins. In this pathway CarB catalyzes the formation of carboxymethylproline from malonyl-CoA and L-glutamate semialdehyde. This is accomplished through the formation of C-C and C-N bonds, decarboxylation and thioester hydrolysis. Other members of the superfamily catalyze these individual reactions, though CarB is the only enzyme capable of catalyzing them all.²³

Structural studies depict CarB as a trimeric structure with C-terminal domains that enclose the active site of their own monomer. As with all other members of the superfamily, CarB utilizes an oxyanion hole formed from the backbone amide groups of Glycine⁶² and Methionine¹⁰⁸ for substrate activation and intermediate stabilization. Structural analysis also identified Glutamate¹³¹, which may serve as a catalytic nucleophile and is structurally equivalent to Aspartate¹⁴² from 4CBD.²⁴

$\Delta^{3,5}$, $\Delta^{2,4}$ Dienoyl CoA Isomerase

While ECH and the other enzymes in the β -oxidation pathway degrade saturated fatty acids, auxiliary enzymes are necessary to convert unsaturated fatty acids into acceptable

substrates. One such enzyme is $\Delta^{3,5}$, $\Delta^{2,4}$ dienoyl CoA isomerase (DCI), which catalyzes the isomerization of 3,5-dienoyl CoA to 2,4-dienoyl-CoA. This fatty acid is then reduced and isomerized to generate 2-enoyl-CoA, the substrate for ECH.²⁵ Enzymes from this family have been identified in the mitochondria and peroxisomes of eukaryotes and several species of prokaryotes.²⁶

The three-dimensional structure of this enzyme has been solved and depicts DCI as a dimer of trimers with a deep hydrophobic pocket and a C-terminal loop that covers the active site of its neighboring subunit within the trimer. Mutagenesis studies have shown that replacement of Glutamate¹⁹⁶ and Aspartate²⁰⁴ with neutral residues mitigates the reaction rate by a factor of 10^5 . Interestingly, Glutamate¹⁹⁶ is found in the same position as Glutamate¹⁶⁴ in ECH while Aspartate²⁰⁴ is homologous to Aspartate¹⁴² from 4CBD.²⁷ In addition to DCI, Δ^3, Δ^2 dodecenoyl CoA isomerase (DDI) is also involved in the metabolism of unsaturated fatty acids.

Δ^3, Δ^2 Dodecenoyl CoA Isomerase

DDI catalyzes the isomerization of 3-trans enoyl-CoA into 2-trans enoyl-CoA – the substrate for ECH in the β -oxidation pathway. This family of enzymes can be grouped into three classes: monofunctional mitochondrial, monofunctional peroxisomal, and multifunctional (the N terminal has isomerase and hydratase activity while C terminal has dehydrogenase activity). The monofunctional enzymes have been identified in eukaryotic organisms while the multifunctional enzymes exist in bacteria and the peroxisomes of eukaryotes. Both the mitochondrial and peroxisomal monofunctional enzymes have been crystallized and found to fold into trimers with C-terminal domains that cover the active site of their own subunit rather than crossing onto another monomer as found in other members

of the superfamily. The isomerase reaction involves the shifting of a double bond by proton abstraction from the C-2 atom leading to an anionic intermediate, which collapses to form product following proton donation to atom C-4. This mechanism is catalyzed by one Glutamate residue. In the peroxisomal variant this residue is Glutamate¹⁵⁸, which is analogous to the catalytic Aspartate²⁰⁴ in DCI and Aspartate¹⁴² in 4CBD. The mitochondrial homologue catalyzes its reaction by using Glutamate¹⁵⁰, which is equivalent to Glutamate¹⁶⁴ in ECH and Glutamate¹⁹⁶ in DCI.²⁸

Histone Acetyltransferase

Enzymes from the superfamily also catalyze reactions outside of central metabolism. Histone acetyltransferase (HisA) has been shown to be involved in the chromatin remodeling process during spermatogenesis. This enzyme is involved in the hyperacetylation of histone H4 subunits to facilitate the exchange of these molecules with protamines. This process is necessary for the maturation of spermatids into spermatozoa.²⁹ Further research into the mechanism of this enzyme may provide insight into the chromatin remodeling process and lead to the design of drugs to control spermatogenesis.

Crotonobetainyl CoA Hydratase

Several species of *enterobacteriaceae* are able to metabolize carnitine via crotonobetaine to produce γ -butyrobetaine under anaerobic conditions. In this pathway crotonobetainyl-CoA hydratase (CaiD) catalyzes the reversible hydration of L-carnitiny-CoA to crotonobetainyl-CoA, which is then hydrolyzed to form γ -butyrobetaine.³⁰ The metabolic context of this pathway is currently unknown.

3-Hydroxyisobutyryl CoA Hydrolase

Another auxiliary enzyme family is 3-Hydroxyisobutyryl CoA Hydrolase (HICH); these enzymes are involved in valine catabolism in mammals and plants to generate metabolites for oxidative respiration. During this pathway methacrylyl-CoA is produced, which reacts with nucleophiles such as protein cysteines and cofactors CoA and glutathione. To mitigate the accumulation of this toxic compound, it is reversibly hydrated and then hydrolyzed by HICH to β -hydroxyisobutyrate. This compound is non-toxic and transported for further metabolism. Mechanistic studies have identified two catalytic residues for this reaction: Glutamate¹⁴¹, which aligns with Glutamate¹⁶⁴ from ECH and Aspartate¹⁴⁹. Mutation of Aspartate¹⁴⁹ to Glycine resulted in a 97% loss of activity relative to wild type, while an E141A mutation affected a complete loss of activity.³¹ In accordance with mutagenic data and sequence alignments, Glutamate¹⁴¹ is predicted to generate the enolate intermediate by attacking the thioester carbonyl, which is hydrolyzed by a water molecule activated by Aspartate¹⁴⁹.³²

Methylmalonyl CoA Decarboxylase

Similarly to HICH, methylmalonyl CoA decarboxylase (MMCD) is involved in the synthesis of propionate though through a different pathway. This enzyme is part of a four-gene operon encoding enzymes to convert succinate into propionate. The metabolic context of this pathway is unknown, though there are multiple pathways to degrade propionate into metabolites for the citric acid cycle.

Mechanistic and structural studies have been published on this enzyme. The decarboxylation of methylmalonyl CoA is proposed to begin with the transfer of electrons from the substrate's carboxylate to its thioester group, forming an anionic intermediate. The intermediate then collapses, abstracts a proton from the enzyme, and expels the substrate's

carboxylate, yielding product. Mutagenesis studies have shown that Tyrosine¹⁷⁰ is necessary for catalysis and is proposed to orient the substrate's carboxylate group through hydrogen-bonding interactions. Additionally important are Histidine⁶⁶ and Glycine¹¹⁰, which form the oxyanion hole that activate the substrate and stabilize the intermediate.³³ Interestingly, the enzyme has an active site Glutamate that is analogous to Glutamate¹⁴⁴ in ECH, but does not appear to interact with the substrate. In addition to identifying catalytic residues, structural studies have shown that this enzyme forms a dimer of trimers with each C-terminal covering the active site of its own subunit.³⁴

Methylglutaconyl CoA Hydratase

Methylglutaconyl-CoA hydratase (MGCH) catalyzes the hydration of 3-methylglutaconyl-CoA to form 3-hydroxy-3-methylglutaconyl-CoA. This activity has been identified in the leucine catabolic pathway of humans and bacteria. While both enzymes function in leucine catabolism, the human homologue also has an RNA binding region of unknown function.

Structural studies depict mammalian methylglutaconyl CoA hydratase as a dimer of trimers with C-terminal domains that fold onto the active site of neighboring subunits within the trimer. While both homologues catalyze the same reaction they differ in their catalytic residues.³⁵ The mammalian enzyme has two glutamate residues in positions homologous to those in ECH while the bacterial one has only an analog to Glutamate¹⁶⁴. To facilitate the activation of a water molecule, these studies suggest that the bacterial enzyme utilizes the terminal carboxylate in the substrate in place of an additional glutamic acid residue.³⁶

Ferulyl CoA Hydratase/lyase

Ferulyl-CoA hydratase/lyase (FHL) is another industrially relevant enzyme in the enoyl CoA hydratase/isomerase superfamily. These enzymes catalyze the formation of vanillin and acetyl CoA through the degradation of lignin polymers from plant debris. Thus, these enzymes have utility both for biodegradation of plant waste and the formation of the flavoring agent vanillin.³⁷

The three-dimensional structure of this enzyme identifies it as a trimer with C-terminal domains that complete neighboring active sites. Comparison of this structure with those of other superfamily members has led to the prediction of mechanistically-important residues. The oxyanion hole of this enzyme is formed by the backbone amide groups of Methionine⁷⁰ and Glycine¹²⁰. The proposed reaction mechanism uses Glutamate¹⁴³ as a catalytic nucleophile, which is homologous to Glutamate¹⁶⁴ from ECH, to activate a water molecule that attacks the substrate and generates an anionic intermediate. This intermediate then collapses and abstracts a proton from the enzyme. A proton is abstracted from the substrate's hydroxyl group, leading to carbonyl formation and release of the acetyl CoA group. Tyrosine²³⁹ is predicted to play a role in this mechanism, though further research is necessary to elucidate its contribution.³⁸

Benzoic acid degradation enzymes

Aromatic compounds are often converted into benzoyl-CoA, which is further metabolized by the benzoate degradation pathway to generate 3 equivalents of acetyl-CoA and 1 CO₂.³⁹ Cyclohex-1-enecarboxyl-CoA hydratase (BadI), cyclohexa-1,5-dienecarbonyl-CoA hydratase (1,5CH) and 2-ketocyclohexanecarboxyl-CoA hydrolase (BadK) all function in this pathway within anaerobic bacteria to assimilate aromatic compounds. Analysis of the enzymes constituting the benzoate degradation pathway has implications for bioremediation

of toxic aromatic hydrocarbons, as well as recycling aromatic plant polymers. Within this pathway BadK and 1,5CH catalyze the hydration of benzoic acid compounds, while BadI occurs toward the end of the pathway and breaks the ring open.⁴⁰ A mechanistic study has been performed for BadI that identified Serine¹³⁸, Aspartate¹⁴⁰ (homologous to Glutamate¹⁶⁴ in ECH) and Tyrosine²³⁵ as the catalytic residues, though their mechanistic functions are unknown.⁴¹

1,4-dihydroxy-2-naphthoyl-CoA synthase

1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB) catalyzes the condensation of o-succinylbenzoyl-CoA to 1,4-dihydroxy-2-naphthoyl-CoA, which is the reverse of the reaction catalyzed by BadI. This enzyme is involved in the menaquinone (vitamin K) biosynthetic pathway found in gram-negative bacteria. In bacteria vitamin K is utilized under anaerobic conditions to shuttle electrons between dehydrogenases in the inner mitochondria. Vitamin K is also important in humans where it is involved in blood coagulation. Because this pathway is only found in gram-negative bacteria, the enzymes that constitute it are potential drug targets.⁴²

Structural and mechanistic studies of this enzyme have shown that it forms a dimer of trimers with a deep hydrophobic pocket and oxyanion hole akin to other structurally-characterized members of the superfamily. Unique to this enzyme is the positioning of its C-terminal domain, which covers the active site of a subunit from a different trimer. The proposed catalytic residues are Aspartate¹⁹², which is analogous to Glutamate¹⁴⁴ in ECH and Tyrosine²⁸⁷. Interestingly, according to sequence alignments both of these residues appear to be conserved in BADI. Mutagenesis studies have shown that these residues are necessary for

function and are hypothesized to participate in proton transfer steps at the beginning and end of the mechanism.⁴³

Identifying promiscuity in the Enoyl-CoA Hydratase/Isomerase Superfamily

If functional promiscuity assists the evolution of new activities, a group of related enzymes may contain proteins that catalyze each other's reactions. In the enoyl CoA hydratase/isomerase superfamily, MGCH and CaiD have already been shown to catalyze the similar ECH reaction.^{31,35} Both of these enzymes use the two glutamate residues for their native reaction. 4CBD, which lacks these catalytic glutamates does not exhibit promiscuous ECH activity, though was engineered to catalyze the reaction by inclusion of two glutamate residues.⁴⁴ These cases of functional plasticity suggest that the barrier to promiscuity within the enoyl-CoA hydratase/isomerase superfamily may largely be placement of appropriate catalytic residues.

Because all superfamily members share fundamental aspects of their structures and catalytic mechanisms, additional promiscuous activities may exist. To investigate catalytic promiscuity in the enoyl CoA hydratase/isomerase superfamily, enzymes have been cloned into histidine-tagged protein expression vectors, over-expressed and purified using nickel-affinity chromatography. The purified enzymes were characterized for promiscuous catalysis of the ECH reaction. This reaction was chosen for several reasons. ECH is a fundamental and ubiquitous enzyme found throughout all forms of life, functioning in β -oxidation as well as many other pathways. Because of its role in primary metabolism, it is hypothesized to be the progenitor of the fourteen other families.

