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.
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, 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. 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. 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. 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. 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
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 Km value for the nonvaried substrate (AcCoA or KIV). Initial velocities were determined using DTP to detect the formation of CoA at 324 nm (ε = 19.8 mM⁻¹ 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 5–10 times the Km value for the nonvaried substrate (AcCoA or KIV). Activity could not be determined above the detection limit of the assay. Determined using 1 mM AcCoA.

**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, and 5–10 times the Km value for the nonvaried substrate (AcCoA or KIV), and varied concentrations of the inhibitor, respectively, and β is the fractional velocity at saturating concentrations of the inhibitor, I.¹⁹

\[
\frac{v}{[E]_t} = \frac{k_{cat} \times [S]}{K_m + [S]} \tag{1}
\]

**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 Km value for the nonvaried substrate (AcCoA or KIV), and varied concentrations of the inhibitor, respectively, and β is the fractional velocity at saturating concentrations of the inhibitor, I.¹⁹

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

\[
\frac{v_i}{[E]_t} = \frac{k_{cat}K_i \times \beta k_{cat}[1]}{[1 + K_i]} \tag{3}
\]

**Table 1. Kinetic Parameters Determined for MtIPMS Variants**

<table>
<thead>
<tr>
<th></th>
<th>k_{cat} (s⁻¹)</th>
<th>K_{AcCoA} (μM)</th>
<th>k_{cat}/K_{AcCoA} (μM⁻¹ s⁻¹)</th>
<th>k_{KIV} (μM)</th>
<th>k_{cat}/K_{KIV} (μM⁻¹ s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>3.4 ± 0.3</td>
<td>42 ± 12</td>
<td>0.08 ± 0.02</td>
<td>6.4 ± 1.4</td>
<td>0.50 ± 0.13</td>
</tr>
<tr>
<td>L79A</td>
<td>1.0 ± 0.1</td>
<td>30 ± 5</td>
<td>0.03 ± 0.01</td>
<td>120 ± 30</td>
<td>0.008 ± 0.002</td>
</tr>
<tr>
<td>R80A/K</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>D81A/H</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>N83A</td>
<td>0.31 ± 0.04</td>
<td>990 ± 290</td>
<td>0.0003 ± 0.0001</td>
<td>51 ± 14*</td>
<td>0.006 ± 0.002</td>
</tr>
<tr>
<td>N83E</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>Q84A</td>
<td>0.10 ± 0.02</td>
<td>170 ± 60</td>
<td>0.0006 ± 0.0002</td>
<td>19 ± 6</td>
<td>0.005 ± 0.002</td>
</tr>
</tbody>
</table>

²Detected using 4,4-dithiopyridine (DTP) to detect the formation of CoA at 324 nm (ε = 19.8 mM⁻¹ 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 Km value for the nonvaried substrate (AcCoA or KIV). Activity could not be determined above the detection limit of the assay. Determined using 1 mM AcCoA.

**Sequence Similarity Network Data Set Sources and Curation.** Four structures representing the known reaction diversity in the DRE-TIM metallolyase superfamiliy (pdb ids: 20 Invm, 1rqb, 3hq1, and 2q7f) were used as a starting point for identifying new homologous sequences through a series of BLAST searches against UniRef100 (Uniprot Release 2012_02). All hits from an initial BLAST search with an E-value less than 1 × 10⁻¹⁰ were kept. This new data set of 13230 sequences was filtered using HMMER 3.0 beta such that any sequence that did not score against the Pfam 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, 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. Briefly, the final superfamily set of proteins (8817 unique sequences) were imported into a MySQL database and defined as nodes. All-by-all Blast2seq 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 Cytoscape-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 v4.5.6 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: 3u6w, LiCMS (3blf), and SpHCS (2zyf)) from the Claisen condensation-like (CC-like) subgroup were aligned using the Needleman–Wunsch algorithm as implemented in the Matchmaker program in the Cytoscape program.
Chimera. 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.

**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 ($k_{cat} = 4.6 \pm 0.3$ min$^{-1}$ and $K_{AcCoA} = 283 \pm 54$ μM). The L79A variant displayed a 60-fold decrease in $k_{cat}/K_{i}$ when compared to that of the wild-type enzyme. The N83A and Q84A variants displayed a 150–250-fold decrease in $k_{cat}/K_{AcCoA}$, while the $k_{cat}/K_{i}$ 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. Inhibition parameters for two-step, slow-onset inhibitors are described by two terms, $K_{i0}$ 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. 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**

