Human Molecular Genetics, 2014, Vol. 23, No. 9 **2313–2323**
doi:10.1093/hmg/ddt624
Advance Access published on December 11, 2013

CLYBL is a polymorphic human enzyme with malate synthase and β-methylmalate synthase activity

Laura Strittmatter[1,2,4], Yang Li[1,5], Nathan J. Nakatsuka[1,3], Sarah E. Calvo[1,2,4], Zenon Grabarek[1,4], and Vamsi K. Mootha[1,2,4,*]

[1]Department of Molecular Biology, Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, MA 02114, USA [2]Department of Systems Biology, [3]Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, USA [4]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02141, USA [5]Department of Statistics, Harvard University, Cambridge, MA 02138, USA

Received September 30, 2013; Revised November 15, 2013; Accepted December 9, 2013

CLYBL is a human mitochondrial enzyme of unknown function that is found in multiple eukaryotic taxa and conserved to bacteria. The protein is expressed in the mitochondria of all mammalian organs, with highest expression in brown fat and kidney. Approximately 5% of all humans harbor a premature stop polymorphism in CLYBL that has been associated with reduced levels of circulating vitamin B12. Using comparative genomics, we now show that CLYBL is strongly co-expressed with and co-evolved specifically with other components of the mitochondrial B12 pathway. We confirm that the premature stop polymorphism in CLYBL leads to a loss of protein expression. To elucidate the molecular function of CLYBL, we used comparative operon analysis, structural modeling and enzyme kinetics. We report that CLYBL encodes a malate/β-methylmalate synthase, converting glyoxylate and acetyl-CoA to malate, or glyoxylate and propionyl-CoA to β-methylmalate. Malate synthases are best known for their established role in the glyoxylate shunt of plants and lower organisms and are traditionally described as not occurring in humans. The broader role of a malate/β-methylmalate synthase in human physiology and its mechanistic link to vitamin B12 metabolism remain unknown.

INTRODUCTION

Vitamin B12, or cobalamin, is an essential cofactor required for the activity of two human enzymes: mitochondrial methylmalonyl-CoA mutase (MUT) and cytoplasmic methionine synthase (MTR). Humans obtain cobalamin from diet or supplements, and once absorbed, the vitamin is further processed to its active forms—adenosylcobalamin for MUT and methylcobalamin for MTR. Vitamin B12 metabolism and utilization have been extensively studied, and with the exception of the mitochondrial transporter, the molecular identities of all proteins necessary for B12 maturation have been identified (1,2).

Two recent human genetic association studies have begun to define the genetic loci that control B12 levels (3,4). Of the 12 non-intergenic loci collectively identified by these two studies, 11 lie near genes with established or purported roles in vitamin B12 biology, including B12 absorption (*MS4A3, FUT2, FUT6, TCN1, TCN2, CUBN, CD320)*, B12 maturation (*ABCD4*, *MMAA*, *MMACHC)* and B12-dependent catalysis (*MUT*). The 12th locus corresponds to a stop polymorphism within CLYBL, an uncharacterized protein. The B12-associated polymorphism (rs41281112) changes Arg259 to a stop codon, predicted to produce a truncated CLYBL protein. Notably, this polymorphism had the largest effect on B12 levels of all the SNPs reported: Chinese men homozygous for the premature stop had 3-fold reduced concentrations of circulating B12, with heterozygotes exhibiting an intermediate phenotype (3).

Although CLYBL appears to be a metabolic enzyme, its activity remains entirely unknown. CLYBL is annotated as 'citrate lyase beta like' because it shows sequence similarity to citE, a component of an enzymatic complex (citE/citD/citF) found in bacteria that cleaves citrate to oxaloacetate and acetyl-CoA (ACoA) (5,6). All three subunits are necessary for complete lyase activity and often co-occur in the same operon (6,7). Humans, as well as other eukaryotes and certain bacteria and archea, possess homologs of citE but not of citD or citF, hinting at an alternate function for this gene (7). Moreover, citrate cleavage in humans is mediated by the cytosolic, ATP-dependent ATP

*To whom correspondence should be addressed at: Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street 6th Floor, Boston, MA 02114, USA. Tel: +1 6176439710; Email: vamsi@hms.harvard.edu

citrate lyase (8). Although the crystal structures of citE homologs in certain organisms lacking citD and citF have been elucidated, the enzymatic function remains unknown.

CLYBL is among a growing number of human genes that harbor polymorphic loss-of-function (LOF) mutations in seemingly healthy individuals. Recent exome sequencing studies indicate that a typical human genome will harbor ~150 LOF variants (9,10). The physiology of some LOF-tolerant genes is well studied, including mitochondrial *ALDH2*, which is associated with 'flushing' in response to alcohol (11,12), and *G6PD*, deficiency of which confers protection against malaria (13). Many other LOF-tolerant genes, like *CLYBL*, remain poorly understood, but if investigated, could yield valuable insights into gene–environment interactions, potential sites of natural selection and novel drug targets.

Here, we explore the molecular function of CLYBL. Using comparative genomics, structural homology modeling and enzymology, we narrowed the list of potential enzymatic activities for CLYBL. We report that CLYBL exhibits malate synthase and β-methylmalate synthase activities, which have never before been reported for a human protein.

RESULTS

CLYBL domain organization, subcellular localization and tissue expression

Human CLYBL is 340 amino acids in length and includes a canonical amino-terminus mitochondrial targeting sequence (MTS, Fig. 1A). It is ubiquitously expressed in mitochondria from all mammalian tissues (14). Adjacent to its MTS, CLYBL contains one Pfam domain, the HpcH/HpaI aldolase/citrate lyase family domain (15), which is the basis for the gene's name (citrate lyase beta like).

