

Mechanistic and Bioinformatic Investigation of a Conserved Active Site Helix in α -Isopropylmalate Synthase from *Mycobacterium tuberculosis*, a Member of the DRE-TIM Metallolyase Superfamily

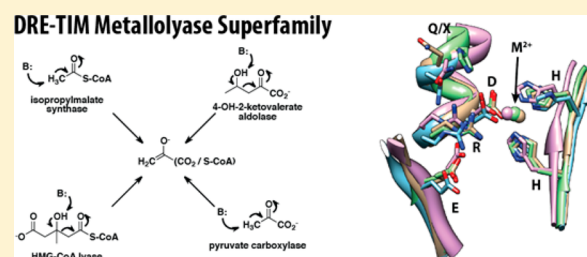
Ashley K. Casey,^{†,⊥} Michael A. Hicks,^{‡,⊥,¶} Jordyn L. Johnson,[†] Patricia C. Babbitt,^{‡,§,||} and Patrick A. Frantom^{*,†}

[†]Department of Chemistry, The University of Alabama, 250 Hackberry Lane, Tuscaloosa, Alabama 35406, United States

[‡]Departments of Bioengineering and Therapeutic Sciences and [§]Pharmaceutical Chemistry, and the ^{||}California Institute for Quantitative Biosciences, University of California, San Francisco, California 94158, United States

Supporting Information

ABSTRACT: The characterization of functionally diverse enzyme superfamilies provides the opportunity to identify evolutionarily conserved catalytic strategies, as well as amino acid substitutions responsible for the evolution of new functions or specificities. Isopropylmalate synthase (IPMS) belongs to the DRE-TIM metallolyase superfamily. Members of this superfamily share common active site elements, including a conserved active site helix and an HXH divalent metal binding motif, associated with stabilization of a common enolate anion intermediate. These common elements are overlaid by variations in active site architecture resulting in the evolution of a diverse set of reactions that include condensation, lyase/aldolase, and carboxyl transfer activities. Here, using IPMS, an integrated biochemical and bioinformatics approach has been utilized to investigate the catalytic role of residues on an active site helix that is conserved across the superfamily. The construction of a sequence similarity network for the DRE-TIM metallolyase superfamily allows for the biochemical results obtained with IPMS variants to be compared across superfamily members and within other condensation-catalyzing enzymes related to IPMS. A comparison of our results with previous biochemical data indicates an active site arginine residue (R80 in IPMS) is strictly required for activity across the superfamily, suggesting that it plays a key role in catalysis, most likely through enolate stabilization. In contrast, differential results obtained from substitution of the C-terminal residue of the helix (Q84 in IPMS) suggest that this residue plays a role in reaction specificity within the superfamily.



The enzyme isopropylmalate synthase (IPMS) catalyzes the first step in the biosynthesis of L-leucine in bacteria, archaea, and some eukaryotes. This pathway is absent in mammals, making IPMS a possible target for the development of novel antibiotic and antifungal therapeutics.¹ IPMS also serves as a model system for the study of allosteric mechanisms, as it is subject to allosteric feedback inhibition by L-leucine.² The enzyme catalyzes a Claisen-like condensation between α -ketoisovalerate (KIV) and acetyl-CoA (AcCoA) forming α -isopropylmalate and CoA (Scheme 1). Structural studies show that IPMS utilizes a distinct active site architecture to accomplish this type of chemistry when compared with malate synthase³ and citrate synthase,⁴ which catalyze similar reactions. In fact, the active site architecture exhibited by IPMS is more similar to a collection of enzymes catalyzing a diverse set of reactions including 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) lyase, 2-hydroxy-4-ketovalerate aldolase, and pyruvate carboxylase (Scheme 1). It has been proposed that these enzymes belong to a mechanistically diverse group known as the DRE-TIM metallolyase superfamily, a group of evolutionarily

related enzymes that catalyze different reactions using distinct mechanisms.⁵

Despite this diversity in function, enzymes in a superfamily share a common mechanistic aspect in the stabilization of an intermediate mediated through a set of conserved residues.⁶ Members of the DRE-TIM metallolyase superfamily share a TIM-barrel fold (Figure 1A), a D-R-E active site motif, and rely on a divalent cation for activity. Catalytically, members of this superfamily are hypothesized to stabilize an enolate intermediate in their respective reactions using the conserved arginine in the DRE motif (Scheme 2). Additionally, several characterized members of the superfamily respond to allosteric regulation.^{7,8} The active site architecture for the DRE-TIM metallolyase superfamily is highlighted by a D-R-E motif composed of a well-conserved active site α -helix containing the R and D residues adjacent to one another (Figure 1B). The conserved glutamate residue is found on an adjacent β sheet

Received: February 25, 2014

Revised: April 10, 2014

Published: April 10, 2014

Scheme 1

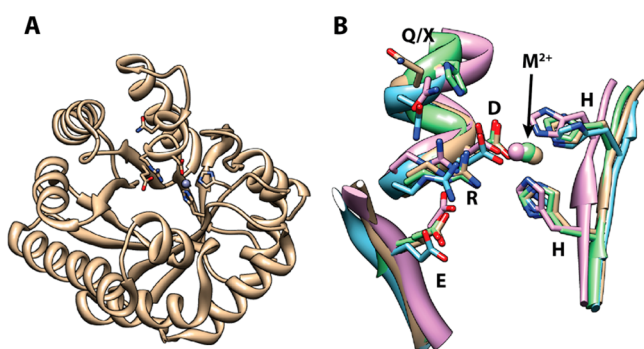
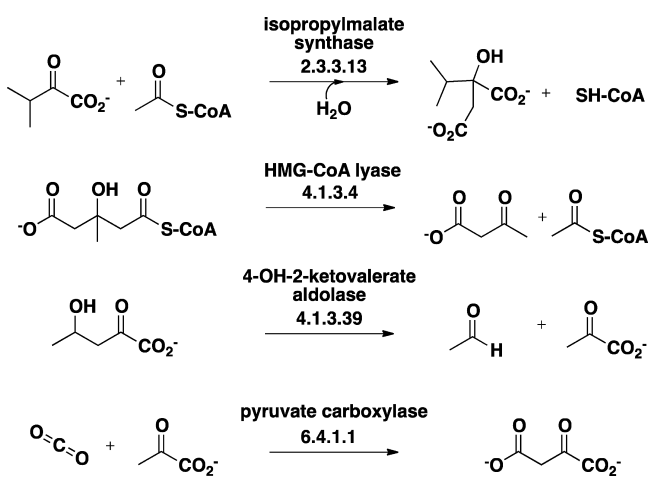
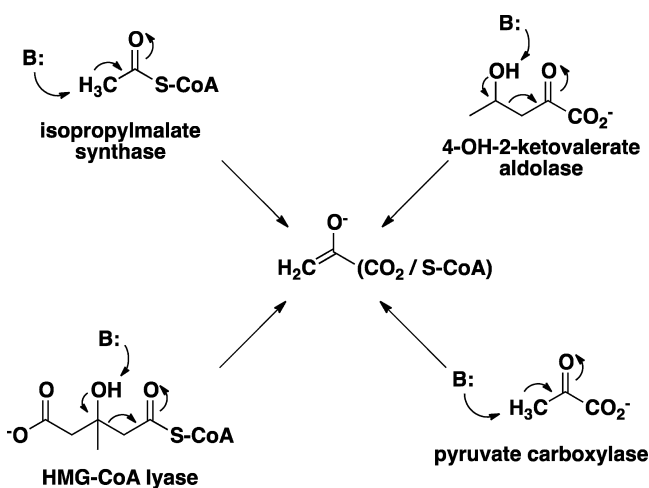


Figure 1. (A) Ribbon diagram of the DRE-TIM metallolyase catalytic domain from MtIPMS (1pdb id: 1sr9⁶³). Conserved active site residues are shown as sticks. (B) Conserved active site architecture of the DRE-TIM metallolyase superfamily. The superposition was created using the Matchmaker algorithm in Chimera with the following pdb files: 1sr9,⁶³ IPMS (tan); 1ydo,⁵ HMG-CoA lyase (blue); 2qf7,²³ pyruvate carboxylase (pink); 1nmv,⁷ 4-hydroxy-2-ketovalerate aldolase (green). Labels designate the residues of the DRE-motif, the differentially conserved Q/X residue, and the HXH motif. The divalent metal is also indicated.