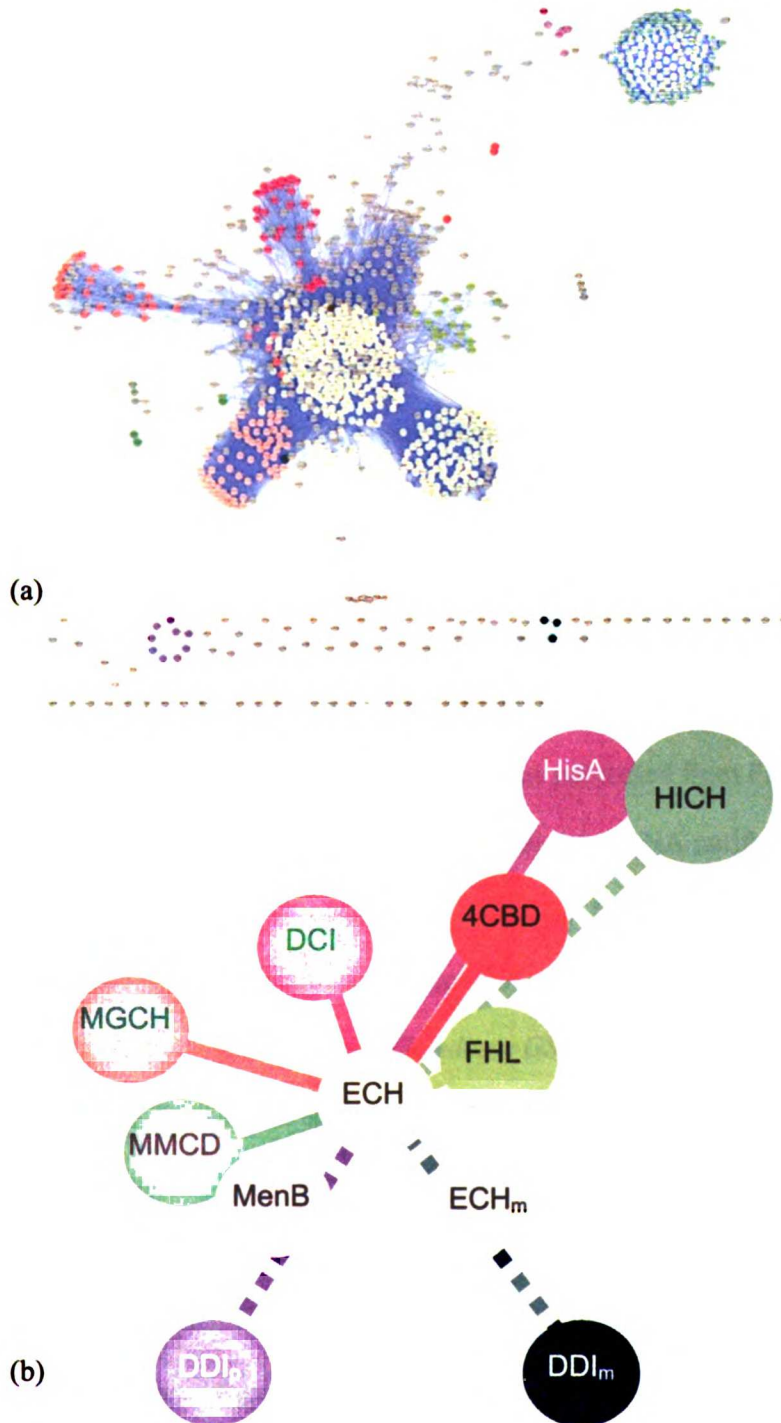


Figure 4: Network analysis of the enoyl CoA hydratase/isomerase superfamily. (a) Nodes depict sequences and are colored according to their family designation; blue lines are edges between nodes with sequence homology scores at or below 10^{-30} . (b) Solid lines indicate homology scores between families of 10^{-30} ; dashed lines indicate expectation values above or equal to 10^{-27} . ECH_m denotes the hydratase domain on the multifunctional β -oxidation protein. DDI_p and DDI_m are the peroxisomal and mitochondrial homologues respectively.

To investigate the evolutionary relationship amongst families in the superfamily, a network analysis was performed (Figure 4). This study identified ECH as the central family to which all other families exhibit sequence homology with an expectation value below 10^{-27} . We hypothesized that a likely source of promiscuity could be retention of a progenitor's activity, making ECH an ideal reaction with which to identify promiscuity. In addition to evolutionary relationships, ECH was chosen because it is the most well characterized enzyme in the superfamily, facilitating the analysis of any identified promiscuous reactions.

Materials

All reagents were purchased from Sigma-Aldrich. As reported by the manufacturer, crotonyl-CoA contained 5-10% cis isomer and was used without further purification. Bicinchoninic acid (BCA) protein assay reagents were purchased from Pierce. BPER reagent was purchased from Pierce. Phusion DNA polymerase was purchased from Finnzymes. Restriction enzymes were purchased from New England Biolabs, DNA purification kits from Qiagen, and T4 DNA ligase from Invitrogen. 5 mL nickel fast flow liquid chromatography columns were purchased from GE Healthcare. *Saccharomyces cerevisiae* genomic DNA was purchased from Invitrogen and *geobacter metallirudcens GS-15* genomic DNA was purchased from American Type Culture Collection. *Pseudomonas syringae* genomic DNA, *Escherichia coli* K-12 genomic DNA, and the PHUG60 plasmid were generous gifts from the Lindow (<http://nature.berkeley.edu/icelab>), Voigt (<http://www.voigtlab.ucsf.edu>), and Sylvestre (Michel.Sylvestre@iaf.inrs.ca) laboratories, respectively. The gene encoding HisA was purchased from American Type Culture Collection in the pRSET(B) vector. BADK in the pDPHisK1 vector was a gift from the Harwood laboratory (csh5@u.washington.edu). The Tonge laboratory (<http://ms.cc.sunysb.edu/~ptonge>) provided MenB and ECH, both in

pET15b vectors. The Gerlt laboratory (j-gerlt@uiuc.edu) provided BADI in pET15b, MGCH in pET16b, MMCD in Tom15b and HICH in Tom15b.

Methods

Superfamily network analysis

To gain insight into evolutionary relationships between members of the enoyl CoA hydratase/isomerase superfamily the program Cytoscape⁴⁵ was used to construct a sequence-based protein network. Sequences used for this analysis were taken from the Gold Standard set in the Structure-Function Linkage Database, which limits superfamily members to sequences containing the conserved fold and active site motif. These sequences were subdivided into families based on family-specific active site residues.¹⁰ Following a 95% sequence homology filter the resulting set of sequences were subjected to an all-by-all alignment to calculate expectation values (E-values) for every sequence pair. These E-values were used as input for Cytoscape to generate a protein network. To resolve the network into family clusters, edges between nodes were restricted to sequence pairs with E-values below 10^{-27} .

Cloning

Genes were amplified from DNA using Phusion DNA polymerase according to the manufacturer's protocol. Polymerase chain reaction (PCR) was initiated with a 2-minute denaturation step at 94°C, followed by 35 cycles of 1 minute at 94°C, annealing at 60°C for 1 minute and 15 seconds, and fifteen seconds at 70°C. Following 35 rounds samples were subjected to a 10-minute elongation step at 70°C. PCR products were then stored at 4°C. Following amplification PCR products were purified and restriction digested between 4 and 16 hours alongside the corresponding vector DNA. Double-digested fragments were purified

through gel extraction and ligated overnight at 16°C using T4 DNA ligase. Ligated constructs were electroporated into competent cells and streaked on Luria-Bertani (LB) plates containing 50 mg/mL Carbenicillin. Plasmid DNA from a selection of resultant colonies was then sequenced to identify cloned genes.

Overexpression

Plasmids were transformed into BL21(DE3) cells. Prior to purification, all constructs were examined for overexpression of their target gene. To this end cells were grown in 250 mL of (LB) containing 50 mg/mL carbenicillin until they reached log phase, which was identified by an optical density of 0.6 to 0.8 absorbance units at 600 nanometers. Cultures were then halved; one sample received 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) to induce protein overexpression while the other did not. Both cultures were grown for an additional 16 hours. Following this period cultures were centrifuged at 13,000 rpm for 10 minutes, chemically lysed using BPER reagent and visualized using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Following the overexpression analysis, bacteria carrying plasmid encoding superfamily enzymes were grown in 1L of LB at 37°C with 50 mg/mL of carbenicillin. Once cells reached log phase, they were treated in accordance with the determined conditions to maximize production of soluble protein. Following this period cells were centrifuged at 6,000 rpm for 30 minutes. The resultant supernatant was decanted and the pellet resuspended in 40 mL of nickel binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) and 1 mM phenylmethanesulphonylfluoride to inhibit serine proteases. The cells were lysed by 6 passages through a microfluidizer and the resultant debris pelleted by centrifugation at

30,000xg for 30 minutes. Clarified lysate was filtered through a 20-micron filter prior to affinity purification.

Purification

Protein purification was carried out using a 5mL fast flow Nickel column. Following loading of clarified lysate, protein was washed and then eluted using a 100 mL gradient of histidine-elution buffer (500 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9). Fractions containing protein were immediately pooled and dialyzed against 4L of dialysis buffer (20 mM KH₂PO₄, 3 mM EDTA, pH 7.4 and 1 mM DTT) for 24 hours. Purity was assessed using SDS-PAGE. The concentration of purified protein was determined using the BCA protein assay. With the determined protein concentration and Beer's law (1), extinction coefficients for each enzyme were calculated using each enzyme's absorbance at 280 nm.

$$A = \epsilon \left(\text{M}^{-1} \text{cm}^{-1} \right) \times b(\text{cm}) \times C(\text{M}) \quad (1)$$

Figure 5: Beer's law formula where A equals absorbance, ϵ the extinction coefficient, b the path length and C the concentration of the enzyme.

Enzymatic Assay

ECH activity was followed by measuring the decrease in absorbance at 280nm associated with the hydration of the C-(2) double bond in crotonyl-CoA (Figure 6). Reactions were performed at 25°C in 20 mM sodium phosphate buffer, pH 7.4, 3 mM EDTA and .0044% albumin. Initial velocities for each enzyme were measured using 20 mM of crotonyl-CoA and variable amounts of enzyme. Substrate concentration was determined by the absorbance at 280 nm using an extinction coefficient of 3600 M⁻¹cm⁻¹.^{46,47} To determine the range of detectable ECH activities, assay measurements were made in triplicate using 20 mM crotonyl-CoA and ECH concentrations from 33 pM to 20.22 nM.

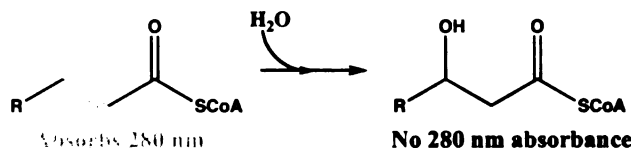


Figure 6: Enoyl CoA Hydratase activity is measured by following the decrease in absorbance at 280 nm affected by the hydration of the C-(2) double bond in crotonyl CoA. The double bond that absorbs at 280 nm is depicted in red.

Specific activity calculations

To calculate specific activity for the ECH reaction the milligrams of enzyme used in the assay was calculated from the absorbance at 280 nm using formulas 2, 3, and 4.

Additionally the rate of the enzymatic assay was converted from milli-absorbance units (mAU) minute^{-1} to $\mu\text{moles minute}^{-1}$ using formulas 5 and 6. The specific activity was calculated using formula 7.

$$\text{Concentration (M)} = \frac{280\text{nm absorbance (AU)}}{\text{Extinction coefficient (M}^{-1}\text{cm}^{-1}) \times \text{Path length (cm)}} \quad (2)$$

$$\text{Moles of protein} = \text{Concentration} \left(\frac{\text{moles}}{\text{liter}} \right) \times \text{Cuvette volume (} 100 \times 10^{-6}\text{L)} \quad (3)$$

$$\text{(a) Milligrams of protein} = \text{Moles of protein} \times \text{Formula weight} \left(\frac{\text{gram}}{\text{mole}} \right) \times \left(\frac{1000 \text{ mg}}{1 \text{ g}} \right) \quad (4)$$

$$\text{Rate of activity} \left(\frac{\text{Molar}}{\text{minute}} \right) = \text{Rate} \left(\frac{\text{mAU}}{\text{minute}} \right) \times \left(\frac{1 \text{ AU}}{1000 \text{ mAU}} \right) \times \left(\frac{1}{3600 (\text{M}^{-1}\text{cm}^{-1}) \times \text{Path length (cm)}} \right) \quad (5)$$

$$\text{(b) Rate of activity} \left(\frac{\mu\text{moles}}{\text{minute}} \right) = \text{Rate} \left(\frac{\text{Molar}}{\text{minute}} \right) \times (100 \times 10^{-6}\text{L}) \times \left(\frac{10^6 \mu\text{moles}}{1 \text{ mole}} \right) \quad (6)$$

$$\text{(c) Specific activity} = \left(\frac{\mu\text{moles}}{\text{minute}} \right) \text{ mg protein} \quad (7)$$

Figure 7: Formulas to calculate the specific activity for the enoyl CoA hydratase reaction. a) Formulas used to calculate the milligrams of protein from molar concentration. b) Formulas to convert the reaction rate from mAU minute^{-1} into $\mu\text{moles minute}^{-1}$. c) Calculation of specific activity from the reaction rate and mg protein.