We examined RNA profiles from 91 tissues and cell types, primarily from C57BL/6 adult male mice, measured with Affymetrix MOE430_2 microarrays (16), as well as protein expression profiles across 28 mouse tissues from wild-type C57BL/6 adult mice (17). Focusing on the 18 tissues common to both datasets, we observed that Clybl expression was highest in brown fat and kidney (Fig. 1B).

Impact of the rs41281112 polymorphism on CLYBL expression

The loss-of-function polymorphism (rs41281112) changes Arg259 to a premature stop. Out of the 1092 individuals included in the 1000 Genomes Project Phase 1 data, two individuals—one European and one East Asian—were homozygous null for CLYBL (rs41281112), and 49 individuals were heterozygous. The allele frequency varies among populations—5.8% in Europeans, 7.7% in East Asians, 2.8% in Admixed Americans and 0% in Africans (9). Not included in Phase 1 but part of the larger 1000 Genomes Project, we identified a trio with a segregating rs41281112 polymorphism (father homozygous for the predicted LOF T|T genotype, mother homozygous for the wild-type C|C genotype and heterozygous daughter) (Fig. 1C). At both the mRNA (Fig. 1D) and protein (Fig. 1E) levels, the father exhibited the lowest expression of CLYBL, while the heterozygous daughter had an intermediate phenotype. These results demonstrate that the predicted LOF polymorphism does indeed lead to loss of the protein product.

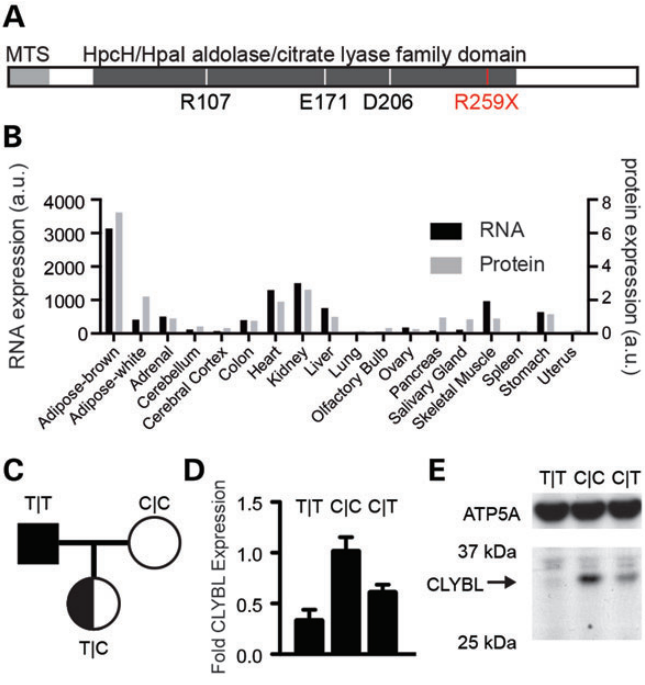

Figure 1. CLYBL domain organization and expression. (**A**) CLYBL contains a predicted 22 amino acid MTS and an 'HpcH/HpaI aldolase/citrate lyase family' protein domain. Predicted magnesium-(E171, D206) and substrate-(R107, E171) binding sites are shown in white. The naturally occurring R259X stop polymorphism (rs41281112) is denoted in red. (**B**) Expression levels of Clybl RNA and protein across 18 mouse tissues in common between the RNA and protein datasets used for co-expression analysis. (**C**) A trio from the 1000 Genomes Project harboring a segregating CLYBL LOF allele, from whose lymphoblasts CLYBL (**D**) normalized RNA based on qPCR and (**E**) protein expression based on immunoblotting is shown.

Co-evolution and co-expression of CLYBL with mitochondrial B12 metabolism

To better understand the function of CLYBL, we considered genes that are co-evolved and co-expressed with CLYBL. First, we performed a phylogenetic profiling analysis of CLYBL, using a new tree-based method (Y. Li *et. al.*, manuscript in review). We utilized a previously published eukaryotic tree of life (10), and used this new algorithm to infer the ancestral states for CLYBL (Fig. 2A). We then scored all human proteins for the likelihood of having evolved under the inferred model of CLYBL evolution compared with a random model. This genome-wide analysis spotlighted nine proteins with a positive log-likelihood ratio (LLR), including all mitochondrial vitamin B12 processing and utilization components (MMAA, MMAB, MUT).

We next sought to identify mouse genes co-expressed with CLYBL at the RNA and protein levels. We calculated Pearson's correlation values for co-expression of CLYBL with all other genes, and found that at both the RNA and protein levels, only a small number of genes were highly correlated with CLYBL, including *Mut*, *Mmaa* and *Mmab* (Fig. 2B,C). We found that *Mmaa* and *Mmab* were highly correlated (Pearson's correlation coefficients >0.9 for RNA and protein) with *Mut* not far behind (Pearson's