Scheme 2



and is proposed to orient the arginine residue. The aspartate acts as a ligand to a required divalent cation along with two well-conserved histidine residues in a HXH motif. At least one

additional metal ligand in each enzyme is provided by a substrate; however, the nature of the substrate–metal interaction is not conserved.

While the aspartate and arginine residues have been subjected to site-directed mutagenesis in several DRE-TIM metallolyase superfamily members,^{9–13} they have not been investigated in IPMS. Recently, the active site helix containing the conserved aspartate and arginine was implicated as the target of an inhibitory allosteric signal in IPMS from *Mycobacterium tuberculosis* (MtIPMS), raising additional questions about the role of the helix in catalysis and regulation in this enzyme.¹⁴ To address these questions, site-directed mutagenesis has been carried out on MtIPMS, and the effects of substitutions on catalysis and regulation have been determined.

Analysis of the effects of residue substitution with respect to other superfamily members provides a mechanism for the identification of conserved catalytic strategies and characterization of structure/function relationships responsible for differences in reactivity, substrate selectivity, and regulation. Thus, parallel to the biochemistry studies, a bioinformatics investigation of the DRE-TIM metallolyase superfamily has been initiated and the results illustrated using sequence similarity networks for the DRE-TIM metallolyase superfamily. Sequence similarity networks have been successfully used to organize functionally diverse enzyme superfamilies into subgroups and families of sequences representing discrete reaction specificities.¹⁵ The language of superfamily hierarchies used here is as follows: superfamily, a set of evolutionary related enzymes that share a common mechanistic step, such as stabilization of the same type of intermediate, but whose overall reactions may be different; subgroup, a subset of a superfamily whose members share more similarity in sequence with one another than they do with proteins in other subgroups; family, a subset of a subgroup whose members catalyze the same reaction in essentially the same way. This organization allows for the rapid detection of conserved residues at differing hierarchies within the superfamily. For instance, more recently evolved residues (such as those conserved at the subgroup or family level) may be critical specificity determinants or provide information for unique regulatory mechanisms.¹⁶ Applying this methodology to the DRE-TIM metallolyase superfamily provides insight into the conservation and diversity of residues in the DRE active site helix and aids in teasing out differentially conserved interactions in each reaction class.

MATERIALS AND METHODS

Materials. Oligonucleotides for the mutagenesis of MtIPMS were obtained from Eurofins MWG Operon (Huntsville, AL). Acetyl CoA (AcCoA) and ketoisovalerate (KIV) were purchased from Sigma-Aldrich. 4,4'-Dithiodipyridine (DTP) was purchased from Acros Organics. All other buffers and reagents were obtained from VWR or were of the highest quality available. The HisTrap HP column was purchased from GE Healthcare. Competent cells (BL21(DE3)pLysS and Top 10) were obtained from Invitrogen.

MtIPMS Variant Construction and Purification. Wild type MtIPMS and all variants reported here were constructed and isolated as previously described.¹⁷ Briefly, QuikChange Lightning site-directed mutagenesis (Stratagene) was used to create point mutations in the pET28a(+):*leuA* plasmid. Primers used for mutagenesis are shown in Table S1 (Supporting Information). BL21(DE3)pLysS cells containing

Table 1. Kinetic Parameters Determined for MtIPMS Variants^a

enzyme	k_{cat} (s ⁻¹)	K_{AcCoA} (μM)	$k_{\text{cat}}/K_{\text{AcCoA}}$ (μM ⁻¹ s ⁻¹)	K_{KIV} (μM)	$k_{\text{cat}}/K_{\text{KIV}}$ (μM ⁻¹ s ⁻¹)
WT	3.4 ± 0.3	42 ± 12	0.08 ± 0.02	6.4 ± 1.4	0.50 ± 0.13
L79A	1.0 ± 0.1	30 ± 5	0.03 ± 0.01	120 ± 30	0.008 ± 0.002
R80A/K	nd ^b	nd	nd	nd	nd
D81A/H	nd	nd	nd	nd	nd
N83A	0.31 ± 0.04	990 ± 290	0.0003 ± 0.0001	51 ± 14 ^c	0.006 ± 0.002
N83E	nd	nd	nd	nd	nd
Q84A	0.10 ± 0.02	170 ± 60	0.0006 ± 0.0002	19 ± 6	0.005 ± 0.002

^aDetermined using 4,4-dithiopyridine (DTP) to detect the formation of CoA at 324 nm ($\epsilon = 19.8 \text{ mM}^{-1} \text{ cm}^{-1}$) at 25 °C. Standard reaction conditions consisted of 50 mM potassium phosphate buffer (pH 7.5), 12 mM MgCl₂, 50 μM DTP, and at least 5–10 times the K_m value for the nonvaried substrate (AcCoA or KIV). ^bActivity could not be determined above the detection limit of the assay. ^cDetermined using 1 mM AcCoA.

the plasmids were grown in autoinduction media. Overexpressed proteins were purified via metal affinity chromatography using a HisTrap HP column (5 mL). Protein expression and purity were checked by SDS–PAGE analysis. The peak fractions were pooled and dialyzed against 1 L of 20 mM triethanolamine (TEA) (pH 7.8) and repeated three times. The protein was stored in 10% glycerol at –20 °C.

Enzymatic Assays. A typical reaction mixture contained 50 mM potassium phosphate buffer (pH 7.5), 12 mM MgCl₂, 50 μM DTP, and 5–10 times the K_m value for the nonvaried substrate (AcCoA or KIV). Initial velocities were determined using DTP to detect the formation of CoA at 324 nm ($\epsilon = 19.8 \text{ mM}^{-1} \text{ cm}^{-1}$). Kinetic constants were determined by fitting the data to the Michaelis–Menten equation (eq 1) using Kaleidagraph (Synergy Software), where v is the velocity, $[E]_t$ is the total enzyme concentration, $[S]$ is the concentration of the substrate being varied, K_m is the Michaelis–Menten constant, and k_{cat} is the maximal velocity.

$$\frac{v}{[E]_t} = \frac{k_{\text{cat}} \times [S]}{K_m + [S]} \quad (1)$$

Inhibition Assays. Inhibition assays were performed using a standard reaction mixture of 50 mM potassium phosphate buffer (pH 7.5), 12 mM MgCl₂, 50 μM DTP, 5–10 times the K_m value of the nonvaried substrate (AcCoA or KIV), and varied concentrations of leucine (0–120 μM). Reactions were initiated by the addition of MtIPMS. For assays that displayed nonlinear (biphasic) kinetics, the steady state and final velocities were determined using eq 2, where $[P]_t$ is the total product formed (CoA) at time, t , $[E]_t$ is the total enzyme concentration, v_i is the initial velocity, v_f is the final velocity, t is time, k_b is the exponential rate constant, and C is a constant.¹⁸ The inhibition parameters were then determined by replotting the velocities versus leucine concentration and fit to eq 3 (for K_i values) or eq 4 (for K_i^* values) where K_i is the dissociation constant for the initial enzyme–inhibitor complex, K_i^* is the overall dissociation constant for the two-step slow-onset mechanism, v_i and v_f are the initial and final velocity, respectively, and β is the fractional velocity at saturating concentrations of the inhibitor, I .¹⁹

$$\frac{[P]_t}{[E]_t} = v_f t + \frac{(v_i - v_f)(1 - \exp^{-k_b t})}{k_b} + C \quad (2)$$

$$\frac{v_i}{[E]_t} = \frac{k_{\text{cat}} K_i \times \beta k_{\text{cat}} [I]}{[I] + K_i} \quad (3)$$

$$\frac{v_f}{[E]_t} = \frac{k_{\text{cat}} K_i^* \times \beta k_{\text{cat}} [I]}{[I] + K_i^*} \quad (4)$$

Sequence Similarity Network Data Set Sources and

Curation. Four structures representing the known reaction diversity in the DRE-TIM metallolyase superfamily⁵ (pdb_ids: 20 Invm,⁷ 1rqb,²¹ 3hq1,²² and 2qf7²³) were used as a starting point for identifying new homologous sequences through a series of BLAST²⁴ searches against UniRef100 (Uniprot Release 2012_02).²⁵ All hits from an initial BLAST search with an E -value less than 1×10^{-5} were kept. This new data set of 13230 sequences was filtered using HMMER 3.0 beta²⁶ such that any sequence that did not score against the Pfam²⁷ HMM for PF00682: HMGL-like was dropped. This new data set (4889 unique sequences) was then filtered to 40% identity using CD-HIT v4.5.6,²⁸ resulting in 54 sequences which were used as a query against UniRef100 again (73764 hits) and filtered in the same manner to produce the final superfamily set (8817 unique sequences).