Results

Enzyme	Expression condition	Purified (mg protein (g cell) ⁻¹)	Extinction Coefficient (M ⁻¹ cm ⁻¹)	Enoyl CoA Hydratase activity (μmol min ⁻¹ (mg of protein) ⁻¹)
Enoyl CoA Hydratase	+1mM IPTG	6.21	12611	864
4-Chlorobenzoyl CoA Dehalogenase	Undetected	-	NA	NA
Δ ³ ,Δ ² Dodecenoyl CoA Isomerase	+1mM IPTG	-	NA	NA
Histone Acetyltransferase	+1mM IPTG	-	NA	NA
Crotonobetainyl CoA Hydratase	+1mM IPTG	1.96	20833	0.857
3-Hydroxyisobutyryl CoA Hydrolase	+1mM IPTG	8.35	54729	Undetected
Methylmalonyl CoA Decarboxylase	+1mM IPTG	8.37	17095	Undetected
Methylglutaconyl CoA Hydratase	+1mM IPTG	1.74	21230	Undetected
Feruloyl CoA Hydratase/Lyase	+1mM IPTG	0.735	56170	Undetected
Cyclohex-1-enecarboxyl-CoA Hydratase	Undetected	-	NA	NA
2-Ketocyclohexanecarboxyl-CoA Hydrolase	Undetected	-	NA	NA
Cyclohexa-1,5-dienecarbonyl-CoA Hydratase	+1mM IPTG	0.34	14192	Undetected
1,4-dihydroxy-2-Naphthoyl-CoA Synthase	+1mM IPTG	5.03	31685	Undetected

Table 1: Overview of results

Enoyl CoA Hydratase

Amplification from plasmid DNA

For purposes of sequence identification, primers were designed to amplify ECH from plasmid DNA. To this end, the forward primer 5'- ATGGGTGCTAACTTTCAGTACATCA C -3' and reverse primer 5'- TCAGTGGTCTTTGAAGTTGGCC -3' were designed. Using PCR plasmid DNA was amplified producing a fragment of approximately 786 nucleotides. Sequencing of the resultant gene identified it as a structurally characterized variant of mitochondrial ECH from *rattus novegicus*.

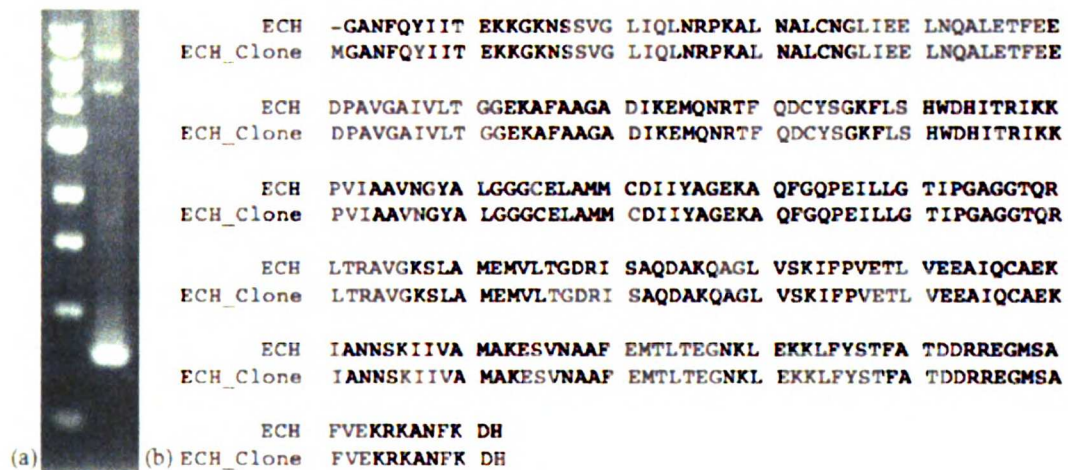


Figure 8 (a) 10% Agarose gel: DNA standard (left lane), ECH (right lane). (b) Sequenced gene is homologous to a shortened variant of mitochondrial ECH from *rattus novegicus* used for structural determination.

Over-expression of recombinant protein

ECH was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS-PAGE illustrated that IPTG increases over-

Figure 9: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. ECH circled in red.



expression of ECH relative to empty vector and uninduced bacteria.

Purification of recombinant protein

According to the absorbance at 280nm, extraneous protein was washed off the column using an imidazole gradient to a final concentration of 200 mM.

Following this wash step, imidazole concentration was raised to 500 mM and

ECH was eluted in a minimal volume of buffer.

The eluted protein was visualized with SDS-PAGE,

(a)

(b)

Figure 10: (a) Nickel-affinity purification chromatogram of ECH, blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively. (b) SDS-PAGE analysis of ECH purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

and found purified to homogeneity. Using the BCA protein assay, 43.45 mg were determined to have been purified from 7 grams of *Escherichia coli* cells.

Enoyl CoA Hydratase enzymatic activity assay

ECH was measured for its native reaction in triplicate and the specific activity was calculated using equations 1-7. To assay ECH for activity, 0.842 nM of enzyme and 20 mM substrate were used. Under these conditions ECH exhibited a specific activity of 864 $\mu\text{mol min}^{-1} (\text{mg protein}^{-1})$.

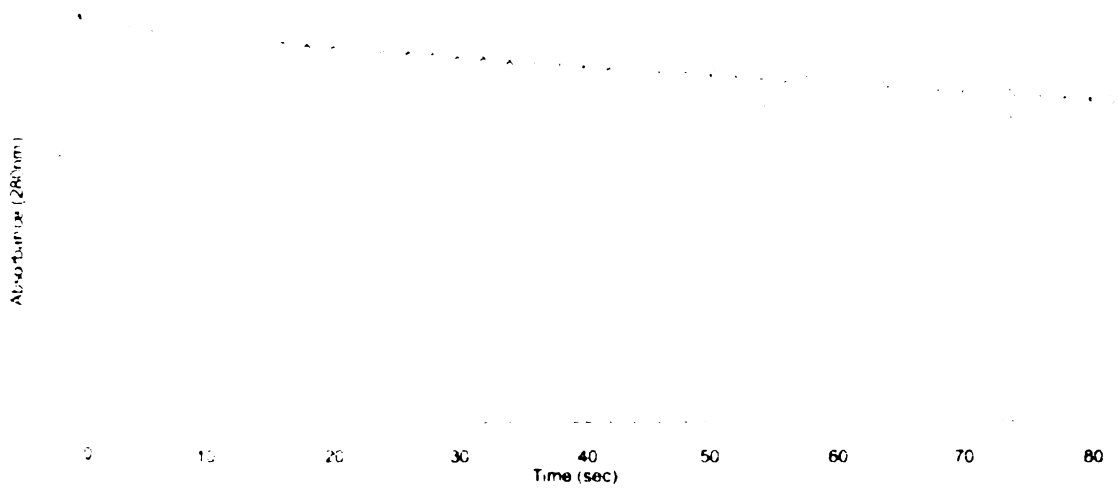


Figure 11: ECH kinetic spectra. Square and circle samples are negative controls with no enzyme. Triangle and diamond assays have 0.842 nM ECH. Using formulas 1-7 a specific activity of $864 \mu\text{mol minute}^{-1} (\text{mg protein})^{-1}$ was calculated.

4-Chlorobenzoyl-CoA Dehalogenase

Cloning from plasmid DNA

For purposes of cloning and sequence identification, primers were designed to amplify 4CBD from plasmid DNA. To this end, the forward primer 5'- GAATTAATTC GGATCCATGTATGAGGCAATTGGTCACC -3' and reverse primer 5'- GGTGGTGG TGCTCGAGCTACACGCCCGCCG -3' were designed. Using PCR plasmid DNA was amplified producing a fragment of 810 base pairs. This gene was subsequently digested and ligated into the pET20b expression vector. Sequencing of the resultant gene identified it as 4CBD from *pseudomonas strain Cbs3*.

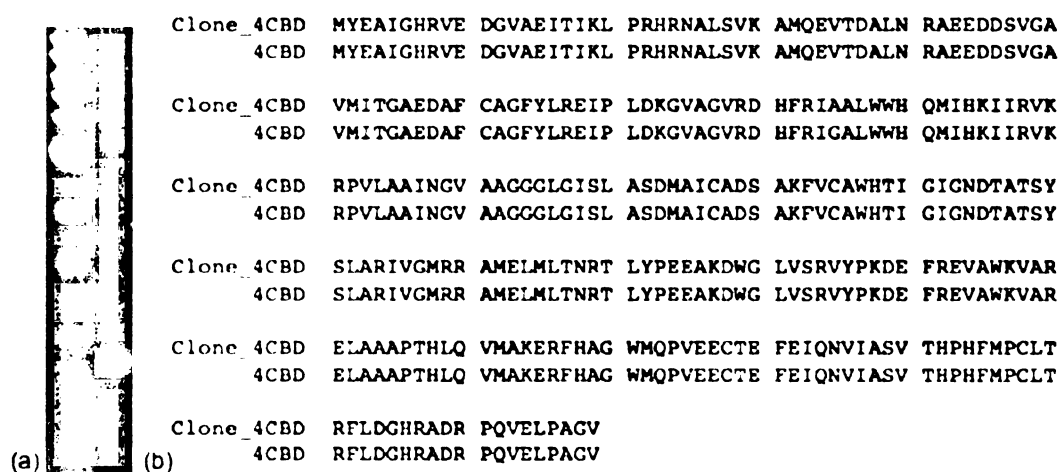


Figure 12 (a) 10% Agarose gel: DNA standard (left lane), 4CBD (right lane). (b) Sequenced gene is homologous to 4CBD from *pseudomonas strain Cbs3*.

Overexpression of recombinant protein

4CBD was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the

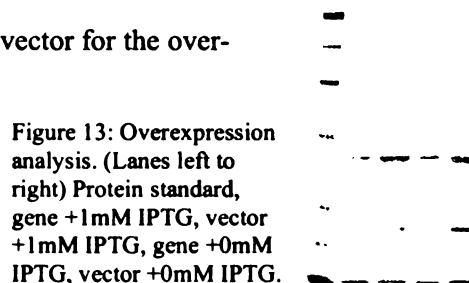


Figure 13: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG.

lactose promoter by 1mM IPTG. SDS-PAGE was unable to detect an overexpressed protein band under either sample condition.

Purification of recombinant protein

To purify 4CBD from whole cellular lysate, nickel affinity chromatography was used. After loading the protein onto to the column, an imidazole gradient from 5 mM to 500 mM was used to wash and elute the target protein. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The gradient continued up to 500 mM imidazole but no protein eluted. Further analysis is necessary to determine conditions to obtain pure, soluble protein.

Figure 14: Purification of 4-CBD by nickel affinity chromatography where the blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively.

$\Delta^{3,5}$ -Dodecenoyl-CoA Isomerase

Cloning from genomic DNA

For purposes of cloning and sequence identification, primers were designed to amplify DDI from *sacchromyces cerevisiae* genomic DNA. To this end, the forward primer 5'- GAATTAATTCGGATCCATGTCGCAAGAAATTAGGC-3' and reverse primer 5'- GACGGAGCTCGAATTCTCATAAACGATGCTTCCTTTG -3' were used. Using PCR genomic was amplified producing a fragment of 840 base pairs. This gene was subsequently digested and ligated into pET20b expression vector. Sequencing of the resultant gene identified it as DDI from *sacchromyces cerevisiae*.

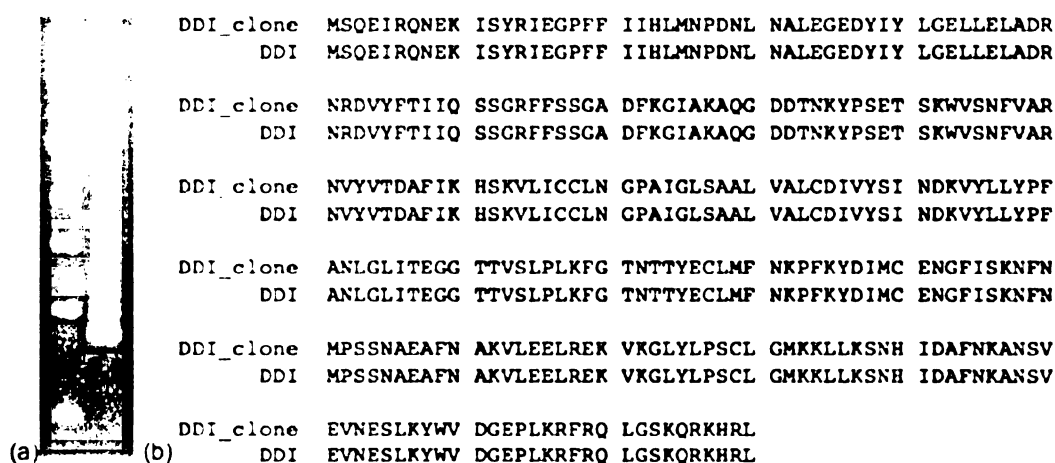
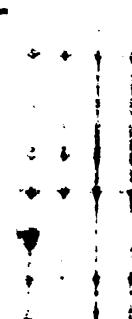


Figure 15 (a) 10% Agarose gel: DNA standard (left lane), DDI (right lane). (b) Sequenced gene is homologous to DDI from *sacchromyces cerevisiae*.