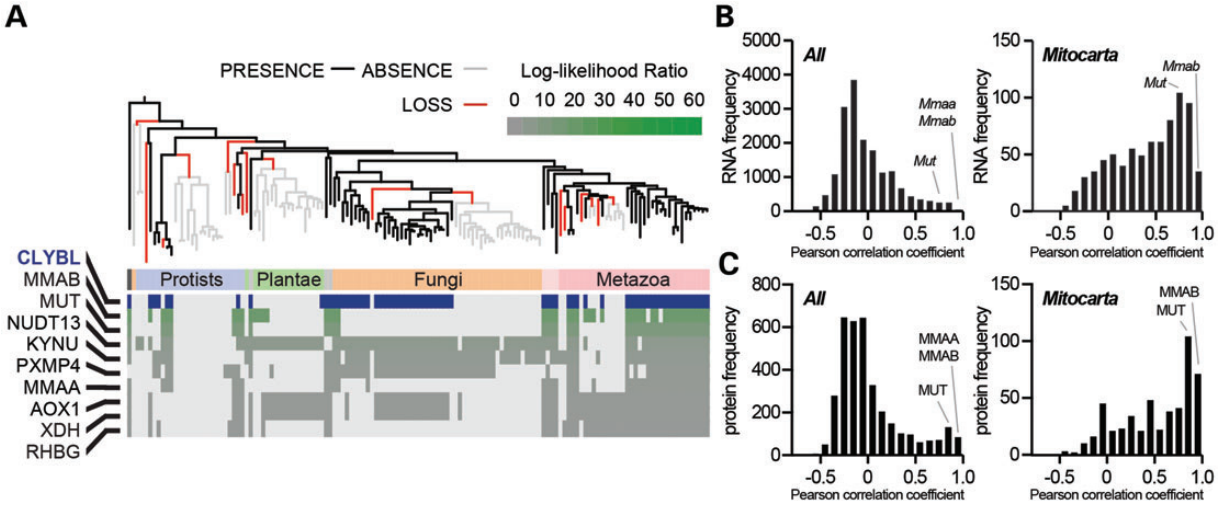

Figure 2. CLYBL co-evolution and co-expression. (**A**) Phylogenetic profile for CLYBL showing genes predicted to share an evolutionary history with a LLR of >0. Blue indicates a CLYBL homolog. (**B**) Distribution of RNA co-expression of CLYBL with all genes (left) and MitoCarta genes (right), based on a public gene expression atlas and focused on the 18 mouse tissues shown in Figure 1B (16). (**C**) Distribution of protein co-expression of CLYBL with all genes (left) and MitoCarta genes (right), based on publicly available protein expression data for 28 mouse tissues (17).

correlation coefficient >0.7 for RNA and >0.8 for protein). Among these highly correlated gene products, we also found the alpha and beta subunits of propionyl-CoA carboxylase (PCCA, PCCB), the alpha and beta subunits of methylcrotonyl-CoA carboxylase (MCCC1, MCCC2) and methylmalonyl-CoA epimerase (MCEE). We sought to determine whether the observed correlation represented a simple consequence of the fact that mitochondrial proteins tend to be co-expressed with each other (18). Even among genes encoding mitochondrial proteins, we observed that the mitochondrial vitamin B12 processing proteins are among the most closely correlated to CLYBL (Fig. 2B,C).

These genome-wide co-evolution and co-expression analyses (Fig. 2) strongly support a functional link between CLYBL and mitochondrial B12 metabolism, in further support of the human genetic association studies (3,4).

Operon analysis of bacterial homologs of CLYBL

CLYBL is annotated as 'citrate lyase beta-like' due to sequence similarity to bacterial citE, which specifically performs the citryl-acyl carrier protein cleavage step of the citrate lyase reaction. CLYBL is unlikely to have the same function because bacterial citE requires two additional proteins, citD and citF, to perform the citrate lyase reaction, and these proteins are not found in humans (7).

We sought to determine whether bacterial CLYBL homologs always co-occur with citD and citF. Of 942 diverse bacterial species represented within the Cluster of Orthologous Groups (COGs) database (19), we found that 121 contain CLYBL homologs as well as citD/citF homologs, while 281 species contain CLYBL homologs but lack citD/F homologs (Fig. 3Ab).

To predict eukaryotic CLYBL's function, we focused on CLYBL homolog-containing operons within the 281 bacterial species lacking citD/F homologs. These 587 CLYBL-containing operons comprised a total of 3947 different genes, corresponding to 576 unique COGs. Interestingly, no single operon structure was overrepresented. However, the top 10 COGs that co-occurred

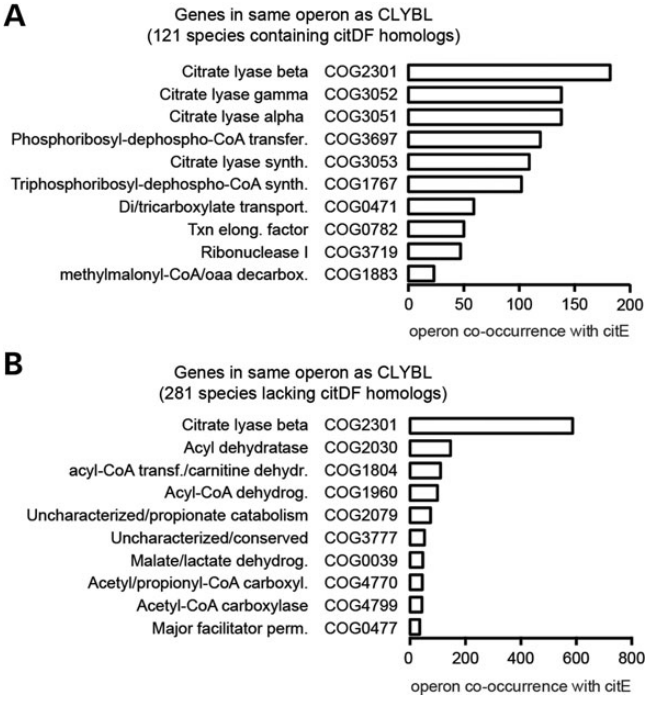

Figure 3. Comparative operon analysis for CLYBL. (**A**) Top 10 clusters of orthologous genes (COGs) co-occurring in CLYBL-containing operons in 121 species containing CLYBL (citE) and citD/F homologs. (**B**) Top 10 COGs co-occurring in CLYBL-containing operons in 281 species containing a CLYBL homolog but lacking citD/F homologs.

with CLYBL included 3 involved in acyl CoA reactions, 2 with ACoA metabolism, 2 with propionate/propionyl-CoA (PCoA) catabolism and 1 with malate/lactate metabolism (Fig. 3B).