Generation of Sequence Similarity Networks. Sequence similarity networks were generated using the Pythoscape program.²⁹ Briefly, the final superfamily set of proteins (8817 unique sequences) were imported into a MySQL database and defined as nodes. All-by-all Blast2seq²⁴ runs were performed to define edges between nodes, and this information was uploaded into the MySQL database. Additional information associated with the proteins, such as taxonomic information and pdb structures, was mapped to the MySQL database through the Pythoscape interface. A Cytoscape³⁰-readable xgmml file was exported from Pythoscape to produce networks that could be explored interactively using Cytoscape. Mapping the network nodes with additional data, such as kinetic characterization data, was done within the Cytoscape program.

Because of the large number of edges at the superfamily level, representative nodes and edges were required for visualization of the full network in Cytoscape. A representative node was defined by CD-HIT v4.5.6²⁸ as a cluster of sequences sharing greater than 60% identity. Representative edges, defined here as the mean BLAST E -value between the set of sequences contained within two connected representative nodes, are shown only if their BLAST scores have a statistical significance value less than (better than) a user-defined E -value cutoff. (See network figure legends for specific cutoffs used.)

Structure-Guided Sequence Alignments. The best aligning chains of MtIPMS (pdb_id: 3u6w³¹), LiCMS (3blf⁹), and SpHCS (2zyf³²) from the Claisen condensation-like (CC-like) subgroup were aligned using the Needleman–Wunsch algorithm³³ as implemented in the Matchmaker program³⁴ in

Chimera.³⁵ A multiple sequence alignment based on the structure alignment was created through a companion program, Match → Align. The sequence alignment was then refined by the eye using the aligned structures as a guide. Sequence-based alignments of each cluster were generated based on cluster membership using MAFFT, version 6.³⁶

RESULTS

Steady State Kinetics. In order to investigate the catalytic role of the active site helical residues, site-directed mutagenesis was performed. All enzyme variants were characterized using circular dichroism spectroscopy to ensure that substitutions did not affect the overall fold of the enzyme (Figure S1, Supporting Information). The kinetic parameters determined for the enzyme variants are listed in Table 1. The substitutions made to R80 and D81 abolished IPMS activity. However, D81A MtIPMS has the ability to catalyze the hydrolysis of AcCoA producing acetate, KIV, and free CoA (data not shown). In the absence of KIV, kinetic parameters for the hydrolysis of AcCoA were determined ($k_{\text{cat}} = 4.6 \pm 0.3 \text{ min}^{-1}$ and $K_{\text{AcCoA}} = 283 \pm 54 \mu\text{M}$). The L79A variant displayed a 60-fold decrease in $k_{\text{cat}}/K_{\text{KIV}}$ when compared to that of the wild-type enzyme. The N83A and Q84A variants displayed a 150–250-fold decrease in $k_{\text{cat}}/K_{\text{AcCoA}}$, while the $k_{\text{cat}}/K_{\text{KIV}}$ parameters are relatively unchanged. Q84A MtIPMS also exhibits a 30-fold decrease in k_{cat} . N83E MtIPMS was determined to be inactive.

Effect of Mutations on L-Leucine Inhibition. Kinetic assays were performed to determine the effect of L-leucine binding on the active enzyme variants. In the presence of L-leucine, all of the enzyme variants displayed slow-onset inhibition kinetics (nonlinear biphasic progression curves) similar to the results seen with the wild-type enzyme.³⁷ Inhibition parameters for two-step, slow-onset inhibitors are described by two terms, K_i , which describes the inhibition constant for the initial enzyme–inhibitor complex, and K_i^* , which describes the overall inhibition constant for the two-step inhibition mechanism.¹⁸ The initial and steady state velocities were determined by varying concentrations of leucine (as described in the Materials and Methods) and fit to eqs 3 and 4 to determine the inhibition parameters K_i and K_i^* , respectively (Figure S2, Supporting Information). Inhibition parameters are shown in Table 2. There is no drastic change in the values for the inhibition constants K_i and K_i^* relative to those determined with the wild-type enzyme.

Table 2. Leucine Inhibition Parameters^a

enzyme	K_i (μM)	K_i^* (μM)	β
WT	12 ± 3	2.3 ± 0.2	0.1
L79A	19 ± 3	4.0 ± 0.5	0
N83A ^b	8.3 ± 2.2	3.2 ± 0.5	0.2
Q84A	13 ± 4	4.8 ± 0.5	0.04

^aDetermined using a standard reaction mixture of 50 mM potassium phosphate buffer (pH 7.5), 12 mM MgCl₂, 50 μM DTP, 5–10 times the K_m values of AcCoA and KIV, and varied concentrations of leucine (0–120 μM). ^bDetermined using 750 μM AcCoA.

Sequence Similarity Network for the DRE-TIM Metallolyase Superfamily. The overall sequence similarity network for representative members of the DRE-TIM metallolyase superfamily (1261 representative nodes representing 8817 unique sequences) is shown in Figure 2. Nodes represent sets of proteins sharing greater than 60% identity and are colored if

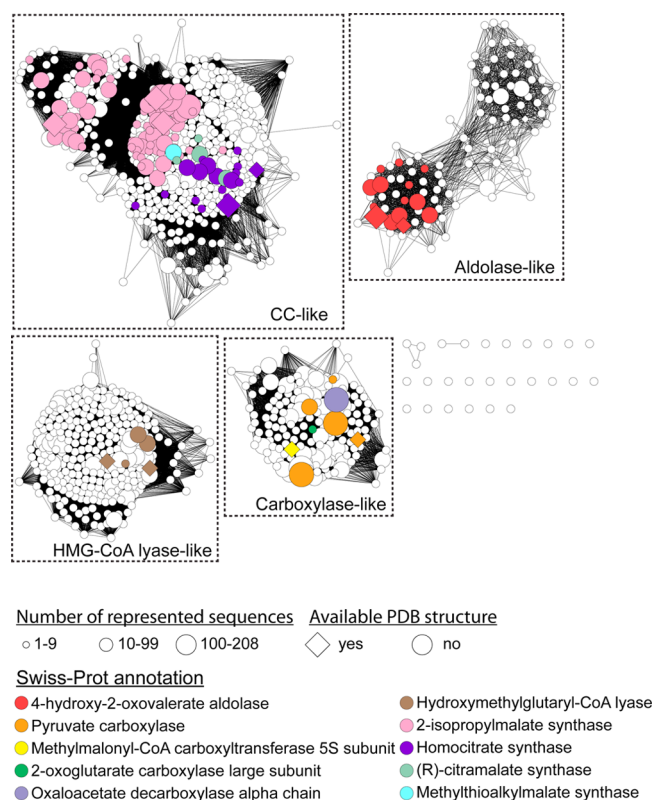


Figure 2. Representative sequence similarity network for the DRE-TIM metallolyase superfamily. Each node (1261 representative nodes) represents a group of protein sequences sharing greater than 60% identity (8817 unique sequences). Edges are drawn if the mean similarity between a pair of nodes is less than an E -value threshold cutoff of 1×10^{-26} (median alignment length = 379, and median percent identity of pairwise comparisons = 29%). The network is displayed using the organic layout in Cytoscape. Disconnected nodes have been moved for clarity, as the distance between disconnected nodes has no meaning in this layout. The relative sizes of the nodes represent the number of sequences in the representative cluster (small nodes, 1–9 sequences; medium nodes, 10–99 sequences; large nodes, 100–208 sequences). Colored nodes represent sets of proteins in which at least one protein has been annotated in Swiss-Prot with a functional activity according to the inset color key. Conversely, nodes are left uncolored if no protein has been reported to be characterized in Swiss-Prot. Diamond shaped nodes contain at least one protein with a solved crystal structure in the PDB.²⁰ Dotted boxes display the four subgroups of the superfamily, as defined by clustering patterns: HMG-CoA lyase-like, CC-like (comprising proteins of the IPMS, homocitrate synthase, and citramalate synthase families, along with unknowns), aldolase-like, and carboxylase-like.

at least one protein in the node has been annotated in Swiss-Prot³⁸ with a functional activity. Nodes shown as diamonds contain at least one sequence for which a three-dimensional structure has been reported in the Protein Databank.²⁰ The vast majority of proteins have not been biochemically characterized and are thus left as proteins of unknown function (“unknowns”, uncolored nodes). Correct annotation of these unknowns will require identification of their specificity determinants and is beyond the scope of this work. Four identifiable subgroups are defined based on clustering patterns and characterized functions: Claisen condensation-like (CC-like), carboxylase-like, lyase-like, and aldolase-like.