Overexpression of recombinant protein

DDI was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the

Figure 16: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. Protein band circled in red.



lactose promoter by 1mM IPTG. SDS-PAGE demonstrated that IPTG affects the over-expression of DDI relative to empty vector and uninduced bacteria.

Purification of recombinant protein

To purify DDI from whole cellular lysate, nickel affinity chromatography was used. After loading the protein onto to the column, an imidazole gradient from 5 mM to 500 mM was used to wash and elute the target protein. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The gradient continued up to 500 mM imidazole, but no protein eluted. Further analysis is necessary to determine conditions to obtain pure, soluble protein.

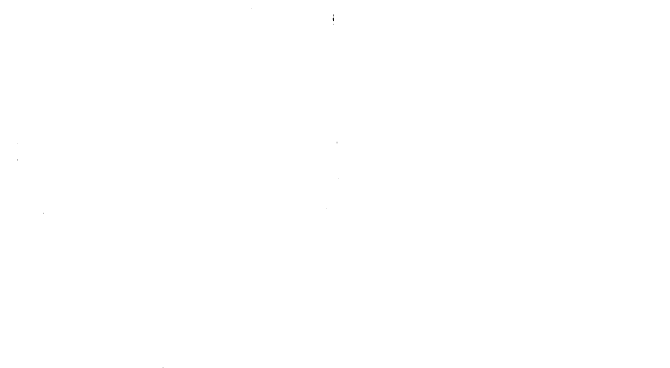


Figure 17: Purification of DDI by nickel affinity chromatography where the blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively.

Histone Acetyltransferase

Amplification and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify HisA from plasmid DNA. To this end, the forward primer 5'- ATGGCTTCCGAGGAGCTGT AC -3' and reverse primer 5'-TCAGAACTCATCGATCTTCCTCT -3' were designed. Using PCR plasmid DNA was amplified producing a fragment of approximately 1630 nucleotides.

Sequencing of the resultant gene identified it as HisA from *homosapien*.

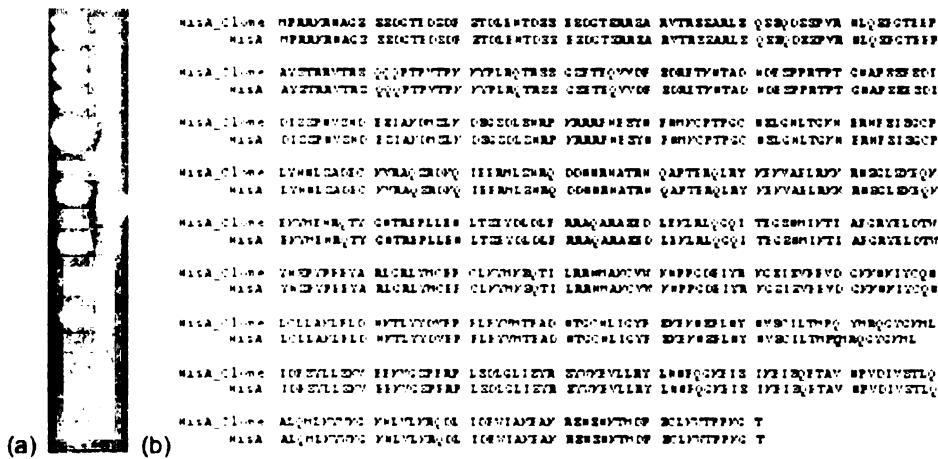
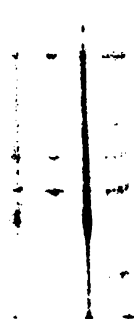


Figure 18 (a) 10% Agarose gel: DNA standard (left lane), Histone acetyltransferase (right lane). (b) Sequenced gene is homologous to histone acetyltransferase from *homosapien*.

Overexpression of recombinant protein

HisA was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS-PAGE illustrated that IPTG increases the over-expression of HisA relative to empty vector

Figure 19: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. HisA band circled in red.



and uninduced bacteria.

Purification of recombinant protein

To purify HisA from whole cellular lysate nickel affinity chromatography was used. After loading the protein onto to the column, an imidazole gradient from 5 mM to 500 mM was used to wash and elute the target protein. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The gradient continued up to 500 mM imidazole, but no protein eluted. Further analysis is necessary to determine conditions to obtain pure, soluble protein.

Figure 20: Purification of HisA by nickel affinity chromatography where the blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively.

Crotonobetainyl-CoA Hydratase

Cloning from genomic DNA

For purposes of cloning and sequence identification, the forward primer 5'-TTTATTACATATGAAACGGCAGGGAAC-3' and reverse primer 5'-ATTA ACTCGAGTTAACGTCCTTCCACA -3' primers were used to amplify CaiD from *Escherichia coli* K12 genomic DNA. Using PCR a gene of approximately 900 nucleotides was amplified, which was subsequently digested and ligated into a pET20b expression vector. Sequencing of the resultant gene identified it as CaiD from *Escherichia coli* K12.

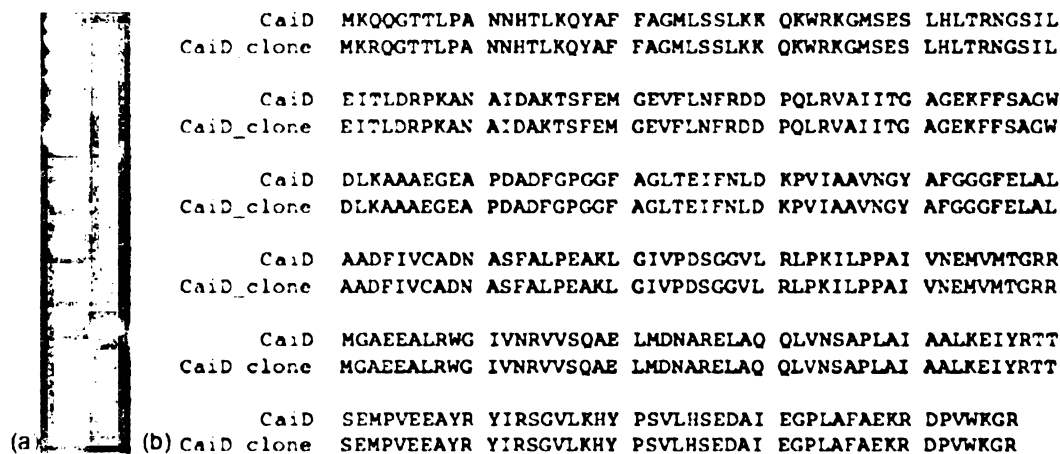


Figure 21 (a) 10% Agarose gel: DNA standard (left lane), CaiD (right lane). (b) Sequenced gene is homologous to CaiD from *Escherichia coli* K12.

Overexpression of recombinant protein

CaiD was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS-PAGE

Figure 22: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. CaiD circled in red.

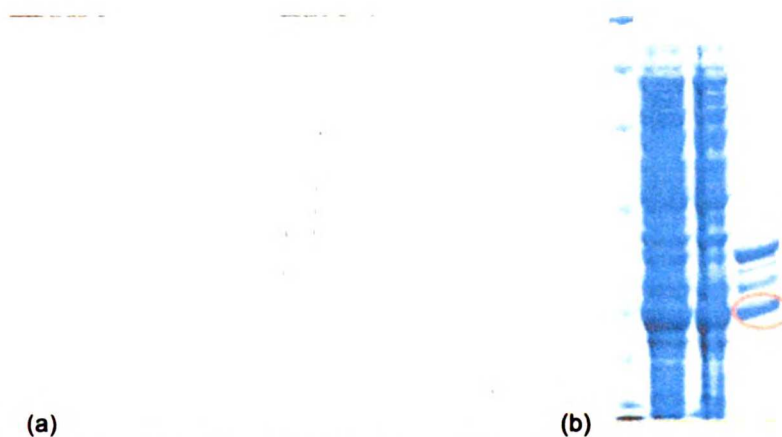
illustrated that IPTG increases the over-expression of CaiD relative to empty vector and uninduced bacteria.

Purification of recombinant protein

CaiD was purified from whole cellular lysate using nickel affinity chromatography.

To wash the column and elute the target protein, an imidazole gradient from 5 mM to 500 mM was used.

According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The



(a) Figure 23: Purification of CaiD by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm. and concentration of imidazole, respectively. (b) SDS-PAGE analysis of CaiD purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein. The soluble CaiD band is circled in red.

target protein began to elute with 200 mM imidazole. Using the BCA protein assay, 10.37 mg were determined to have been purified from 5.3 g of *Escherichia coli*. As indicated by the SDS-PAGE, CaiD was purified to homogeneity with two other bands. The additional protein bands were considered precipitated CaiD.

Enoyl CoA Hydratase activity assay

CaiD was measured for promiscuous ECH activity in triplicate. The specific activity was calculated using equations 2-7. To assay CaiD activity 20 mM substrate was used with enzyme concentrations between 0.1 to 0.5 μ M to make sure the enzyme concentration was

within the linear range of the assay. Under these conditions CaiD exhibited a specific activity of $0.857 \mu\text{mol min}^{-1} (\text{mg of protein})^{-1}$.

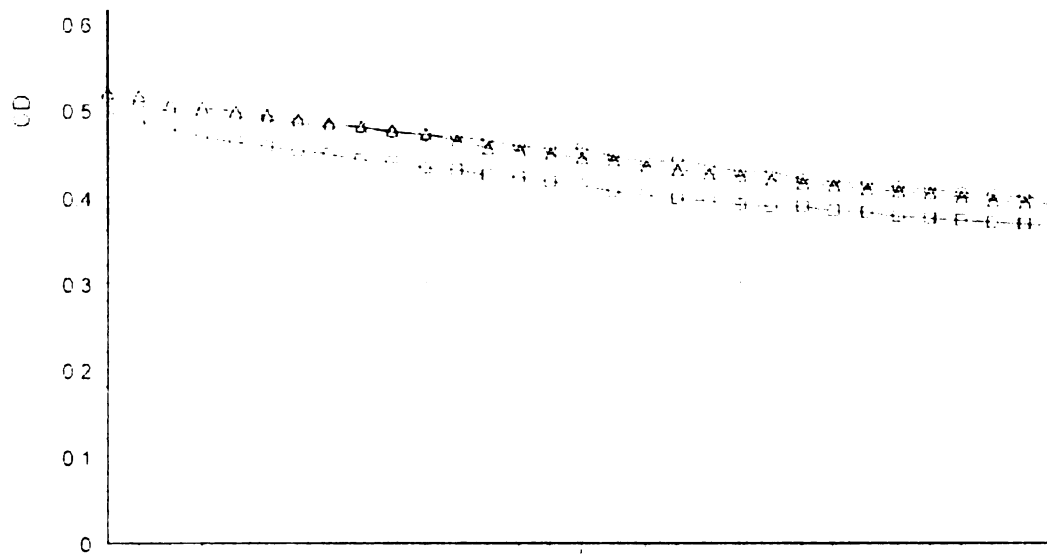


Figure 24: CaiD kinetic spectra using $0.27 \mu\text{M}$ enzyme concentration and $20 \mu\text{M}$ substrate.

3-Hydroxyisobutyryl-CoA Hydrolase

Amplification and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify 3-HICH from plasmid DNA. To this end, the forward primer 5'- TCAGAGCCCCGCCAG CG -3' and reverse primer 5'- ATGAACGTGCTTTTCGAAGAACG -3' were designed. Using PCR plasmid DNA was amplified producing a fragment of approximately 1110 base pairs. Sequencing of the resultant gene identified it as HICH from *pseudomonas aeruginosa*.

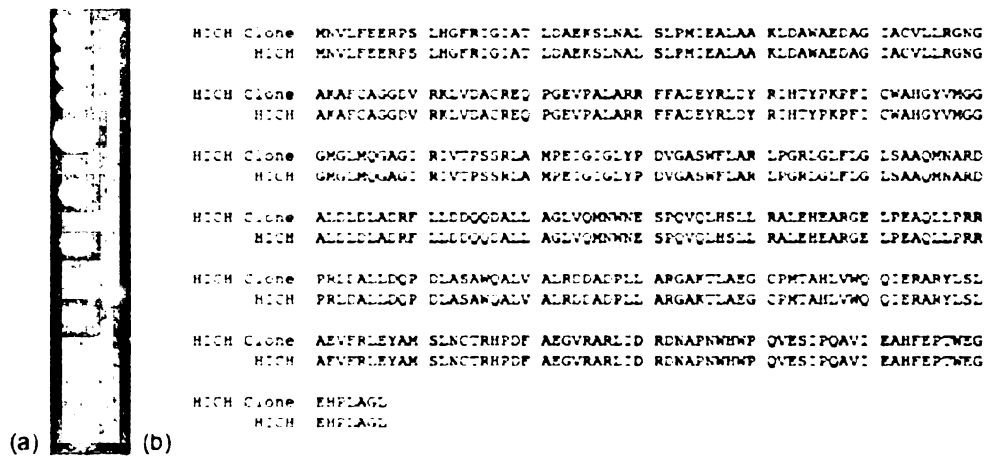


Figure 25 (a) 10% Agarose gel: DNA standard (left lane), HICH (right lane). (b) Sequenced gene is homologous to HICH from *pseudomonas aeruginosa*.