In a few instances, bacterial CLYBL homologs found in operons lacking citD/F have been characterized at the structural and biochemical levels. The structures of RipC from *Yersinia pestis KIM 10 (biovar Mediaevalis)* (20) and CitE from

Mycobacterium tuberculosis H37Rv (7) have been elucidated, but the enzymatic activities have not been tested. Mcl1 and Mcl2 are homologs of CLYBL in *Rhodobacter sphaeroides strain 2.4.1* (21) and operate together to produce malate synthase activity. With the exception of Mcl1, which appears in an operon on its own, the rest of these genes appear in operons with genes in COGs similar or identical to those shown in Figure 3B. In particular, all the three operons contained enzymes predicted to be involved in acyl-CoA metabolism.

Enzymatic activity of CLYBL

Based on the above comparative operon analysis, the reported biochemical activity observed in *R. sphaeroides*, and published crystal structures (7,20,21), we hypothesized that the human CLYBL might perform similar catalytic reactions. Therefore, we designed, bacterially expressed, and purified human CLYBL with an N-terminal His-tag, but without the predicted MTS (22). We used Ellman's reagent (DTNB) to monitor the production of CoA-SH in the presence of pairs of substrates: (i) an acyl-CoA coupled with (ii) either glyoxylate or a larger alpha-ketoacid. Our assay was based on a previously reported assay for malate synthase activity (23), though using potassium phosphate buffer to avoid glyoxylate complexation (24). We tested CLYBL with a variety of pairs of substrates, including ACoA, PCoA, butyryl-CoA or racemic methylmalonyl-CoA, coupled with glyoxylate, pyruvate, 2-oxobutyrate or alpha-ketoglutarate. CLYBL only showed substantial activity when incubated with glyoxylate and either ACoA or PCoA, reactions predicted to produce malate and β-methylmalate, respectively. CLYBL also showed some activity, though lower, when incubated with ACoA and pyruvate, a substrate pair predicted to produce citramalate. As a positive control, we also tested GlcB, a previously characterized malate synthase from *M. tuberculosis* (25). We expressed and purified GlcB in *Escherichia coli,* in parallel with CLYBL, and found that GlcB was specific for ACoA/glyoxylate and had kinetic values in our assay on par with those reported in the literature (unpublished data), demonstrating the fidelity of our assay.

A CoA-SH-producing reaction between glyoxylate and ACoA appeared most likely to be a malate synthase reaction (Fig. 4A), so we further characterized this activity. We found that our recombinant protein had an apparent K_m of 3.6 mM for glyoxylate and 74 μM for ACoA, similar to the K_ms observed for the condensation of ACoA and glyoxylate by *R. sphaeroides* Mcl1, though Mcl1 does not perform the complete malate synthase reaction, instead producing malyl-CoA, which is cleaved to malate and CoA by Mcl2. CLYBL had a maximum specific activity of 180 pmol/μg protein/minute, and a turnover per subunit of 0.12 s^{-1}, slow compared with Mcl1 at 8.6 s^{-1}. Other divalent cations can substitute for magnesium in this reaction, but CLYBL has essentially no activity in the presence of EDTA or Ca^{2+} (Table 1; Supplementary Material, Table S1).

We confirmed that CLYBL indeed produced malate by monitoring malate production over time via liquid chromatography/mass spectrometry (LC/MS) (Fig. 4B). We confirmed the identity of malate as the reaction product by matching its retention time, exact mass and MS/MS spectra to a standard (Fig. 4B, inset). After confirming that our method could linearly detect