Sequence Similarity Network for the Claisen Condensation-Like Subgroup. The 583 representative nodes

(representing 4298 nonidentical proteins) of the CC-like subgroup defined in Figure 2 were selected for the creation of a new network that includes all of the representative nodes of the subgroup and subjected to a more stringent *E*-value cutoff filter (Figure 3). The edges shown in this figure represent a

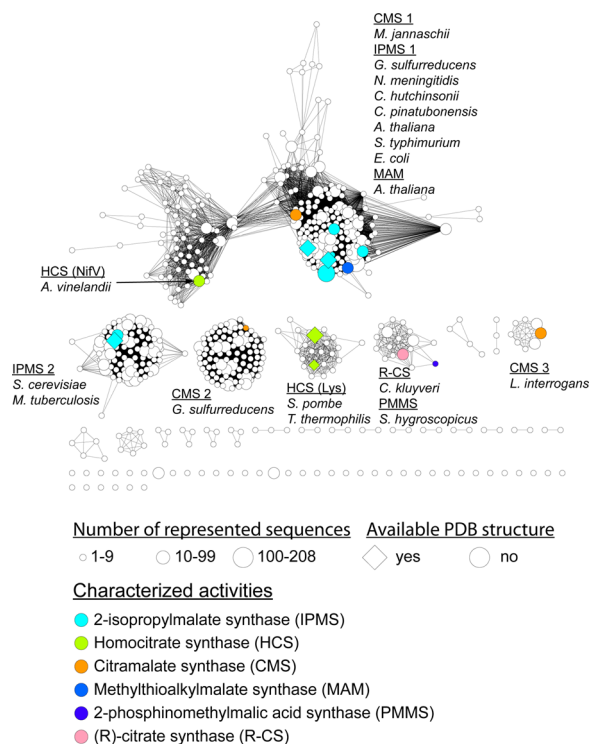
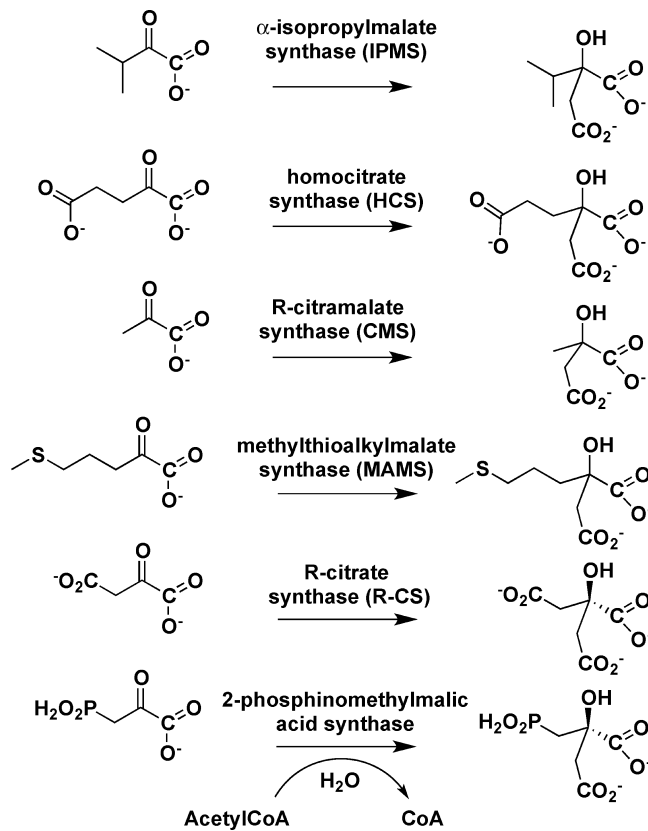


Figure 3. Representative sequence similarity network for the CC-like subgroup. Each node (583 representative nodes) represents a group of protein sequences sharing greater than 60% identity (4298 unique sequences). Edges are drawn if the similarity between a pair of nodes is better than an *E*-value threshold cutoff of 1×10^{-80} (median alignment length = 377, and median percent identity of pairwise comparisons = 43%). Colored nodes represent sets of proteins in which at least one protein has confirmed *in vitro* activity according to the inset color key. Organism names indicate which protein within the representative node has been characterized. Conversely, nodes are left uncolored if no protein has been reported to be characterized in the literature. Diamond shaped nodes contain at least one protein with a solved crystal structure in the PDB.²⁰

mean pairwise alignment score for all-by-all comparisons of these 4298 sequences that is better than an *E*-value cutoff of 1×10^{-80} , with a median percent identity of 43% and a median alignment length of 377 residues. Analyses of the sequences found in this network reveal enzymes with at least six unique substrate specificities for the Claisen condensation reaction with AcCoA (Scheme 3). Representative nodes in Figure 3 are colored on the basis of having at least one protein reported to have *in vitro* characterization of enzymatic activity for IPMS,^{39–44} citramalate synthase (CMS),^{9,45,46} homocitrate synthase (HCS),^{47,48} methylthiolalkylmalate synthase (MAM),⁴⁹ R-citrate synthase (R-CS),⁵⁰ and 2-phosphinomethylmalic synthase (PMMS).⁵¹ A full table of characterized enzymes with Uniprot identifiers is shown in Table S2 (Supporting Information). Functional assignments shown in Figure 2 are in good agreement with reported Swiss-Prot functional annotation (Figure S3, Supporting Information). The largest cluster contains significant functional diversity, with

Scheme 3



IPMS, CMS, MAM, and HCS activity represented. Interestingly, reported IPMS, CMS, and HCS activities can be found in multiple clusters. This is consistent with a report proposing multiple origins for IPMS⁵² and could be suggestive of additional functional promiscuity.

DISCUSSION

Targeting evolutionarily conserved residues for site-directed mutagenesis is a common approach used to investigate the chemical mechanism and specificity determinants of an enzyme. In light of the rapid increase in genomic sequences available, the identification of “strictly” conserved residues in a given enzyme is complicated by misannotation in public databases⁵³ and the identification of functionally diverse enzyme superfamilies that utilize conserved active site architecture associated with a fundamental catalytic strategy or partial reaction to catalyze a variety of chemical reactions. The addition of “genomic enzymology,”⁶ describing enzyme catalysis from the context of structure–function relationships among homologous members of enzyme superfamilies, to the toolbox of mechanistic enzymology provides an organizational framework, aids in our understanding of the evolution of function, and helps interpret structure-based mechanisms studies. Here, sequence-similarity networks have been used to summarize some of these relationships across the entire membership of the DRE-TIM metallolyase superfamily and assist in the interpretation of our mechanistic analysis of a conserved active-site helix in MtIPMS.

Analysis of the DRE-TIM Metallolyase Superfamily Network. The superfamily network shown in Figure 2 depicts some of the benefits and challenges associated with a network-based interpretation. The four enzyme activities originally

proposed in the DRE-TIM metallolyase superfamily are easily identified in the Claisen condensation-like (CC-like), carboxylase-like, lyase-like, and aldolase-like subgroups. Functional diversity within each subgroup is well established for the CC-like and carboxylase-like subgroups including the identification of two less-documented activities: 2-phosphinomethylmalic acid synthase⁵¹ and α -ketoglutarate carboxylase.⁵⁴ Identification of reported functional diversity provides a strong foundation for the discovery of differentially conserved residues specific to each reported activity. Once the roles of these residues are confirmed, sequences of unknown function containing the confirmed residues can be annotated. Additionally, identification of sequences containing unique residues at positions important for substrate selectivity can offer new hypotheses for screening for new functionalities. Functional diversity is also suggested by inspection of the aldolase-like cluster in Figure 2. All of the nodes containing a Swiss-Prot reviewed sequence for 4-OH-2-ketovalerate aldolase activity are located in a single region, while the other half of the subgroup remains unexplored and may contain new functionalities.