Overexpression of recombinant protein

HICH was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG.

Figure 26: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. HICH band circled in red.



SDS-PAGE demonstrated that IPTG increases the overexpression of HICH relative to empty vector and bacteria grown without IPTG.

Purification of recombinant protein

HICH was purified from whole cellular lysate using nickel affinity chromatography. To wash the column and elute the target protein an imidazole gradient from 5 to 500 mM was used. According to the absorbance at 280nm, extraneous protein was washed off the column with about 70 mM imidazole.

The target protein began to elute with 250 mM imidazole, at which point the concentration of imidazole was increased to 500 mM

to elute the protein in a minimal volume of buffer.

Using the BCA protein

assay, 37.75 mg were determined to have been purified from 3.8 g of *Escherichia coli*.

According to SDS-PAGE HICH was purified to homogeneity.

Promiscuous Enoyl CoA hydratase assay

Purified HICH was measured for promiscuous ECH activity and no detectable amount of activity was observed using enzyme concentrations up to 18.42 μ M.



(a) (b)
Figure 27: Purification of HICH by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm. and concentration of imidazole, respectively. (b) SDS-PAGE analysis of HICH purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

Methylmalonyl-CoA Decarboxylase

Amplification and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify MMCD from plasmid DNA. To this end, the forward primer 5'- GTGAATTTATCCAG ACGCAATATTTTG -3' and reverse primer 5'- TTAATGACCAACGAAATTAGGTTT ACG -3' were used. Using PCR MMCD was amplified producing a fragment of approximately 840 nucleotides. Sequencing of the resultant gene identified it as MMCD from *Escherichia. coli*.

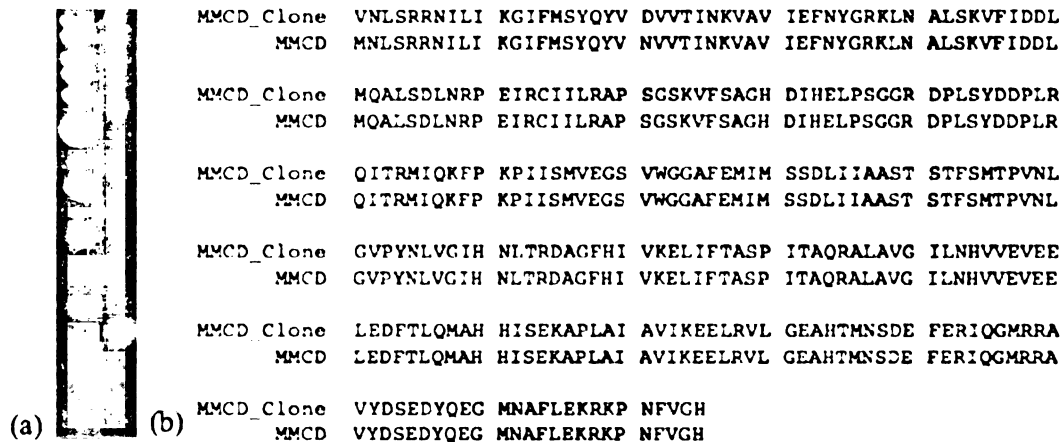
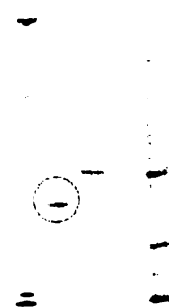


Figure 28 (a) 10% Agarose gel: DNA standard (left lane), MMCD (right lane). (b) Sequenced gene is homologous to MMCD from *escherichia coli*.

Overexpression of recombinant protein

MMCD was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS-PAGE illustrates that IPTG increases the

Figure 29: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. MMCD band circled in red.



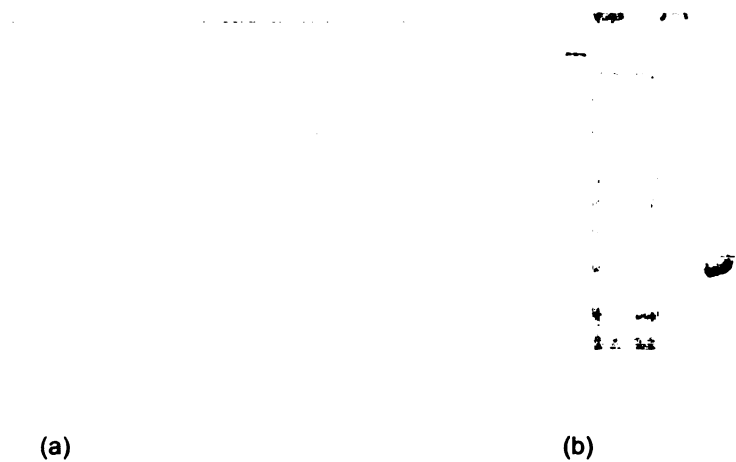
over-expression of MMCD relative to empty vector and uninduced bacteria.

Purification of recombinant protein

MMCD was purified from whole cellular lysate using nickel affinity chromatography. To wash the column and elute the target protein an imidazole gradient from 5 mM to 500 mM was used. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The target protein began to elute with 300 mM imidazole, at which point the concentration of imidazole was increased to 500 mM to elute the protein in a

minimal volume of buffer. Using the BCA protein assay, 59.46 mg were determined to have been purified from 7.1 g of *Escherichia coli*. As

indicated by SDS-PAGE, MMCD was purified to homogeneity.



(a) (b)
Figure 30: Purification of MMCD by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively. (b) SDS-PAGE analysis of MMCD purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

Promiscuous Enoyl CoA hydratase assay

Purified MMCD was measured for promiscuous ECH activity and no detectable amount of activity was observed using enzyme concentrations up to 58.5 μ M.

Methylglutaconyl-CoA Hydratase

Amplification and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify 3-MGCH from plasmid DNA. To this end, the forward primer 5'- ATGAGCGATTTTCAG CACCCTC -3' and reverse primer 5'- TTAATGACTCATGGGCGCGGCTCCTT -3' were designed. Using PCR MGCH was amplified, producing a fragment of approximately 810 nucleotides. Sequencing of the resultant gene identified it as MGCH from *pseudomonas putida*.

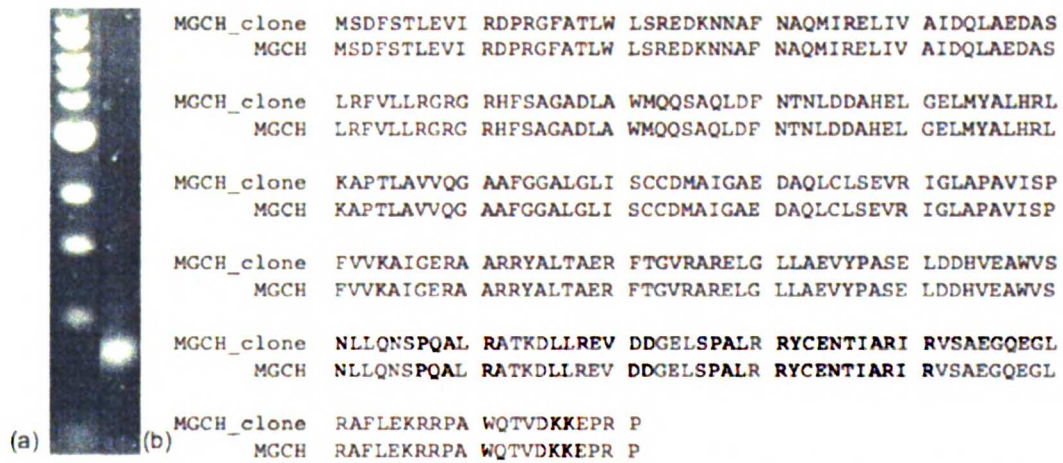


Figure 31 (a) 10% Agarose gel: DNA standard (left lane), MGCH (right lane). (b) Sequenced gene is homologous to MGCH from *pseudomonas putida*.

Overexpression of recombinant protein

MGCH was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG.

SDS-PAGE illustrates that IPTG increases the overexpression of MGCH relative to empty vector and uninduced bacteria.

Figure 32: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. MGCH band circled in red.



Purification of recombinant protein

MGCH was purified from whole cellular lysate using nickel affinity chromatography. To wash the column and elute the target protein an imidazole gradient from 5 mM to 500 mM was used. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately

70 mM imidazole. The

target protein began to

elute with 300 mM

imidazole, at which point

the concentration of

imidazole was increased to

500 mM to elute the

protein in a minimal

volume of buffer. Using

the BCA protein assay, 10.99 mg were determined to have been purified from 6.3 g of

Escherichia coli. As indicated by SDS-PAGE, MGCH was purified to homogeneity.

Promiscuous Enoyl CoA hydratase assay

Purified MGCH was measured for promiscuous ECH activity and no detectable amount of activity was observed using enzyme concentrations up to 47.1 μ M.



Figure 33: Purification of MGCH by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm. and concentration of imidazole, respectively. (b) SDS-PAGE analysis of MGCH purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

Feruloyl-CoA Hydratase/Lyase

Cloning from genomic DNA

For purposes of cloning and sequence identification, primers were designed to amplify FHL from *pseudomonas syringae* genomic DNA. To this end the forward primer 5'-AAGGAGATATACATATGAGCAAATATGAAGGTCGC -3' and reverse primer 5'-GACGGAGCTCGAATTCTAGCGTTTATAAGCTTCCAGGC -3' were designed. Using PCR genomic DNA was amplified producing a fragment of approximately 820 base pairs, which was subsequently digested and ligated into the pET20b expression vector. Sequencing of the resultant gene identified it as FHL from *pseudomonas syringae*.

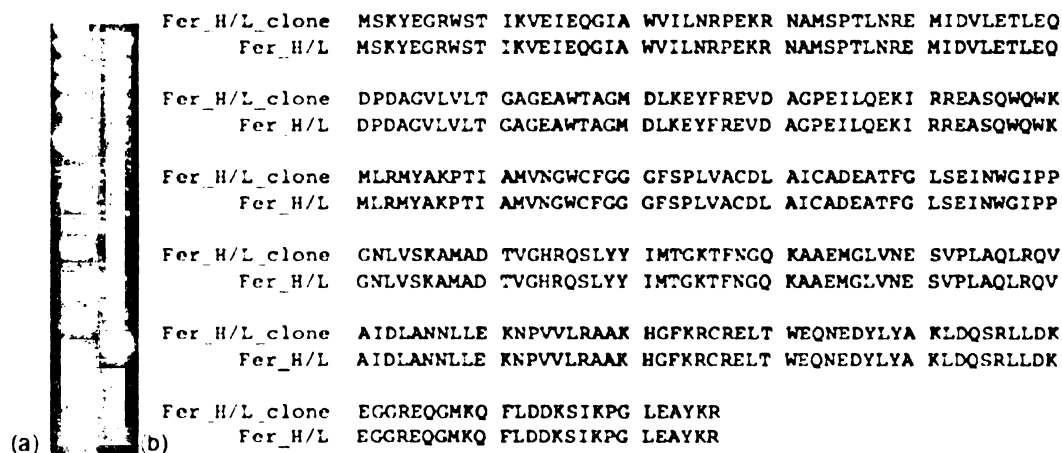


Figure 34 (a) 10% Agarose gel: DNA standard (left lane), FHL (right lane). (b) Sequenced gene is homologous to FHL from *pseudomonas syringae*.

Overexpression of recombinant protein

FHL was cloned into a lactose inducible vector for protein over-expression. To identify optimal expression conditions, protein measurements were made with and without induction of the

Figure 35: Overexpression analysis. (Lanes left to right) Protein standard, gene +0mM IPTG, vector +0mM IPTG, gene +1mM IPTG, vector +1mM IPTG. FHL band circled in red.



lactose promoter by 1mM IPTG. SDS-PAGE was used to visualize the overexpressed protein relative to empty vector. This analysis demonstrates that IPTG increases the over-expression of FHL relative to empty vector and uninduced bacteria.

Purification of recombinant protein

FHL was purified from whole cellular lysate using nickel affinity chromatography. To wash the column and elute the target protein an imidazole gradient from 5 mM to 500 mM was used. According

to the absorbance at

280nm, extraneous protein

was washed off the column

with approximately 70 mM

imidazole. The target

protein eluted earlier than

other enzymes with

approximately 100 mM

imidazole. Using the BCA

protein assay, 2.94 mg

were determined to have been purified from 4 g *Escherichia coli*. As indicated by SDS-

PAGE, FHL was purified to homogeneity.

Promiscuous enoyl CoA hydratase assay

Purified FHL was measured for promiscuous ECH activity and no detectable amount of activity was observed using enzyme concentrations up to 18 μ M.