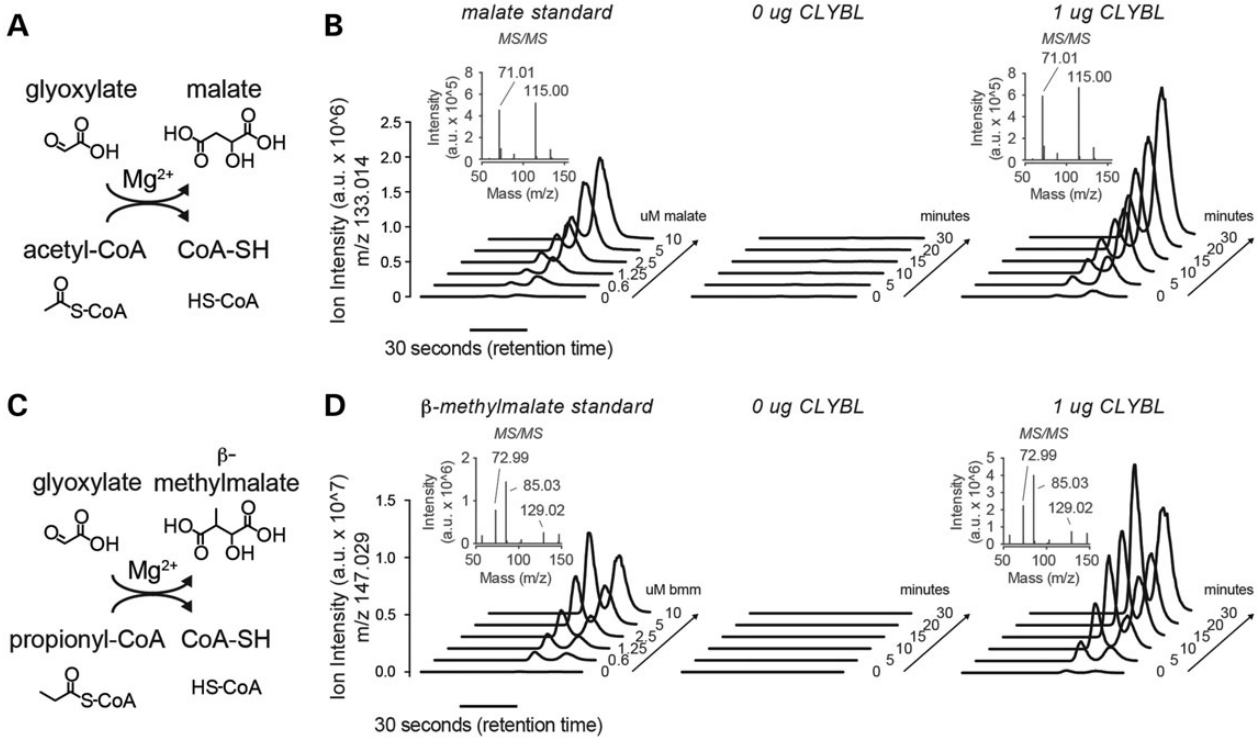

Figure 4. CLYBL has malate synthase and β-methylmalate synthase activity. (**A**) Malate synthase reaction. (**B**) LC-MS extracted ion chromatograms showing detection of malate standards and the accumulation of malate over 30 min when the reaction mixture containing magnesium, glyoxylate and ACoA is incubated with 0 or 1 μg of CLYBL. Insets show representative MS/MS spectra for malate. (**C**) β-methylmalate synthase reaction. (**D**) Same as (b), but for β-methylmalate production with PCoA instead of ACoA. Insets show representative MS/MS spectra for β-methylmalate.

Table 1. Kinetic parameters for malate synthase and β-methylmalate synthase activities of bacterially expressed and purified human CLYBL with an N-terminal His-tag and linker, using DTNB-linked assay. Values are reported with standard error.

Enzymatic activity	Malate synthase	β-Methylmalate synthase
CoA substrate	ACoA	PCoA
Specific activity (pmol/μg/min)		
CoA + glyoxylate	180 ± 12	138 ± 13
Apparent K_m		
Glyoxylate (mM)	3.6 ± 0.7	1.2 ± 0.4
CoA (μM)	74 ± 11	23 ± 5
Turnover per subunit, k_{cat}, s^{-1}		
CoA + glyoxylate	0.12	0.09
Effect of 10 mM alternate divalent cations on specific activity (% Mg^{2+})		
EDTA	2%	0%
Mg^{2+}	100%	100%
Mn^{2+}	198%	0%
Ca^{2+}	0%	0%
Co^{2+}	228%	13%
Ni^{2+}	109%	17%

increasing concentrations of malate standard doped into the reaction mix, we monitored the progress of the reaction over 30 min with or without the addition of 1 μg CLYBL. We observed malate accumulation only in complete reaction mixtures containing CLYBL.

We also characterized the CLYBL-catalyzed reaction coupling glyoxylate and PCoA to produce CoA-SH, which was mostly likely producing β-methylmalate (Fig. 4C). We observed an apparent K_m of 1.2 mM for glyoxylate and 23 μM for PCoA, similar to those observed for condensation of PCoA and glyoxylate to β-methylmalyl-CoA by *R. sphaeroides* Mcl1. This *Rhodobacter* protein is unable to perform the complete β-methylmalate synthase reaction, and Mcl2 has not been shown to cleave β-methylmalyl-CoA (21). CLYBL had a maximum specific activity of 138 pmol/μg protein/minute, and a turnover per subunit of 0.09 s^{-1}, slow compared with Mcl1's condensation of the same substrates (Supplementary Material, Table S1). We confirmed the identity of β-methylmalate as our reaction product using the LC/MS retention time, exact mass and MS/MS spectra to compare the product with a standard (Fig. 4D, inset). Co^{2+} and Ni^{2+} enabled 13 and 17% of magnesium-dependent activity, respectively, while CLYBL had no β-methylmalate synthase activity in the presence of EDTA, Mn^{2+} or Ca^{2+} (Table 1; Supplementary Material, Table S1).

Structural homology modeling of CLYBL

We sought to determine whether the modeled structure of CLYBL is compatible with the malate synthase and β-methylmalate synthase activities. We modeled the structure of CLYBL using as templates the 1.65 Å resolution structure of citrate lyase beta subunit from *M. tuberculosis* (PDB 1u5h) (7) and the 1.95 Å resolution structure of *Haloferax volcanii* malate synthase complexed with ACoA, pyruvate and Mg^{2+} (PDB 3oyz) (26). Using a multistep procedure involving structure-based sequence alignment (Materials and Methods), a model with good geometry was obtained for residues 42–285 of CLYBL despite only 17% amino acid sequence identity with the template. The C-terminal 55 amino acid segment could not be modeled because the corresponding structure in the template (1u5 h) was missing.

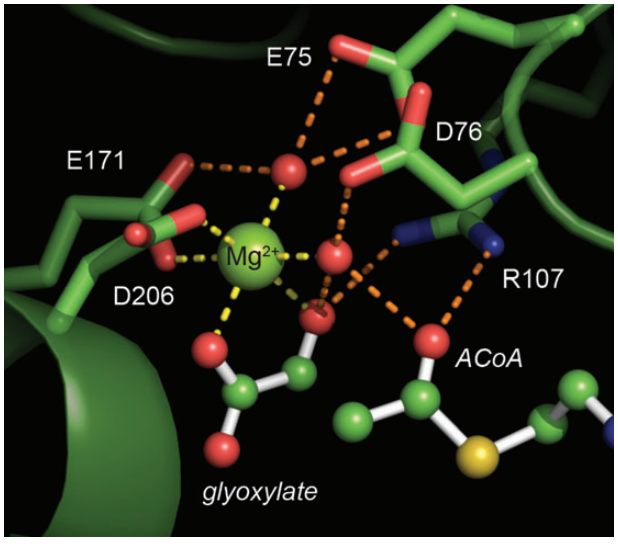

Figure 5. Structural modeling of CLYBL. The model shows that substrates ACoA and glyoxylate fit in the active site, with room for PCoA instead of ACoA. Yellow dashed lines indicate direct bonds, while orange dashed lines depict hydrogen bonds. Structure was visualized in PyMOL (30).