Analysis of the CC-Like Subgroup Network. As the stringency for drawing edges between representative nodes is increased, the CC-like subgroup breaks into multiple clusters (Figure 3). One of the more interesting results is the identification of the main three activities (IPMS, CMS, and HCS) in multiple clusters. Results from the analysis of other superfamilies indicate that multiple clusters of an activity can be the result of multiple evolutions from distinct but related progenitors described as “pseudoconvergent evolution”.^{55,56} In these cases, different cluster memberships correlate with differences in substrate or stereochemical specificity and differential conservation of residues involved in substrate selectivity. Currently, the IPMS clusters contain the most characterized members. Despite differences in sequence identity (<20% average sequence identity between clusters and 50% average sequence identity within a cluster), no significant differences in function or specificity have been identified, and residues shown to interact with KIV are conserved in both clusters. However, membership in the two IPMS clusters is similar to that reported from a phylogenetic analysis of IPMS sequences suggesting multiple origins for IPMS genes,⁵² and differential conservation of residues not directly involved in substrate selectivity can be identified as described below. In the cases of the HCS- and CMS-containing clusters, differential conservation of active site residues predicted to be in contact with the α -ketoacid substrates suggests these activities are examples of pseudoconvergent evolution, but more rigorous phylogenetic and functional analyses are required to further support this hypothesis.

The distribution of regulatory domains in the CC-like subgroup also provides a framework for the identification of boundaries within the LeuA dimer regulatory domain. Recoloring the CC-like subgroup network on the basis of predicted domain architecture indicates there are four main clusters which contain sequences predicted to contain an N-terminal DRE-TIM metallolyase catalytic domain and a C-terminal LeuA dimer regulatory domain: IPMS1/CMS1/MAM, IPMS2, CMS2, and CMS3 (Figure S4, Supporting Information). Unfortunately, only two structures of the LeuA dimer domain have been reported (MtIPMS in IPMS2 cluster and CMS from *Leptospira interrogans* in CMS3 cluster), limiting the ability to draw structural comparisons. Phenomenological mechanisms of allosteric regulation (i.e., V-type or K-type)

have been reported for members of the IPMS1, IPMS2, and CMS3 clusters and suggest that multiple mechanisms of regulation are represented within each cluster ruling out allosteric mechanism as a characteristic driving cluster membership.² The allosteric mechanism of V-type allosteric inhibition in MtIPMS is due to the perturbation of the hydrolysis step in the reaction.⁵⁷ Current work is focused on determining if this mechanism is conserved in members of other clusters exhibiting V-type allosteric regulation.

Superfamily Analysis of Helix Residues in Catalysis.

An examination of conservation patterns of residues in the active-site α helix found in characterized members of the superfamily indicates the only amino acid strictly conserved is an arginine residue (Figure 4), suggesting a critical role in

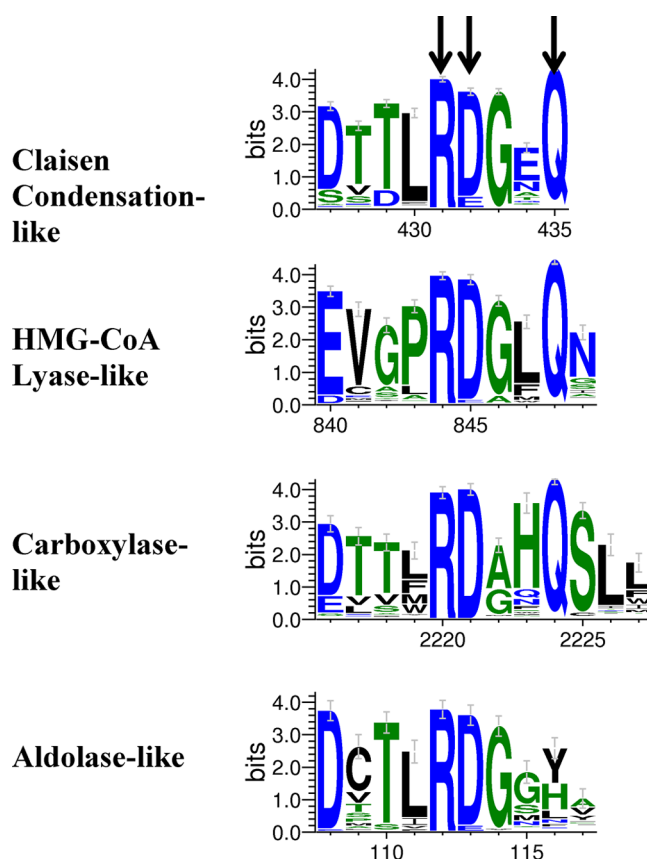


Figure 4. Hidden-Markov Model logos for the active site helix of DRE-TIM metallolyase superfamily members. Logos were generated from alignments of sequences in each boxed cluster in Figure 2 using Web Logo.⁶⁴ Arrows indicate the main residues discussed in the article.

catalysis. As described above, substitution of R80 in MtIPMS with alanine or lysine results in inactive enzymes. Our results are in agreement with similar studies on substitution at the arginine position in representative members of the carboxylase-like,¹² aldolase-like,¹³ and lyase-like subgroups.¹¹ Thus, as predicted by conservation, the active site helix arginine is critical for activity across the superfamily. Mechanistically, the arginine residue is predicted to assist in the stabilization of the enolate intermediate in each of the subgroups. However, structural comparison of four representative superfamily members indicates differences in the location of the enolate ion stabilized by the arginine residue, with aldolase-like enzymes having an alternate location (Figure S5, Supporting

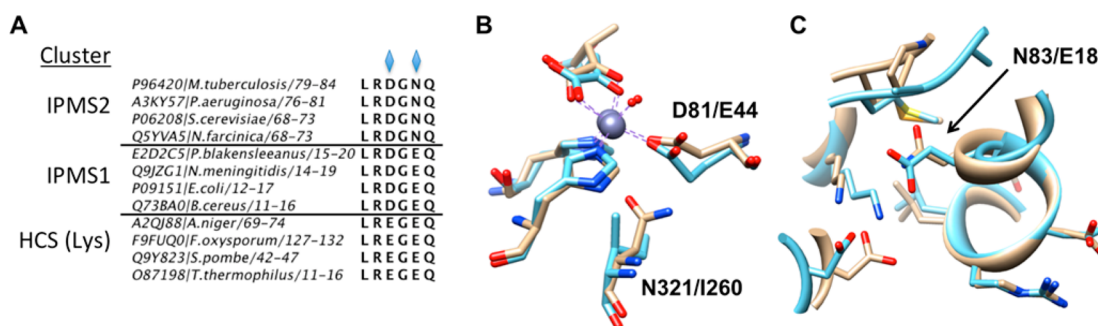


Figure 5. Examples of differential conservation of helix residues in the CC-like subgroup. (A) Alignment of helix residues generated from sequences belonging to IPMS1, IPMS2, and HCS (Lys) clusters. Residues exhibiting differential conservation and mentioned in the discussion are highlighted with a diamond. (B) Superposition of MtIPMS (1sr9,⁶³ brown) and HCS from *S. pombe* (3ivt,¹⁰ blue). (C) Superposition of MtIPMS (1sr9, brown) and NmIPMS (3rmj,³¹ blue).

Information). This suggests that despite differences in substrates and reactions, the arginine is essential to maintain the electronic and catalytic requirements fundamental to catalysis in all members of the superfamily.

A carboxylic group is strictly conserved in the position adjacent to arginine in the DRE-TIM metallolyase superfamily (D81 in MtIPMS). In HMG-CoA lyase, substitution of the aspartate with alanine, glycine, or histidine results in a 10^4 – 10^5 -fold decrease in specific activity in the human enzyme, while substitution with glutamate decreases activity 10-fold.⁵⁸ In MtIPMS, substitution with alanine or histidine results in an inactive enzyme; however, a 10^5 decrease would be at the limit of the spectrophotometric assay for CoA detection. While the aspartate (the “D” in the DRE motif) is favored, in some superfamily sequences a glutamate is found adjacent to the conserved arginine instead (Figure 4). This is discussed below with respect to epistatic interactions within the CC-like subgroup.