(a)

(b)

Figure 36: Purification of FHL by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively. (b) SDS-PAGE analysis of FHL purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

Cyclohex-1-enecarboxyl-CoA Hydratase

Amplification and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify BADK from plasmid DNA. To this end, the forward primer 5'- CTAGCGGTGGGAGAAGCA C-3' and reverse primer 5'- TTGTCGTCCAATCCTATCCTCAC-3' were designed. Using PCR plasmid DNA was amplified producing a fragment of approximately 780 nucleotides.

Sequencing of the resultant gene identified it as BADK from *rhodopseudomonas palustris*.

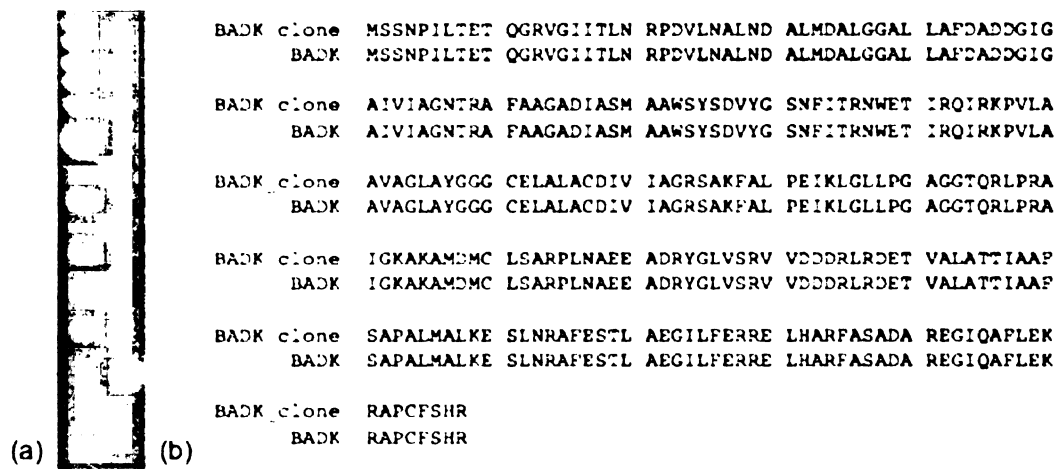


Figure 37 (a) 10% Agarose gel: DNA standard (left lane), BADK (right lane). (b) Sequenced gene is homologous to BADK from *rhodopseudomonas palustris*.

Over-expression of recombinant protein

BADK was cloned into a lactose inducible vector for the over-expression of protein.

To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS-PAGE was used to visualize the over-expressed protein relative to empty vector. This analysis was unable to detect an overexpressed protein band under either of the sample conditions.

Figure 38: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. BADK band circled in red.

Purification of recombinant protein

To purify BADK from whole cellular lysate, nickel affinity chromatography was used. After loading the protein onto to the column, an imidazole gradient from 5 mM to 500 mM was used to wash and elute the target protein. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The gradient continued up to 500 mM imidazole, but no protein eluted. Further analysis is necessary to determine conditions to obtain pure, soluble protein.

Figure 39: Purification of BADK by nickel affinity chromatography where the blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively.

2-Ketocyclohexanecarboxyl-CoA Hydrolase

Amplification and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify BADI plasmid DNA. To this end, the forward primer 5'-TTATTTGATGTATTTGCGGAATTCTGG-3' and reverse primer 5'-ATGCAGTTCGAAGACCTGATCTAT-3' were designed and used. Using PCR plasmid DNA was amplified producing a fragment of approximately 780 base pairs. Sequencing of the resultant gene identified it as BADI from *rhodopseudomonas palustris*.

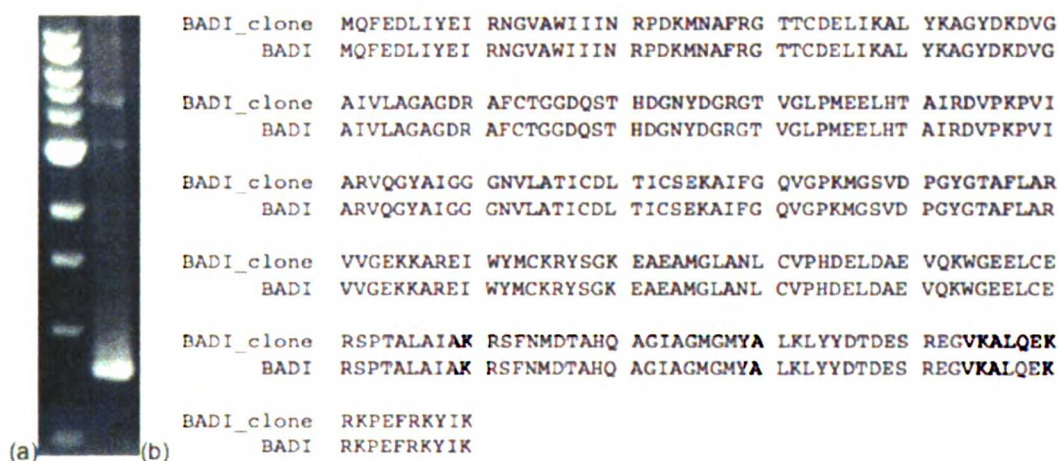


Figure 40 (a) 10% Agarose gel: DNA standard (left lane), BADI (right lane). (b) Sequenced gene is homologous to BADI from *rhodopseudomonas palustris*.

Overexpression of recombinant protein

BADI was cloned into a lactose inducible vector for the over-expression of protein.

To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS-PAGE was used to visualize the over-expressed protein relative to empty vector. This analysis was

Figure 41: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. BADI band circled in red.



unable to detect an overexpressed protein band under either of the sample conditions.

Purification of recombinant protein

To purify BADI from whole cellular lysate, nickel affinity chromatography was used. After loading the protein onto to the column, an imidazole gradient from 5 mM to 500 mM was used to wash and elute the target protein. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The gradient continued up to 500 mM imidazole. At 500 mM protein appeared to elute from the column, though nothing was identified using SDS-PAGE. The peak was determined to be precipitated protein. No soluble protein resulted from the purification. Further analysis is necessary to determine conditions to obtain pure, soluble protein.

Figure 42: Purification of BADI by nickel affinity chromatography where the blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively.

Cyclohexa-1,5-dienecarbonyl-CoA Hydratase

Cloning from genomic DNA

For purposes of cloning and sequence identification, primers were designed to amplify 1,5CH from *geobacter metallireducens* GS-15 genomic DNA. To this end the forward primer 5'- AAGGAGATATACATATGAGCGAGAGCCCTC -3' and reverse primer 5'- GACGGAGCTCGAATTCTCAGCGGTCTTGCCAG -3' were used. Using PCR genomic DNA was amplified producing a fragment of approximately 770 nucleotides, which was subsequently ligated into the pET20b expression vector. Sequencing of the resultant gene identified it as 1,5CH from *geobacter metallireducens*.

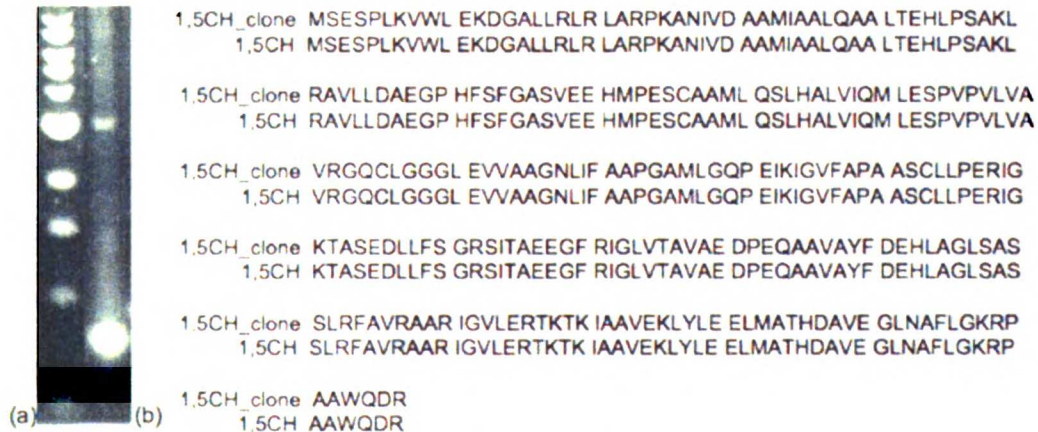


Figure 43 (a) 10% Agarose gel: DNA standard (left lane), 1,5CH (right lane). (b) Sequenced gene is homologous to 1,5CH from *geobacter metallireducens*.

Overexpression of recombinant protein

1,5CH was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG. SDS PAGE was used to

Figure 44: Overexpression analysis. (Lanes left to right) Protein standard, gene +0mM IPTG, vector +0mM IPTG, gene +1mM IPTG, vector +1mM IPTG. 1,5CH protein band circled in red.



visualize the over-expressed protein relative to empty vector. This analysis demonstrates that IPTG increases the over-expression of cyclohexa-1,5-dienecarbonyl-CoA hydratase.

Purification of recombinant protein

1,5CH was purified from whole cellular lysate using nickel affinity chromatography. To wash the column and elute the target protein an imidazole gradient from 5 mM to 500 mM was used. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The target protein eluted with approximately 100 mM imidazole. Using the BCA protein assay, 2.00 mg were determined to have been purified from 5.9 g of *Escherichia coli*. As indicated by SDS-PAGE, 1,5CH was purified to homogeneity.

Promiscuous Enoyl CoA hydratase assay

Purified 1,5CH was measured for promiscuous ECH activity and no detectable amount of activity was observed using enzyme concentrations up to 70.4 μ M.

(a)

(b)

Figure 45: Purification of 1,5CH by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively. (b) SDS-PAGE analysis of 1,5CH purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

1,4-dihydroxy-2-naphthoyl CoA Synthase

Amplification of and sequencing from plasmid DNA

For purposes of sequence identification, primers were designed to amplify MenB from plasmid DNA. To this end, the forward primer 5'- TTAGAAATAGCGCGGAAAC GGG -3' and reverse primer 5'- GTGGTGGCTCCAGCCGG-3' were designed and used. Using PCR plasmid DNA was amplified producing a fragment of approximately 950 nucleotides. Sequencing of the resultant gene identified it as MenB from *mycobacterium tuberculosis*.

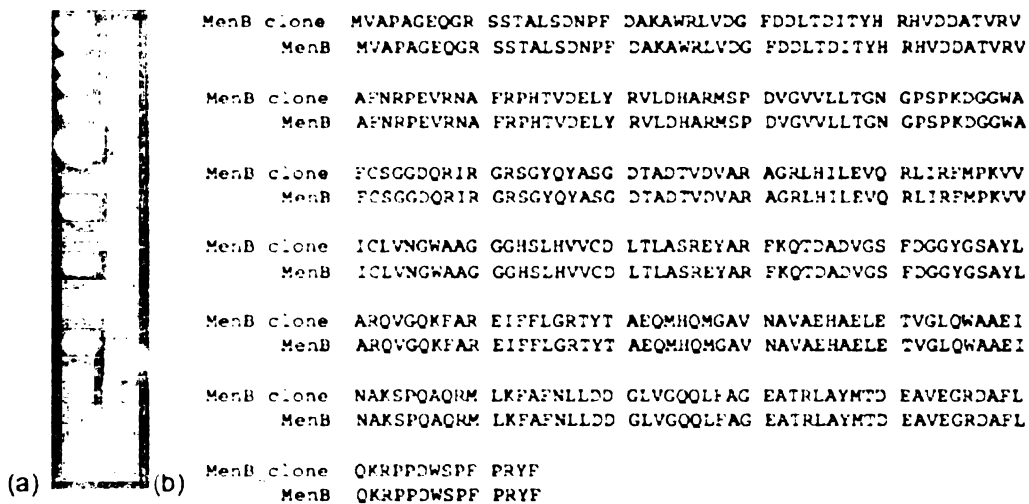


Figure 46 (a) 10% Agarose gel: DNA standard (left lane), MenB (right lane). (b) Sequenced gene is homologous to MenB from *mycobacterium tuberculosis*.

Overexpression of recombinant protein

MenB was cloned into a lactose inducible vector for the over-expression of protein. To identify optimal expression conditions, protein measurements were made with and without induction of the lactose promoter by 1mM IPTG.

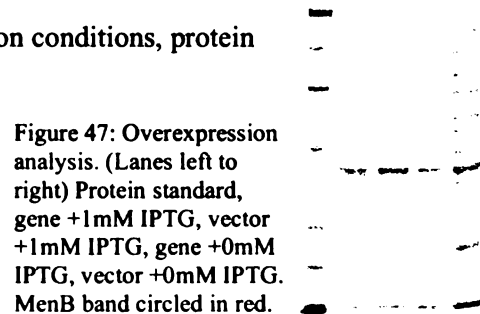


Figure 47: Overexpression analysis. (Lanes left to right) Protein standard, gene +1mM IPTG, vector +1mM IPTG, gene +0mM IPTG, vector +0mM IPTG. MenB band circled in red.