Importantly, due to the presence of ACoA and a substrate analog (pyruvate) in the template structure, the role of several conserved residues critical for the catalytic activity of CLYBL can be evaluated based on the model. The protein has a $(\beta/\alpha)_8$ TIM barrel fold common to a large group of enzymes, including malate synthases (27–29). In this fold, the substrate-binding sites are made of variable loops that connect the C-termini of parallel β-strands with adjacent α-helices. The key element of the catalytic center is Mg^{2+} coordinated in octahedral geometry by two oxygen atoms of the substrate (glyoxylate), the side chain oxygen atoms of Glu171 and Asp206, and two H_2O molecules. Glu75 and Asp76 further stabilize the complex via hydrogen bonds to Mg^{2+}-bound H_2O molecules. An essential element of the catalytic center is Arg107, whose side chain guanidinium group makes strong hydrogen bonds simultaneously with the alpha oxygen of glyoxylate and the acetyl carbonyl oxygen atom of ACoA, thus positioning these substrates precisely for catalysis (Fig. 5). It is also apparent from the model that substitution of a propionyl group for the acetyl group of ACoA can be readily accommodated without affecting any of the key interactions, explaining CLYBL's ability to act as a β-methylmalate synthase as well as a malate synthase.

All the important elements of the catalytic center of CLYBL discussed above are also present in other malate synthase structures and are consistent with CLYBL having a malate synthase catalytic activity. However, in all the three types of malate synthases (A, G and H), there is a C-terminal extension, which apparently plays an important role in catalysis. Specifically, the side chain of Asp633 in *M. tuberculosis* malate synthase isoform G was proposed to function as a catalytic base that abstracts a proton from ACoA (25). Substitution of Asn for Asp631, the corresponding residue in *E. coli* malate synthase, renders the protein inactive (24). There is a conserved Asp residue at the structurally equivalent position also in A and H isoforms of malate synthase (26,31). The absence of an equivalent structural element in the CLYBL monomer may contribute to the low observed catalytic activity.

With the structural model in hand, we sought to further confirm that CLYBL had malate synthase and β-methylmalate synthase activities by mutating specific residues predicted to be critical for various aspects of catalysis. Based on homology modeling and evolutionary conservation, we mutated residues Glu75 and Glu171, predicted to be important for coordinating a magnesium-binding water molecule and directly binding magnesium, respectively, to Gly and found that these mutants had no malate synthase or β-methylmalate synthase activity in our DTNB-linked assay (data not shown). We also mutated Asp250 to Gly to evaluate the effect of a charge mutation at a residue that is conserved but distal to the active site, and found that this mutation reduced but did not eliminate these activities (data not shown). We also attempted to express and purify the Arg259Stop LOF variant (Fig. 1A,C–E) identified from the 1000 Genomes data and associated with low circulating B12 levels, but the expressed protein was insoluble, suggesting that the human mutant may be improperly folded if it is expressed at all, consistent with what we observed in cell lines from individuals with the premature stop polymorphism (Fig. 1E).

DISCUSSION

Our computational, enzymatic and structural homology analyses demonstrate that purified recombinant CLYBL possesses malate synthase and β-methylmalate synthase activities, which, to our knowledge, is the first time that such activities have been ascribed to a human protein. Malate synthase is traditionally described as operating in alternative carbon assimilation pathways such as the glyoxylate shunt. At present, the link between CLYBL and circulating B12 levels remains unclear, but it is notable that the newly identified substrates and products of CLYBL, as well as gene products strongly co-expressed with and co-evolving with CLYBL (Fig. 2), lie in close proximity to the mitochondrial B12 pathway (Fig. 6).

It is important to consider the kinetic parameters of CLYBL in the context of previously reported malate synthases. Previously studied malate synthase enzymes vary widely in their affinity for glyoxylate, ranging from a K_m of 2 mM in *Ricinus comunis* (32) to 21 μM in *E. coli* (24). The 3.6 mM K_m reported here is at the high end of this spectrum, but it is similar to the 3.1 mM K_m reported for Mcl1 in *R. sphaeroides* (Table 1; Supplementary Material, Table S1) (21). Mcl1 is not a complete malate synthase, but rather condenses glyoxylate and ACoA to malyl-CoA before malyl-CoA is cleaved by the paralogous Mcl2. It is also notable that Mcl1 can catalyze condensation of both ACoA and PCoA with glyoxylate with similar kinetics to those observed for CLYBL (21). It is important to note that in our assays CLYBL is relatively inefficient, with low specific activity, but it could be the case that CLYBL requires an additional protein partner or post-translational modification *in vivo* to enhance its activity. When we screened possible substrate pairs, we noted that CLYBL had low activity with pyruvate and ACoA too, suggesting that it has low citramalate synthase activity. It is possible that CLYBL has additional enzymatic activities with greater catalytic efficiency that have not been tested here.