The other strongly conserved residue in the helix is a glutamine residue found three positions after the aspartate (Q84 in MtIPMS) (Figures 1 and 4). The role of the residue at this position appears to be a major factor in the development of reaction diversity for the DRE-TIM metallolyase superfamily. Members of the carboxylase-like and CC-like subgroups both have conserved glutamine residues at this position. However, substitutions at this position produce drastically different catalytic outcomes. In pyruvate carboxylase from *R. etli*, substitution of glutamine with alanine or asparagine inactivates pyruvate carboxylase activity and the ability of the enzyme to enolize pyruvate.¹² In sharp contrast to pyruvate carboxylase, substitution of the conserved glutamine with alanine in MtIPMS and homocitrate synthase from *S. pombe* results in enzymes with only a 10- to 30-fold decrease in the value for k_{cat} relative to the wild-type enzymes.¹⁰ This result suggests the residue plays unique roles in the mechanism of each enzyme. In pyruvate carboxylase, the glutamine is hypothesized to orient pyruvate and maintains an interaction with the helix arginine residue even in the absence of pyruvate.^{59,60} In the CC-like subgroup, the glutamine residue can interact with the carbonyl of AcCoA. AcCoA makes additional binding interactions with the enzyme such that disruption of interaction with Q84 is well tolerated. Additionally, reported crystal structures of MtIPMS exhibit different conformations for Q84 (including interactions with AcCoA, R80, R427, and E317) or are missing electron density for the side chain suggesting flexibility at this position. Residues R427 and E317 are strictly conserved in the CC-like

subgroup, and future mutagenesis studies will investigate the importance of interactions of the glutamine with E317 and R427 in MtIPMS.

A second piece of evidence concerning the importance of the glutamine position in reaction specificity is seen in the aldolase-like subgroup. From the similarity networks, it can be seen that the aldolase-like subgroup contains tyrosine, histidine, or leucine in place of glutamine (Figures 1 and 4). In the class II pyruvate aldolase BphI (see Scheme 1 for the 4-OH-2-ketovalerate aldolase reaction), the histidine residue acts as a general base to deprotonate the C-4 hydroxyl group in the first step of the reaction as substitution with alanine decreases the k_{cat} value by 50-fold relative to that of the wild-type and results in the loss of a pK_a value in the pH rate profile.¹³ The unique use of histidine as the general base by the aldolase-like subgroup is consistent with the alternate position of the enolate ion relative to the other subgroups. Aldolase-like subgroup members with tyrosine or leucine at this position, however, would be unable to utilize this mechanism. This suggests that these systems utilize architecturally distinct amino acids as the catalytic base or that members lacking the histidine residue catalyze a different reaction. In support of this hypothesis, a member of the aldolase-like subgroup from *M. tuberculosis* (Rv3469c) containing a tyrosine at this position was reported to lack aldolase activity.⁶¹ The protein did exhibit oxaloacetate decarboxylase activity, albeit with a very low catalytic efficiency ($1.6 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$). As a β -keto decarboxylation, this reaction does not require a base to generate the enolate intermediate.

The residues flanking the arginine and aspartate on the helix are highly similar in all of the DRE-TIM metallolyase superfamily with a hydrophobic residue preceding the arginine residue and a glycine/alanine following the aspartate residue. These residues are likely important in stabilizing the overall architecture of the active site α helix. The main effect of the L79A substitution in MtIPMS is in agreement with this conjecture as the only significant perturbation is a 20-fold increase in the K_m value determined for KIV. This is the only report of a substitution at this position, and the alanine/glycine position has not previously been investigated; therefore, limited context is available to interpret the significance of this result with respect to other members of the superfamily.

Differential Conservation of Helix Residues in the Claisen Condensation-Like Subgroup. Analysis of the activities reported for members of the CC-like subgroup (Scheme 3) indicate this subgroup is specificity diverse, i.e., these enzymes catalyze a common reaction with varying

substrate specificities. The majority of sequences are annotated as either IPMS, homocitrate synthase (HCS), or R-citramalate synthase (CMS) (Figure S3, Supporting Information). The arginine and aspartate (or glutamate) have been previously investigated in HCS and CMS with kinetic results similar to those reported here.^{9,10} This is not surprising as the helix residues appear to be required for catalysis, while residues involved in substrate selectivity are located elsewhere in the active site. As noted above, each activity has been experimentally demonstrated in multiple clusters of the network shown in Figure 3, suggesting differences in the evolutionary conservation of amino acids between clusters for enzymes catalyzing identical reactions, including possible epistatic constraints within each cluster. Analysis of differential conservation can provide insight into strategies for substrate selectivity or account for differences in other properties, such as allosteric regulation. With respect to the active site helix under investigation, two examples of differential conservation within the CC-like subgroup can be readily identified and are described below.

While the DRE motif aspartate is commonly found as a ligand to the metal ion, enzymes involved in fungal lysine biosynthesis (HCS (Lys) cluster of Figure 3) have a strictly conserved glutamate residue in the analogous position (i.e., ERE-motif) (Figure 5A). Structural studies indicate glutamate acts as a ligand to the divalent cation, similar to the more common aspartate.¹⁰ A comparison of structures from HCS (Lys) and the IPMS clusters indicates that the glutamate substitution is linked to a conserved compensatory/epistatic substitution of an isoleucine in place of an asparagine (D81/N321 in MtIPMS and E44/I251 in HCS from *S. pombe*) (Figure 5B). Without the N → I substitution, the glutamate residue would cause substantial steric clashes in the active site. Moreover, the asparagine residue is well-conserved within the CC-like and HMG-CoA lyase-like subgroups and can act as an additional ligand to the metal ion in IPMS⁴¹ and HMG-CoA lyase.⁶² Future mutagenesis studies on MtIPMS will determine if the (D → E)/(N → I) mutations are tolerated in enzymes from other clusters in the CC-like subgroup and explore a role for the alternative metal architecture.

Comparing active site helices from the two IPMS clusters provides a second example of differential conservation. In IPMS enzymes, the residue corresponding to N83 in MtIPMS is most often an asparagine or glutamate. From the sequence similarity network, it can be concluded that the identity of the residue at this position can be used to categorize each sequence into either the IPMS1 (glutamate) or IPMS2 (asparagine) cluster (Figure 5A). In MtIPMS, the residue is an asparagine (N83) in agreement with its placement in the IPMS2 cluster. In MtIPMS, substitution of N83 with alanine results in a 25-fold elevation in the K_m value for AcCoA and a 10-fold decrease in the k_{cat} value. Surprisingly, when the N83E substitution is made, no activity is detectable. A comparison of structures from MtIPMS and IPMS from *Neisseria meningitidis* (NmIPMS, a member of IPMS1) fails to identify a conserved compensating substitution in the region near the asparagine/glutamate residue (Figure 5C). In fact, *in silico* modeling of the substitution suggests that the glutamate can be accommodated in the MtIPMS scaffold without causing steric clashes. Circular dichroism studies are consistent with a properly folded enzyme ruling out a large change in structure in the loss of activity. One possibility for the loss of activity in the N83E variant of

MtIPMS is that this position is constrained by long-range epistatic factors specific to each cluster.

Role of Helix Residues in the Allosteric Mechanism of MtIPMS. Residues on the conserved helix were recently implicated in the allosteric mechanism of MtIPMS on the basis of L-leucine induced changes in dynamics measured by backbone amide hydrogen/deuterium exchange.¹⁴ These results suggested a plausible mechanism for the allosteric regulation of MtIPMS involving perturbation of the helix by L-leucine binding. However, all active variants exhibit slow-onset inhibition in the presence of L-leucine and give similar values for inhibition parameters relative to those determined with the wild-type enzyme. As the slow-onset mechanism has recently been linked to the movement of a loop in the regulatory domain,¹⁷ it is not surprising that biphasic kinetics is exhibited. However, the lack of perturbation to the inhibition parameters suggests that residues on the helix do not directly participate in the allosteric mechanism, although direct involvement of R80 or D81 cannot be ruled out as substitution at these positions creates inactive enzymes. More recently, kinetic isotope effects were used to identify the hydrolysis step as the allosteric target of inhibition.⁵⁷ From the integrated mutagenesis and bioinformatics results, the main role of the conserved active site helix is stabilization of the common enolate intermediate, suggesting that the catalytic machinery for hydrolysis (and the target for allosteric regulation) in the CC-like subgroup lies elsewhere in the active site.