SDS-PAGE was used to visualize the overexpressed protein relative to empty vector. This analysis was unable to detect an overexpressed protein band under of the sample either conditions.

Purification of recombinant protein

MenB was purified from whole cellular lysate using nickel affinity chromatography. To wash the column and elute the target protein, an imidazole gradient from 5 mM to 500 mM was used. According to the absorbance at 280nm, extraneous protein was washed off the column with approximately 70 mM imidazole. The target protein began to elute with 200 mM imidazole. Rather than increase the elution buffer concentration to 100%, the gradient was continued to 500 mM to prevent MenB from becoming too concentrated and precipitating from solution.

Using the BCA protein assay, 25.14 mg were determined to have been purified from 5 g of *Escherichia coli*. As indicated

by SDS-PAGE, MenB was purified to homogeneity.

Promiscuous Enoyl CoA hydratase assay

Purified MenB was measured for promiscuous ECH activity and no detectable amount of activity was observed using enzyme concentrations up to 31.5 μ M.

(a)

(b)

Figure 48: Purification of MenB by nickel affinity chromatography: (a) Purification chromatogram, blue and yellow traces represent the absorbance at 280nm and concentration of imidazole, respectively. (b) SDS-PAGE analysis MenB purification, lanes left to right: protein standard, cellular lysate, column flow through, purified protein.

Technical Discussion

A couple of technical challenges were encountered in this study that should be considered when evaluating the reported results. With any spectrophotometric assay, the limit of detection is at some point above 0 $\mu\text{mol}/\text{minute}$. To address this limitation the lower bound of the reported assay was determined using decreasing concentrations ECH down to 33 pM. While we cannot say superfamily enzymes are devoid of promiscuous activity; we can say they do not exhibit activity above $1.76 \times 10^{-4} \mu\text{mol}/\text{min}$, which is quite low given the high enzyme concentrations used in this study.

The second obstacle was background interference in the assay. When working with an enzyme of low activity, one can usually increase its concentration to obtain a measurable rate. The assay used here follows the disappearance of the C-(2) double bond in crotonyl-CoA at 280 nm; in this instance, increasing the concentration of protein in the assay causes an increase in background that obscures activity measurements. To address background interference, control assays containing enzyme and buffer were performed. In every case where high enzyme concentration was used the resultant background slope in the control assay was equivalent to the slope of the assay with enzyme plus substrate. This result indicated to us that no ECH activity above $1.76 \times 10^{-4} \mu\text{mol}/\text{min}$, the lower bound of the assay, was present in these enzymes.

Discussion

The work presented here aimed to investigate enzyme evolution. Specifically, we sought to gain insight into catalytic promiscuity by asking three questions: 1) Are promiscuous capabilities abundant? 2) How are these reactions catalyzed? and 3) Could they have facilitated the evolution of new catalytic functions in nature. Previous studies have

identified promiscuous activities, though by and large these activities were limited changes in substrate specificity or minor changes in chemical reaction (e.g. hydrolysis of a phosphate bond instead of an amide bond). These results, in addition to studies that enhanced promiscuous activities, fueled hypotheses that enzymes may be capable of more than just substrate promiscuity; they may be capable of catalyzing entirely different chemical reactions. If true, then could identifying promiscuity be the key to evolving new activities and explain why *de novo* attempts in the laboratory have been so unsuccessful? If so, engineering new activity would become less an exercise in design and instead rely on identifying the correct template.

As an initial step to investigate this hypothesis, we searched for catalytic promiscuity within the enoyl CoA hydratase/isomerase superfamily. This group was chosen primarily because the characterized members use a common active site architecture to catalyze a breadth of reactions, which we felt would lend itself to promiscuously catalyzing diverse reactions. We hypothesized that promiscuity would most likely exist for enzymes directly related to one another; an enzyme may exhibit promiscuity for an activity for which it gave rise or an enzyme may retain promiscuity for a reaction it evolved from.

To predict the evolutionary relationships between superfamily members and to inform our selection of reaction and enzymes to test, we performed an all-by-all network analysis. Two thousand superfamily enzymes were clustered according to sequence homology, enabling the visualization of which enzymes cluster together as well as the linkages between clusters. The result from this analysis expectedly showed that families clustered most tightly amongst themselves. Unexpectedly, each family cluster exhibited linkages to only one other family – the monofunctional ECH's.

Could ECH be the ancestral family that gave rise to this diverse superfamily? Of all the characterized activities in the enoyl-CoA hydratase/isomerase superfamily, ECH functions in the most fundamental and ubiquitous pathway – fatty acid β -oxidation. This family of enzymes is found in nine other pathways, many of which also contain other superfamily members. Additionally, the characterized catalytic scaffolds all show similarities to the ECH active site; each enzyme has a conserved oxyanion hole and all but 4CBD have at least one catalytic glutamate homologous to one of the catalytic glutamates in ECH. Without a more thorough phylogenetic analysis, we cannot speculate further concerning the evolutionary ancestry of the superfamily, though the network results in combination with the biological and catalytic centrality present interesting support for ECH as the ancestral function of the superfamily.

With the results from the network analysis, we were able to choose the enzymes and reaction to measure for promiscuity. To this end we chose the ECH activity, hypothesizing that if ECH is the progenitor family, then several enzymes may retain catalytic capability for this reaction. Additionally, this reaction was chosen because we knew it could be catalyzed with one or both catalytic carboxylates, which we believed hinted at a plasticity that may facilitate promiscuous reactivity. We choose to assay one representative member from each family. Enzyme selection was based on the amount of available published data.

The first step in our experiment was to obtain superfamily genes either as donations from other labs or through cloning. Overexpression was attempted for all constructs, though some did not produce protein and will require further optimization. The protein that was produced was purified to homogeneity and assayed for ECH activity. Surprisingly, of the 7 enzymes measured for ECH activity, only CaiD exhibited promiscuity.

In another study, human MGCH was shown to catalyze the ECH reaction. We chose to check the bacterial variant in this analysis, which was unable to catalyze the ECH reaction at a detectable level. All of the enzymes we tested had at least one active site glutamic acid situated analogously to one of the two in ECH. The only enzymes that have been shown to catalyze the ECH reaction contain homologs to both catalytic nucleophiles and catalyze hydratase reactions on similar substrates. Are these results indicative of chemical promiscuity amongst all modern day enzymes?

The results described here illustrate the specificity necessary to catalyze a chemical reaction. Enzymes have evolved over millennia to make precise interactions with a substrate to initiate a reaction and stabilize its transition state. Having similar active sites and even partially homologous catalytic residues may not be enough for promiscuity. One needs to take into account more sophisticated features of catalysis concerning the pKa of active site residues as well as structural flexibility and dynamics.

In this study a group of highly similar enzymes catalyzing different, yet related reactions did not exhibit any promiscuity, except in the cases where the chemical reactions were identical. We believe this underscores the difficulty in predicting promiscuous reactions and illustrates why the field is currently dominated by cases of substrate promiscuity. Without catalytic promiscuity, what other mutational shortcut is available for enzymes to evolve new activities? Is chemistry-constrained divergent evolution enough to account for the scope of modern day enzymatic activities?

The next step to take along this line of research will be to engineer activity into several related enzymes that do not exhibit promiscuity in order to delineate the minimum requirement for catalysis. Several studies using superfamily enzymes have shown that new

functions can sometimes be engineered with only a couple of mutations.^{44,48} Using this approach one could ask how far off are related enzymes from promiscuity and how likely is it that these reactions would evolve without promiscuity?

Perhaps initial promiscuity is not necessary to evolve a new activity. Instead, balancing on the cusp of functional change could be sufficient. To investigate this scenario a comparative study should be undertaken to evolve a promiscuous enzyme alongside an enzyme that has had a low, promiscuous level of activity engineered into it. This line of research may provide insight into how enzymatic function has evolved in nature. As a practical application, this could inform the selection of starting points to engineer novel enzymatic activities.

Many questions remain concerning enzyme evolution. The aim of the reported work has been to provide insight into the prevalence and properties of catalytic promiscuity in nature. The results from this endeavor suggest that unmitigated promiscuity is unlikely to exist within the proteome, even amongst related enzymes and that a process by which few mutations are necessary to alter functionality may be more realistic.

Works Cited

1. Roodveldt C, Tawfik DS. (2005). Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry*. **44**, 12728-12736.
2. Loftfield RB, Vanderjagt D. (1972). The frequency of errors in protein biosynthesis. *Biochem J*. **128**, 1353-1356.
3. Colby J, S. D., Dalton H (1977). The soluble methane mono-oxygenase of *Methylococcus capuslatus* (Bath). Its ability to oxygenate n-alkanes, n-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *Biochem J*. **165**, 395-402.
4. Yoshikuni Y, Ferrin TE, Keasling JD, *Designed divergent evolution of enzyme function*. *Nature*, 2006. **440**, 1078-1082.
5. David R. J. Palmer, J. B. G., V. Sharma, R. Meganathan, Patricia C. Babbitt, John A. Gerlt (1999). Unexpected Divergence of Enzyme Function and Sequence: "N-Acylamino Acid Racemase" Is o-Succinylbenzoate Synthase. *Biochemistry* **38**, 4252-4258.
6. Torre O, Alfonso I, et al. (2004). Lipase catalysed Michael addition of secondary amines to acrylonitrile. *Chem Commun (Camb)*. **7**, 1724-1725.
7. Yasutake Y, Yao M, et al. (2004). Crystal structure of the *Pyrococcus horikoshii* isopropylmalate isomerase small subunit provides insight into the dual substrate specificity of the enzyme. *Journal of Molecular Biology* **344**. 325-333.
8. O'Brien PJ, Herschlag D. (1999). Catalytic Promiscuity and the Evolution of New Enzymatic Activities. *Chemistry and Biology*. **6**, R91-R105.
9. Zhang C, DeLisi C. (1998). Estimating the Number of Protein Folds. *Journal of Mol. Biol.* **284**, 1301-1305.
10. Brown SD, Gerlt JA, Seffernick JL, Babbitt PC. (2006). A Gold Standard Set of Mechanistically Diverse Enzyme Superfamilies. *Genome Biology*. **7**, R81-R815.
11. Hult K, Berglund P. (2007). Enzyme Promiscuity: mechanism and applications. *Trends in Biotechnology*. **25**, 231-238.
12. Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS. 2005. The 'evolvability' of promiscuous protein functions. *Nature Genetics*. **37**. 73-76.
13. Cooper AJL, Bruschi SA, Iriarte A, Martinez-Carrion M. (2002). Mitochondrial aspartate aminotransferase catalyzes cysteine S-conjugate β -lyase reactions. *Biochem. J*. **368**, pp. 253-261.
14. Holden HM, Benning MM, Haller T, Gerlt JA. (2001). The Crotonase Superfamily: Divergently Related Enzymes That Catalyze Different Reactions Involving Acyl Coenzyme A Thioesters. *Acc. Chem. Res.* **34**, 145-157.
15. Moskowitz GJ, Merrick JM. (1969). Metabolism of poly-beta-hydroxybutyrate. II. Enzymatic synthesis of D-(-)-beta hydroxybutyryl coenzyme A by an enoyl hydratase from *Rhodospirillum rubrum*. *Biochemistry*. **8**, 2748-55.
16. http://www.brenda.uni-koeln.de/php/result_flat.php4?ecno=4.2.1.17
17. Engel CK, Kiema TR, Hiltunen JK, Wierenga RK. (1998). The crystal structure of enoyl-CoA hydratase complexed with octanoyl-CoA reveals the structural adaptations required for binding of a long chain fatty acid-CoA molecule. *Journal of Molecular Biology*. **275**, 847-859.