We note in our structural modeling that CLYBL is missing a large C-terminal domain present in malate synthase enzymes. Critically, CLYBL is missing a conserved aspartate (Asp633 in *M. tuberculosis* malate synthase G) predicted to work as a catalytic base that abstracts a proton from ACoA to initiate the malate synthase reaction (25,26). An Asp residue in a structurally equivalent position is absolutely conserved in all G form (~730 amino acids) malate synthases as well as in their shorter analogs: the A form (~530 amino acids), e.g. the *E. coli* and *B. anthracis* malate synthases (31) and the H form (~430 amino acids), e.g. *H. volcanii* (26). Substitution of this critical Asp with Asn in the *E. coli* G malate synthase was shown to render the protein inactive (24). Thus, a question arises which, if any, part of CLYBL could substitute for the missing catalytic Asp. A comparison of our model of CLYBL with the template *H. volcanii* malate synthase indicates that the C-terminal 50 amino acid segment of the polypeptide chain, which we were unable to model due to the low sequence similarity, is too short to fold into a stable structure required for positioning the critical Asp close to the bound substrate within the same monomer. We hypothesize that this critical contribution must be provided by intermolecular interactions within an oligomeric structure. Strong support for this prediction comes from the structure of the putative citrate lyase beta subunit from *B. xenovorans* (PDB 3r4i). The 339 amino acid polypeptide chain of this protein folds into a TIM barrel structure like CLYBL, but also contains a small flexible C-terminal domain, which extends into the catalytic domain of the neighboring

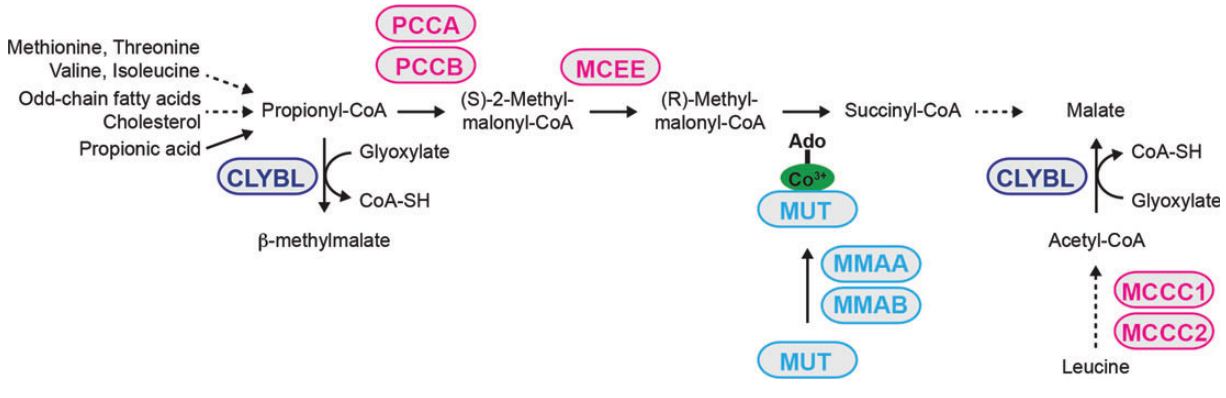

Figure 6. Newly identified enzymatic activities of CLYBL in the context of mitochondrial B12 metabolism. Genes in light blue encode enzymes in the mitochondrial B12 pathway. Genes in magenta encode transcripts and proteins that are highly co-expressed with CLYBL. Solid lines indicate one-step reactions and dashed lines represent multiple enzymatic steps.

monomer in the biological trimer (Supplementary Material, Fig. S1A). Asp298 in that domain is exactly in the same position as the catalytic Asp388 of *H. volcanii* malate synthase (Supplementary Material, Fig. S1B). Based on the amino acid sequence alignment (using MUSCLE) of CLYBL and *B. xenovorans* citrate lyase, we tentatively identify Asp320 as the plausible catalytic residue in CLYBL (Supplementary Material, Fig. S2) (33). Interestingly, other citrate lyase beta proteins are trimers based on their X-ray structures, suggesting that these proteins might also have malate synthase activity. Further studies are needed to verify this prediction.

Malate synthases are traditionally described as part of alternative carbon assimilation pathways. Malate synthase and isocitrate lyase together comprise a glyoxylate shunt that bypasses two CO_2-releasing steps of the TCA cycle (34) to promote biomass accumulation from lipids without the loss of carbon. Malate synthase activity can occur in organisms lacking isocitrate lyase (21) and is thought to function in an acetate assimilation ethylmalonyl-CoA pathway (21,35,36). In principle, CLYBL could be participating in a similar carbon assimilation pathway. Malate synthase and isocitrate lyase activities have been reported in brown adipose tissue from black bears (37), livers of several mammals (38–40) and toad bladder (41), though other groups have disputed these results (42,43). In none of these reports was a molecular basis for these activities identified. In light of these previous studies, it is notable that CLYBL expression is particularly high in brown fat (Fig. 1B).

The β-methylmalate synthase activity of CLYBL may be particularly important in the context of certain inborn errors of metabolism, which are characterized by elevated PCoA (44). In such disorders, it is known that PCoA can compete with ACoA to produce side products, such as methylcitrate (45). It is conceivable that β-methylmalate is similarly produced by CLYBL. Whether β-methylmalate is a toxic metabolite, or perhaps an inert metabolite that is excreted, remains to be determined.

Our work adds CLYBL to the growing number of human enzymes known to utilize glyoxylate, and underscores a broader role for this metabolite in human physiology. Glyoxylate is best studied in the context of calcium oxalate stones that are due to mutations in *AGXT* or *GRHPR* (46). Previous studies have suggested that the intracellular concentration of glyoxylate ranges between 0 and 100 μM (47,48), though these estimates may be low relative to concentrations within subcellular compartments (49). Glyoxylate in humans may originate from the diet, via glycolate in vegetables and hydroxyproline in meat (50). It may also come from the enzymatic activities of glyoxylate reductase, lactate dehydrogenase (51) or alanine:glyoxylate aminotransferase (52).