CONCLUSIONS

This work represents the first large-scale bioinformatics analysis of the DRE-TIM metallolyase superfamily, used here to provide enhanced context for understanding the mechanistic contributions of residues on an active site α helix conserved across characterized members of the superfamily. Taken together, the experimental and bioinformatics results indicate a critical role for the DRE motif arginine residue for all of the different reactions currently known across the superfamily, most likely through the stabilization of the common enolate ion. The DRE motif aspartate is also essential for metal ion interactions, although alternate architectures using glutamate instead of aspartate are possible. It will be of interest to determine if evolution of glutamate-containing helices provides unique properties to enzymes when compared with the canonical aspartate-containing homologues. Finally, with the similarity network in place, hypotheses concerning mechanisms of substrate and reaction specificity can be investigated for members of the other subgroups.

ASSOCIATED CONTENT

Supporting Information

Oligonucleotide sequences for site-directed mutagenesis and references for reported activities in the CC-like subgroup; circular dichroism spectra; L-leucine inhibition kinetics; visualizations of the CC-like subgroup indicating Swiss-Prot reported activities and predicted domain organization; and locations of the enolate ion predicted for each activity. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Department of Chemistry, The University of Alabama, Box 870336, Tuscaloosa, AL 35406. Tel: 205-348-8349. Fax: 205-348-9104. E-mail: pfrantom@ua.edu.

Present Address

[#](M.A.H.) Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E17-539, Cambridge, MA 02139.

Author Contributions

[†]A.K.C. and M.A.H. contributed equally to this study.

Funding

This work was supported by startup funding from The University of Alabama and NSF grant MCB-1254077 (to P.A.F.) and by R01 GM60595 (to P.C.B). J.L.J. was funded in part by an NSF-REU award (CHE-1004098) to the Department of Chemistry at The University of Alabama.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). We thank Dr. Holly Ellis (Auburn University) for her assistance in circular dichroism spectroscopy.

ABBREVIATIONS

AcCoA, acetyl-Coenzyme A; CC-like, Claisen-condensation-like; CMS, citramalate synthase; DTP, 4–4'-dithio-dipyridine; HCS, homocitrate synthase; HDX, hydrogen–deuterium exchange; HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A; IPM, α -isopropylmalate; IPMS, isopropylmalate synthase; KIV, α -ketoisovalerate; LiCMS, citramalate synthase from *Leptospira interrogans*; MtIPMS, α -isopropylmalate synthase from *Mycobacterium tuberculosis*; NmIPMS, α -isopropylmalate synthase from *Neisseria meningitidis*; SpHCS, homocitrate synthase from *Schizosaccharomyces pombe*; TEA, triethanolamine

REFERENCES

- (1) McAdam, R. A., Weisbrod, T. R., Martin, J., Scuderi, J. D., Brown, A. M., Cirillo, J. D., Bloom, B. R., and Jacobs, W. R., Jr. (1995) In vivo growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect. Immun.* 63, 1004–1012.
- (2) Frantom, P. A. (2012) Structural and functional characterization of alpha-isopropylmalate synthase and citramalate synthase, members of the LeuA dimer superfamily. *Arch. Biochem. Biophys.* 519, 202–209.
- (3) Smith, C. V., Huang, C. C., Miczak, A., Russell, D. G., Sacchetti, J. C., and Honer zu Bentrop, K. (2003) Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 278, 1735–1743.
- (4) Remington, S. J. (1992) Structure and mechanism of citrate synthase. *Curr. Top. Cell Regul.* 33, 209–229.
- (5) Forouhar, F., Hussain, M., Farid, R., Benach, J., Abashidze, M., Edstrom, W. C., Vorobiev, S. M., Xiao, R., Acton, T. B., Fu, Z., Kim, J. J., Miziorko, H. M., Montelione, G. T., and Hunt, J. F. (2006) Crystal structures of two bacterial 3-hydroxy-3-methylglutaryl-CoA lyases suggest a common catalytic mechanism among a family of TIM barrel metalloenzymes cleaving carbon-carbon bonds. *J. Biol. Chem.* 281, 7533–7545.
- (6) Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* 70, 209–246.
- (7) Manjasetty, B. A., Powlowski, J., and Vrieland, A. (2003) Crystal structure of a bifunctional aldolase-dehydrogenase: sequestering a

reactive and volatile intermediate. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6992–6997.

(8) Adina-Zada, A., Zeczycki, T. N., St Maurice, M., Jitrapakdee, S., Cleland, W. W., and Attwood, P. V. (2012) Allosteric regulation of the biotin-dependent enzyme pyruvate carboxylase by acetyl-CoA. *Biochem. Soc. Trans.* 40, 567–572.

(9) Ma, J., Zhang, P., Zhang, Z., Zha, M., Xu, H., Zhao, G., and Ding, J. (2008) Molecular basis of the substrate specificity and the catalytic mechanism of citramalate synthase from *Leptospira interrogans*. *Biochem. J.* 415, 45–56.

(10) Bulfer, S. L., Scott, E. M., Couture, J. F., Pillus, L., and Trievel, R. C. (2009) Crystal structure and functional analysis of homocitrate synthase, an essential enzyme in lysine biosynthesis. *J. Biol. Chem.* 284, 35769–35780.

(11) Tuinstra, R. L., Wang, C. Z., Mitchell, G. A., and Miziorko, H. M. (2004) Evaluation of 3-hydroxy-3-methylglutaryl-coenzyme A lyase arginine-41 as a catalytic residue: use of acetyldithio-coenzyme A to monitor product enolization. *Biochemistry* 43, 5287–5295.

(12) Duangpan, S., Jitrapakdee, S., Adina-Zada, A., Byrne, L., Zeczycki, T. N., St Maurice, M., Cleland, W. W., Wallace, J. C., and Attwood, P. V. (2010) Probing the catalytic roles of Arg548 and Gln552 in the carboxyl transferase domain of the *Rhizobium etli* pyruvate carboxylase by site-directed mutagenesis. *Biochemistry* 49, 3296–3304.

(13) Baker, P., Carere, J., and Seah, S. Y. (2011) Probing the molecular basis of substrate specificity, stereospecificity, and catalysis in the class II pyruvate aldolase, BphI. *Biochemistry* 50, 3559–3569.

(14) Frantom, P. A., Zhang, H. M., Emmett, M. R., Marshall, A. G., and Blanchard, J. S. (2009) Mapping of the allosteric network in the regulation of alpha-isopropylmalate synthase from *Mycobacterium tuberculosis* by the feedback inhibitor L-leucine: solution-phase H/D exchange monitored by FT-ICR mass spectrometry. *Biochemistry* 48, 7457–7464.

(15) Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4, e4345.

(16) Zhu, W. W., Wang, C., Jipp, J., Ferguson, L., Lucas, S. N., Hicks, M. A., and Glasner, M. E. (2012) Residues required for activity in *Escherichia coli* o-succinylbenzoate synthase (OSBS) are not conserved in all OSBS enzymes. *Biochemistry* 51, 6171–6181.

(17) Casey, A. K., Baugh, J., and Frantom, P. A. (2012) The slow-onset nature of allosteric inhibition in alpha-isopropylmalate synthase from *Mycobacterium tuberculosis* is mediated by a flexible loop. *Biochemistry* 51, 4773–4775.

(18) Morrison, J. F., and Walsh, C. T. (1988) The behavior and significance of slow-binding enzyme inhibitors. *Adv. Enzymol. Relat. Areas Mol. Biol.* 61, 201–301.

(19) de Carvalho, L. P., Frantom, P. A., Argyrou, A., and Blanchard, J. S. (2009) Kinetic evidence for interdomain communication in the allosteric regulation of alpha-isopropylmalate synthase from *Mycobacterium tuberculosis*. *Biochemistry* 48, 1996–2004.

(20) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.