18. Hofstein HA, Feng Y, Anderson VE, Tonge PJ. (1999). Role of glutamate 144 and glutamate 164 in the catalytic mechanism of enoyl-CoA hydratase. *Biochemistry*. **38**, 9508-9516.
19. Loffler F, Muller R. (1991). Identification of 4-chlorobenzoyl-coenzyme A as intermediate in the dehalogenation catalyzed by 4-chlorobenzoate dehalogenase from *Pseudomonas* sp. CBS3. *FEBS Letters*. **290**, 224-226.
20. Yang G, Liang PH, Dunaway-Mariano D. (1994). Evidence for nucleophilic catalysis in the aromatic substitution reaction catalyzed by (4-chlorobenzoyl)coenzyme A dehalogenase. *Biochemistry*. **33**, 8527-8531.
21. Benning MM, Taylor KL, Liu R-Q, Yang G, Xiang H, Wesenberg G, Dunaway-Mariano D, Holden HM. (1996). Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: an enzyme catalyst generated via adaptive mutation. *Biochemistry*. **35**, 8103-8109.
22. Zhang W, Wei Y, Luo L, Taylor KL, Yang G, Dunaway-Mariano D, Benning MM, Holden HM. (2001). Histidine 90 function in 4-chlorobenzoyl-coenzyme A dehalogenase catalysis. *Biochemistry*. **40**, 13474-13482.
23. Gerratana B, Arnett SO, Stapon A, Townsend CA. (2004). Carboxymethylproline synthase from *Pectobacterium carotorova*: a multifaceted member of the crotonase superfamily. *Biochemistry*. **43**, 15936-45.
24. Sleeman MC, Sorensen JL, Batchelar ET, McDonough MA, Schofield CJ. (2005). Structural and mechanistic studies on carboxymethylproline synthase (CarB), a unique member of the crotonase superfamily catalyzing the first step in carbapenem biosynthesis. *Journal of Biological Chemistry*. **280**, 34956-34965.
25. Luo MJ, Smeland TE, Shoukry K, Schulz H. (1994). Delta 3,5, delta 2,4-dienoyl-CoA isomerase from rat liver mitochondria. Purification and characterization of a new enzyme involved in the beta-oxidation of unsaturated fatty acids. *Journal of Biological Chemistry*. **269**, 2384-2388.
26. He XY, Shoukry K, Chu C, Yang J, Sprecher H, Schulz H. (1995). Peroxisomes contain delta 3,5,delta 2,4-dienoyl-CoA isomerase and thus possess all enzymes required for the beta-oxidation of unsaturated fatty acids by a novel reductase-dependent pathway. *Biochemical and Biophysical Research and Communications*. **215**, 15-22.
27. Zhang D, Liang X, He XY, Alipui OD, Yang SY, Schulz H. (2001). Delta 3,5,delta 2,4-dienoyl-CoA isomerase is a multifunctional isomerase. A structural and mechanistic study. *Journal of Biological Chemistry*. **276**, 13622-13627.
28. Mursula AM, van Aalten DM, Hiltunen JK, Wierenga RK. (2001). The crystal structure of delta(3)-delta(2)-enoyl-CoA isomerase. *Journal of Molecular Biology*. **309**, 845-853.
29. Lahn BT, Tang ZL, Zhou J, Barndt RJ, Parvinen M, Allis CD, Page DC. (2002). Previously uncharacterized histone acetyltransferases implicated in mammalian spermatogenesis. *Proc Natl Acad Sci U S A*. **99**, 8707-8712.
30. Elssner T, Engemann C, Baumgart K, Kleber HP. (2001). Involvement of coenzyme A esters and two new enzymes, an enoyl-CoA hydratase and a CoA-transferase, in the hydration of crotonobetaine to L-carnitine by *Escherichia coli*. *Biochemistry*. **40**, 11140-11148.
31. Zolman BK, Monroe-Augustus M, Thompson B, Hawes JW, Krukenberg KA, Matsuda SP, Bartel B. (2001). chyl1, an *Arabidopsis* mutant with impaired beta-oxidation, is

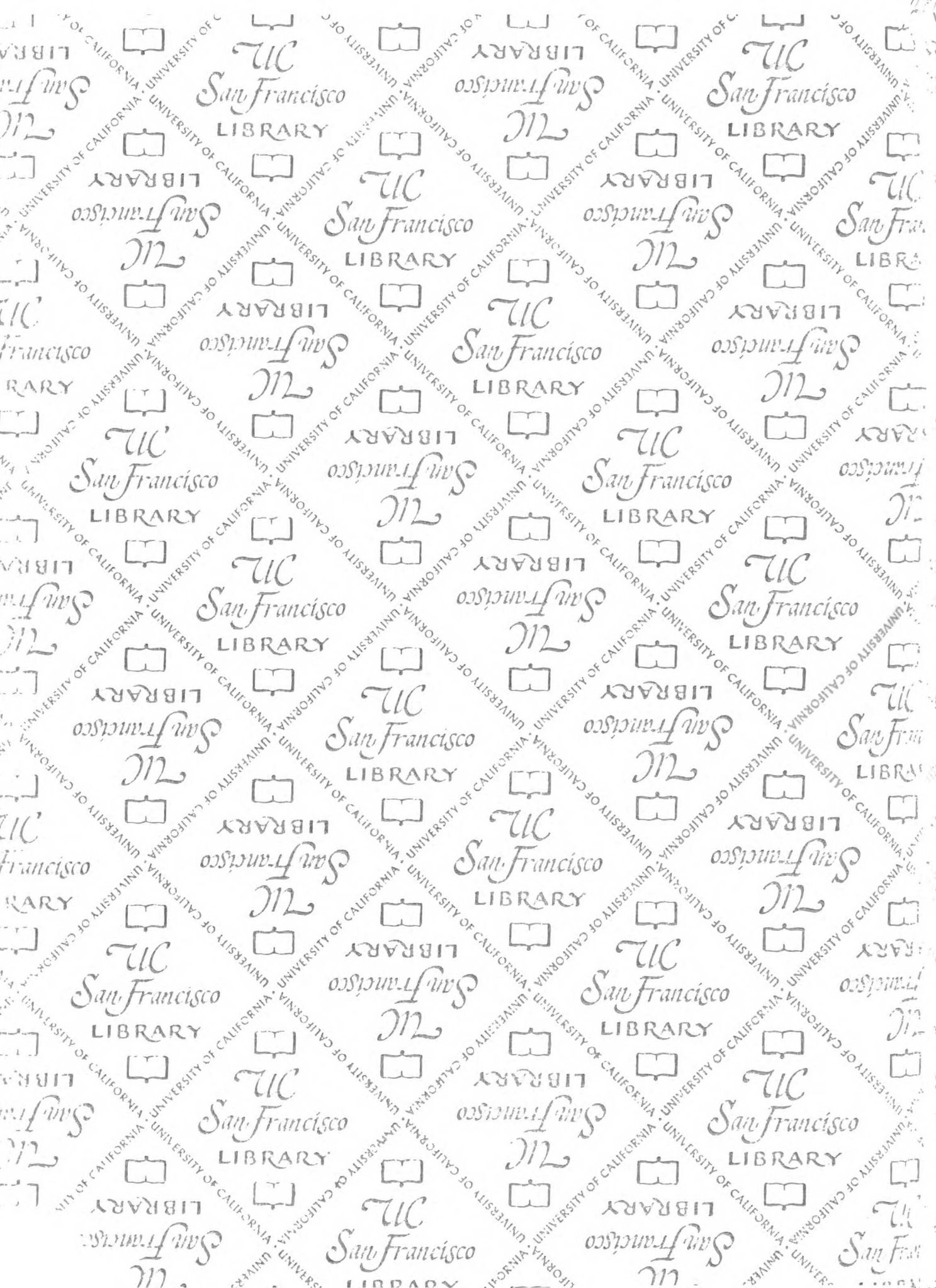
- defective in a peroxisomal beta-hydroxyisobutyryl-CoA hydrolase. *Journal of Biological Chemistry*. **276**, 31037-31046.
32. Wong BJ, Gerlt JA. (2003). Divergent function in the crotonase superfamily: an anhydride intermediate in the reaction catalyzed by 3-hydroxyisobutyryl-CoA hydrolase. *J Am Chem Soc*. **125**, 12076-12077.
 33. Haller T, Buckel T, Retey J, Gerlt JA. (2000). Discovering new enzymes and metabolic pathways: conversion of succinate to propionate by *Escherichia coli*. *Biochemistry*. **39**, 4622-4629.
 34. Benning MM, Haller T, Gerlt JA, Holden HM. (2000). New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from *Escherichia coli*. *Biochemistry*. **39**, 4630-4639.
 35. Kurimoto K, Fukai S, Nureki O, Muto Y, Yokoyama S. (2001). Crystal structure of human AUH protein, a single-stranded RNA binding homolog of enoyl-CoA hydratase. *Structure*. **12**, 1253-1263.
 36. Wong BJ, Gerlt JA. (2004). Evolution of function in the crotonase superfamily: (3S)-methylglutaconyl-CoA hydratase from *Pseudomonas putida*. *Biochemistry*. **43**, 4646-4654.
 37. Gasson MJ, Kitamura Y, McLauchlan WR, Narbad A, Parr AJ, Parsons EL, Payne J, Rhodes MJ, Walton NJ. (1998). Metabolism of ferulic acid to vanillin. A bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. *J Biol Chem*. **273**, 4163-4170.
 38. Leonard PM, Brzozowski AM, Lebedev A, Marshall CM, Smith DJ, Verma CS, Walton NJ, Grogan G. (2006). The 1.8 Å resolution structure of hydroxycinnamoyl-coenzyme A hydratase-lyase (HCHL) from *Pseudomonas fluorescens*, an enzyme that catalyses the transformation of feruloyl-coenzyme A to vanillin. *Acta Crystallogr D Biol Crystallogr*. **62**, 1494-1501.
 39. Laempe D, Eisenreich W, Bacher A, Fuchs G. (1998). Cyclohexa-1,5-diene-1-carbonyl-CoA hydratase [corrected], an enzyme involved in anaerobic metabolism of benzoyl-CoA in the denitrifying bacterium *Thauera aromatica*. *Eur J Biochem*. **255**, 618-627.
 40. Eglund PG, Pelletier DA, Dispensa M, Gibson J, Harwood CS. (1997). A cluster of bacterial genes for anaerobic benzene ring biodegradation. *Proc Natl Acad Sci U S A*. **94**, 6484-9.
 41. Eberhard ED, Gerlt JA. (2004). Evolution of function in the crotonase superfamily: the stereochemical course of the reaction catalyzed by 2-ketocyclohexanecarboxyl-CoA hydrolase. *J Am Chem Soc*. **126**, 7188-7189.
 42. Meganathan R. (2001). Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms. *Vitam Horm*. **61**, 173-218.
 43. Truglio JJ, Theis K, Feng Y, Gajda R, Machutta C, Tonge PJ, Kisker C. (2003). Crystal structure of *Mycobacterium tuberculosis* MenB, a key enzyme in vitamin K2 biosynthesis. *J Biol Chem*. **278**, 42352-42360.
 44. Xiang H, Luo L, Taylor KL, Dunaway-Mariano D. (1999). Interchange of catalytic activity within the 2-enoyl-coenzyme A hydratase/isomerase superfamily based on a common active site template. *Biochemistry*. **38**, 7638-7652.
 45. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. **13**, 2498-2504.

46. Stern JR. (1955). Crystalline crotonase from ox liver. *Methods in Enzymol.* **1**, 559-566.
47. Steinman HM, Hill RL. (1975). Bovine liver crotonase (enoyl coenzyme A hydratase). EC 4.2.1.17 L-3-hydroxyacyl-CoA hydrolyase. *Methods Enzymol.* **35**, 136-151.
48. Schmidt DM, Mundorff EC, Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA. (2003). Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry.* **42**, 8387-8393.

Appendix

Enzyme	Construct Name	Vector	Primer Name	Primer Storage
Enoyl CoA Hydratase	ECH	pET15b	ECH For ECH Rev	100mM in 10 mM TE pH 8.0
4-Chlorobenzoyl CoA Dehalogenase	4CBD	pET20b	4CBD For 4CBD Rev	100mM in 10 mM TE pH 8.0
Δ^3, Δ^2 Dodecenoyl CoA Isomerase	DDI	pET20b	DCI For DCI Rev	100mM in 10 mM TE pH 8.0
Histone Acetyltransferase	HisA	pRSET(B)	HisA For HisA Rev	100mM in 10 mM TE pH 8.0
Crotonobetainyl CoA Hydratase	CaiD	pET20b	CaiD For CaiD Rev	100mM in 10 mM TE pH 8.0
3-Hydroxyisobutyryl CoA Hydrolase	HICH	Tom15b	HICH For HICH Rev	100mM in 10 mM TE pH 8.0
Methylmalonyl CoA Decarboxylase	MMCD	Tom15b	MMCD For MMCD Rev	100mM in 10 mM TE pH 8.0
Methylglutaconyl CoA Hydratase	MGCH	pET16b	MGCH For MGCH Rev	100mM in 10 mM TE pH 8.0
Feruloyl CoA Hydratase/Lyase	FHL	pET20b	FerHL For FerHL Rev	100mM in 10 mM TE pH 8.0
Cyclohex-1-enecarboxyl-CoA Hydratase	BadK	pDPHisK1	BadK For BadK Rev	100mM in 10 mM TE pH 8.0
2-Ketocyclohexane-carboxyl-CoA Hydrolase	BadI	pET15b	BadI For BadI Rev	100mM in 10 mM TE pH 8.0
Cyclohexa-1,5-dienecarbonyl-CoA Hydratase	1,5CH	pET20b	Cyclo For Cyclo Rev	100mM in 10 mM TE pH 8.0
1,4-dihydroxy-2-Naphthoyl-CoA Synthase	MenB	pET15b	MenB For MenB Rev	100mM in 10 mM TE pH 8.0
$\Delta^{3,5}, \Delta^{2,4}$ dienoyl CoA isomerase	DCI	pNDdi	DCI For DCI Rev	100mM in 10 mM TE pH 8.0

Table 2: Superfamily clones: Plasmids containing superfamily genes and their primers are kept at -20°C in boxes labeled “Crotonase Superfamily Plasmids” and “Crotonase Superfamily Primers”, respectively. Plasmids were purified from bacterial cells using minipreps and are kept in 10 mM TE buffer, pH 8.0. Genomic DNA is kept with the primers.



UNIVERSITY OF CALIFORNIA LIBRARY
7733275



3 1378 00773 3275

San Francisco

For Not to be taken
from the room.
reference,