<table>
<thead>
<tr>
<th>enzyme</th>
<th>$K_{i}$ (μM)</th>
<th>$K_{i}^{*}$ (μM)</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>12 ± 3</td>
<td>2.3 ± 0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>L79A</td>
<td>19 ± 3</td>
<td>4.0 ± 0.5</td>
<td>0</td>
</tr>
<tr>
<td>N83A$^b$</td>
<td>8.3 ± 2.2</td>
<td>3.2 ± 0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Q84A</td>
<td>13 ± 4</td>
<td>4.8 ± 0.5</td>
<td>0.04</td>
</tr>
</tbody>
</table>

$^a$Determined using a standard reaction mixture of 50 mM potassium phosphate buffer (pH 7.5), 12 mM MgCl$_2$, 50 μM DTP, 5–10 times the $K_{m}$ values of AcCoA and KIV, and varied concentrations of leucine (0–120 μM). $^b$Determined using 750 μM AcCoA.

**Figure 2.** Representative sequence similarity network for the DRE-TIM metalloylase 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 $1 \times 10^{-20}$ (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 uncoulored 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. 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, homocitrinate synthase, and citramalate synthase families, along with unknowns), aldolase-like, and carboxylase-like.

Sequence Similarity Network for the DRE-TIM Metalloylase Superfamily. The overall sequence similarity network for representative members of the DRE-TIM metalloylase 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 at least one protein in the node has been annotated in Swiss-Prot 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. The vast majority of proteins have not been biochemically characterized and are thus left as proteins of unknown function ("unknowns", uncoulored 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 mean pairwise alignment score for all-by-all comparisons of these 4298 sequences that is better than an E-value threshold cutoff of $1 \times 10^{-80}$ (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.

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 $1 \times 10^{-80}$ (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.

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 IPMS 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 databases 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,” 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 metallolysate superfamily and assist in the interpretation of our mechanistic analysis of a conserved active-site helix in MtIPMS.

**Analysis of the DRE-TIM Metallolysate 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 mettalloylase 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-phosphomethylmalic acid synthase and α-ketoglutarate carboxylase. 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.” 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, 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 mettalloylase catalytic domain and a C-terminal LeuA dimer regulatory domain: IPMS1/CM2/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. The allosteric mechanism of V-type allosteric inhibition in MtIPMS is due to the perturbation of the hydrolysis step in the reaction. 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 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, aldolase-like, and lyase-like superfamilies. 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 Information).

Figure 4. Hidden-Markov Model logos for the active site helix of DRE-TIM mettalloylase superfamily members. Logos were generated from alignments of sequences in each boxed cluster in Figure 2 using Weblogo. Arrows indicate the main residues discussed in the article.
inactive enzyme; however, a 105 decrease would be at the limit.

MtIPMS, substitution with alanine or histidine results in an
pyruvate and maintains an interaction with the helix arginine
residue plays unique roles in the mechanism of each enzyme. In
aspartate (the
of the spectrophotometric assay for CoA detection. While the
substitution with glutamate decreases activity 10-fold.58 In
fold decrease in speci

c 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 103 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 metallolysase 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,
stimulation of glutamate 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
kcat 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.39,60 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 all of the DRE-TIM metallolysase
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 Km 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

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, brown) and HCS from S. pombe (3irt, blue). (C) Superposition of MtIPMS (1sr9, brown) and NmIPMS (3rmj, 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 carboxylase group is strictly conserved in the position
adjacent to arginine in the DRE-TIM metallolysase superfamily
(D81 in MtIPMS). In HMG-CoA lyase, substitution of the
aspartate with alanine, glycine, or histidine results in a 103 to
104-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 103 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 metallolysase 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,
stimulation of glutamate 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
kcat 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.39,60 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-
ketoadolase 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
kcat value by 50-fold relative to that of the wild-type and results
in the loss of a pKc 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 × 102 M−1 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 metallolysase
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 Km 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-citramate 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.\(^2,10\) 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 asparagine.\(^10\) 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/N83 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 IPMS\(^29\) 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 decrease in the k\(_\text{cat}\) value for AcCoA and a 10-fold decrease in the K\(_m\) value. Surprisingly, when the N83E substitution is made, no activity is detectable. A comparison of structures from MtIPMS and IPMS from Neisseria meningitidis (NnIPMS, 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.

\[\text{dx.doi.org/10.1021/bi500246a} \]
Biochemistry

Present Address
8(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
acetate, acetyl-Coenzyme A; CC-like, Clan classification-like; CMS, citramalate synthase; DTP, 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; TFA, triethanolamine

REFERENCES
(22) Koon, N., Squire, C. J., and Baker, E. N. PDB ID: 3HQI, Probing the Active Site of M. tuberculosis LeuA.
Bioinformatics 22, 1658–1659.


Bioinformatics 28, 2845–2846.