(21) Hall, P. R., Zheng, R., Antony, L., Pusztai-Carey, M., Carey, P. R., and Yee, V. C. (2004) Transcarboxylase 5S structures: assembly and catalytic mechanism of a multienzyme complex subunit. *EMBO J.* 23, 3621–3631.

(22) Koon, N., Squire, C. J., and Baker, E. N. PDB ID: 3HQ1, Probing the Active Site of *M. tuberculosis* LeuA.

(23) St Maurice, M., Reinhardt, L., Surinya, K. H., Attwood, P. V., Wallace, J. C., Cleland, W. W., and Rayment, I. (2007) Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* 317, 1076–1079.

(24) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

- (25) Consortium, T. U. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* 41, D43–47.
- (26) Eddy, S. R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inf.* 23, 205–211.
- (27) Punta, M., Coghill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., and Finn, R. D. (2012) The Pfam protein families database. *Nucleic Acids Res.* 40, D290–301.
- (28) Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659.
- (29) Barber, A. E., II, and Babbitt, P. C. (2012) Pythoscape: a framework for generation of large protein similarity networks. *Bioinformatics* 28, 2845–2846.
- (30) Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- (31) Huisman, F. H., Koon, N., Bulloch, E. M., Baker, H. M., Baker, E. N., Squire, C. J., and Parker, E. J. (2012) Removal of the C-terminal regulatory domain of alpha-isopropylmalate synthase disrupts functional substrate binding. *Biochemistry* 51, 2289–2297.
- (32) Okada, T., Tomita, T., Wulandari, A. P., Kuzuyama, T., and Nishiyama, M. (2010) Mechanism of substrate recognition and insight into feedback inhibition of homocitrate synthase from *Thermus thermophilus*. *J. Biol. Chem.* 285, 4195–4205.
- (33) Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453.
- (34) Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., and Ferrin, T. E. (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* 7, 339.
- (35) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- (36) Katoh, K., and Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings Bioinf.* 9, 286–298.
- (37) de Carvalho, L. P., Argyrou, A., and Blanchard, J. S. (2005) Slow-onset feedback inhibition: inhibition of *Mycobacterium tuberculosis* alpha-isopropylmalate synthase by L-leucine. *J. Am. Chem. Soc.* 127, 10004–10005.
- (38) UniProt, C. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42, D191–198.
- (39) de Carvalho, L. P., and Blanchard, J. S. (2006) Kinetic and chemical mechanism of alpha-isopropylmalate synthase from *Mycobacterium tuberculosis*. *Biochemistry* 45, 8988–8999.
- (40) de Kraker, J. W., Luck, K., Textor, S., Tokuhisa, J. G., and Gershenzon, J. (2007) Two *Arabidopsis* genes (IPMS1 and IPMS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. *Plant Physiol.* 143, 970–986.
- (41) Huisman, F. H., Hunter, M. F., Devenish, S. R., Gerrard, J. A., and Parker, E. J. (2010) The C-terminal regulatory domain is required for catalysis by *Neisseria meningitidis* alpha-isopropylmalate synthase. *Biochem. Biophys. Res. Commun.* 393, 168–173.
- (42) Ulm, E. H., Bohme, R., and Kohlhaw, G. (1972) Alpha-isopropylmalate synthase from yeast: purification, kinetic studies, and effect of ligands on stability. *J. Bacteriol.* 110, 1118–1126.
- (43) Webster, R. E., and Gross, S. R. (1965) The alpha-isopropylmalate synthetase of neurospora. I. The kinetics and end product control of alpha-isopropylmalate synthetase function. *Biochemistry* 4, 2309–2318.
- (44) Wiegel, J., and Schlegel, H. G. (1977) Alpha-isopropylmalate synthase from *Alcaligenes eutrophus* H 16. III. Endproduct inhibition and its relief by valine and isoleucine. *Arch. Microbiol.* 114, 203–210.
- (45) Howell, D. M., Xu, H., and White, R. H. (1999) (R)-Citramalate synthase in methanogenic archaea. *J. Bacteriol.* 181, 331–333.
- (46) Risso, C., Van Dien, S. J., Orloff, A., Lovley, D. R., and Coppi, M. V. (2008) Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. *J. Bacteriol.* 190, 2266–2274.
- (47) Andi, B., West, A. H., and Cook, P. F. (2005) Regulatory mechanism of histidine-tagged homocitrate synthase from *Saccharomyces cerevisiae*. I. Kinetic studies. *J. Biol. Chem.* 280, 31624–31632.
- (48) Kumar, V. P., West, A. H., and Cook, P. F. (2011) Kinetic and chemical mechanisms of homocitrate synthase from *Thermus thermophilus*. *J. Biol. Chem.* 286, 29428–29439.
- (49) de Kraker, J. W., and Gershenzon, J. (2011) From amino acid to glucosinolate biosynthesis: protein sequence changes in the evolution of methylthioalkylmalate synthase in *Arabidopsis*. *Plant Cell* 23, 38–53.
- (50) Li, F., Hagemeyer, C. H., Seedorf, H., Gottschalk, G., and Thauer, R. K. (2007) Re-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. *J. Bacteriol.* 189, 4299–4304.
- (51) Hidaka, T., Shimotohno, K. W., Morishita, T., and Seto, H. (1999) Studies on the biosynthesis of bialaphos (SF-1293). 18. 2-phosphinomethylmalic acid synthase: a descendant of (R)-citrate synthase? *J. Antibiot.* 52, 925–931.
- (52) Larson, E. M., and Idnurm, A. (2010) Two origins for the gene encoding alpha-isopropylmalate synthase in fungi. *PLoS One* 5, e11605.
- (53) Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* 5, e1000605.
- (54) Aoshima, M., and Igarashi, Y. (2006) A novel oxalosuccinate-forming enzyme involved in the reductive carboxylation of 2-oxoglutarate in *Hydrogenobacter thermophilus* TK-6. *Mol. Microbiol.* 62, 748–759.
- (55) Sakai, A., Fedorov, A. A., Fedorov, E. V., Schnoes, A. M., Glasner, M. E., Brown, S., Rutter, M. E., Bain, K., Chang, S., Gheyi, T., Sauder, J. M., Burley, S. K., Babbitt, P. C., Almo, S. C., and Gerlt, J. A. (2009) Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis,cis-muconate lactonizing enzymes. *Biochemistry* 48, 1445–1453.
- (56) Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat. Chem. Biol.* 3, 486–491.
- (57) Casey, A. K., Schwalm, E. L., Hays, B. N., and Frantom, P. A. (2013) V-type allosteric inhibition is described by a shift in the rate-determining step for alpha-isopropylmalate synthase from *Mycobacterium tuberculosis*. *Biochemistry* 52, 6737–6739.
- (58) Tuinstra, R. L., and Mizioro, H. M. (2003) Investigation of conserved acidic residues in 3-hydroxy-3-methylglutaryl-CoA lyase: implications for human disease and for functional roles in a family of related proteins. *J. Biol. Chem.* 278, 37092–37098.
- (59) Lietzan, A. D., and St Maurice, M. (2013) Insights into the carboxyltransferase reaction of pyruvate carboxylase from the structures of bound product and intermediate analogs. *Biochem. Biophys. Res. Commun.* 441, 377–382.
- (60) Lietzan, A. D., and St Maurice, M. (2013) A substrate-induced biotin binding pocket in the carboxyltransferase domain of pyruvate carboxylase. *J. Biol. Chem.* 288, 19915–19925.
- (61) Carere, J., McKenna, S. E., Kimber, M. S., and Seah, S. Y. (2013) Characterization of an aldolase-dehydrogenase complex from the cholesterol degradation pathway of *Mycobacterium tuberculosis*. *Biochemistry* 52, 3502–3511.
- (62) Fu, Z., Runquist, J. A., Forouhar, F., Hussain, M., Hunt, J. F., Mizioro, H. M., and Kim, J. J. (2006) Crystal structure of human 3-hydroxy-3-methylglutaryl-CoA Lyase: insights into catalysis and the molecular basis for hydroxymethylglutaric aciduria. *J. Biol. Chem.* 281, 7526–7532.

- (63) Koon, N., Squire, C. J., and Baker, E. N. (2004) Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* *101*, 8295–8300.
- (64) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* *14*, 1188–1190.