Our work provides unequivocal support for a physiological link between CLYBL and B12 metabolism. We have shown that CLYBL is strongly co-expressed with and co-evolving with all the three known members of the mitochondrial B12 pathway: MUT, MMAA and MMAB (Fig. 2). These analyses complement previous association studies that found the CLYBL LOF polymorphism was associated with low B12 levels in human plasma (3,4). In the current work, we have shown that this polymorphism leads to a loss of the CLYBL protein (Fig. 1E). It is notable that the substrates and products of malate/β-methylmalate synthase lie within close metabolic proximity to the mitochondrial B12-dependent enzyme MUT (Fig. 6). Why polymorphisms in MUT and CLYBL alter circulating B12 levels is not known. One possibility is that B12 levels are regulated by sensing the substrates/products of the reactions in Figure 6, analogous to other human metabolic pathways (53).

Although our work has been focused on human CLYBL, it has potential implications for the development of novel therapeutics for *M. tuberculosis*. While the crystal structure of the *M. tuberculosis* CLYBL homolog, CitE (accession Rv2498c), has been solved, its enzymatic activity has remained enigmatic (7). We hypothesize that this protein has the same activity as CLYBL, and it would be straightforward to screen for inhibitors of the protein using the assays we have developed. *Mycobacterium tuberculosis* malate synthase GlcB is already being pursued as an anti-tuberculosis drug target (54), but if the CLYBL homolog is able to substitute for this enzyme, it may be crucial to target both enzymes. Any small-molecule inhibitors discovered would be particularly attractive candidates for therapeutic development; even if they also inhibited the human CLYBL, we would not expect overt toxicity, given that humans can evidently tolerate loss of CLYBL.

MATERIALS AND METHODS

Cellular studies

The following B-lymphocyte cell lines were ordered from the Coriell Cell Repository: HG01970 (T|T for rs41281112), HG01971 (C|C), HG01972 (C|T). Cells were cultured in Sigma RPMI 1640 with 2 mM glutamine, 15% Gibco 16000 fetal bovine serum at 37°C. For mRNA expression analysis, mRNA was isolated from cells using a Qiagen RNeasy kit, and qRT–PCR was performed for CLYBL and ACTB using Taqman assays (Applied Biosystems, assay ID Hs00370518_m1 and 4352935E, respectively), according to the manufacturer's protocol. The qRT-PCR data are expressed as $\Delta\Delta C_T$ values with respect to ACTB and HG01971. For western blot analysis, whole cell lysates were prepared and resolved by SDS–PAGE followed by transfer to a PVDF membrane using the Trans-Blot Turbo Transfer System. Membranes were probed using a primary antibody for CLYBL (Abnova CLYBL MaxPab mouse polyclonal antibody H00171425-B01P, 1:500) or ATP5A (Abcam Mouse Monoclonal, no. AB14748, 1:50 000), followed by a secondary antibody (Polyclonal Sheep anti-IgG mouse antibody, no. NA931V, 1:10 000). The signal was detected using a Pierce ECL western blotting substrate.

Phylogenetic profiling

To study the evolutionary history of CLYBL, we utilized a phylogenetic profiling algorithm (Li *et al.*, manuscript in review). The algorithm accepts as input a binary phylogenetic tree, a binary homology matrix and a query gene set *G*. It then partitions *G* into disjoint clusters in which genes share an inferred model of evolutionary history. For each cluster, the algorithm then identifies genes not in *G* but that share the same evolutionary history. A published phylogenetic tree (10) with 138 eukaryotic species and a single prokaryote outgroup, a phylogenetic matrix of 20 834 human genes, and *CLYBL* as a single gene in *G* were used as input. The human-centric phylogenetic matrix $X = \{X_{g,s}\}_{20834 \times 138}$ was constructed to contain 1 if the human reference gene *g* shared sequence similarity to

any protein in species *s* (BLASTP, Expect $<$ 1e $-$ 3) and 0 otherwise. The algorithm estimated the evolutionary model of gain and loss events for *CLYBL*, and scored all other genes for the likelihood of having arisen under the same inferred model of evolution compared with the background null model, using an LLR. Genes with an LLR score exceeding threshold 0 were concluded to share the same history with *CLYBL* and included in the final output shown in Figure 2.

Co-expression analysis

We analyzed publicly available RNA and protein expression data generated by Lattin *et al*. (16) and Geiger *et al*. (17), respectively. In brief, Lattin *et al*. quantified the RNA expression of over 17 000 genes across 91 C57BL/6 mouse tissues or cell types using Affymetrix MOE430_2 microarrays. Geiger *et al*. quantified the expression of over 6900 proteins across 28 C57BL/6 mouse tissues using LC-MS/MS with mouse tissues labeled via stable isotope labeling with amino acids in cell culture (SILAC) as an internal standard. Pearson's correlations were calculated comparing *Clybl* RNA expression with those of all genes across the 18 mouse tissues (kidney was the average of two sections) in common with the protein dataset and comparing Clybl protein expression with those of all genes across the 28 mouse tissues in the protein dataset. Missing values in the protein expression dataset were ignored in the calculations, and if greater than half (14) of the values for a protein were missing, the protein was removed from the dataset. The mouse MitoCarta (14) was used to attain all genes encoding proteins with strong support of mitochondrial localization.

Bacterial operon analysis

Genomic data for 942 bacterial species were downloaded from eggNOG version 3.0 (19). Consecutive genes on the same strand were annotated to reside in the same operon. Cross-species homologs were determined using pre-computed COG assignments (eggNOG version 3.0), including annotation of homologs to CLYBL/citE (COG2301), citD (COG3052), and citF (COG3051). We identified the species that contained COG2301 but lacked COG3051 and COG3052. For these 281 species, we computed how many times each COG was present within an operon that also contained COG2301. We applied the same analysis to the 121 species that contained all three COGs: COG2301, COG3052 and COG3051.

Protein expression and purification

Starting with the full-length consensus sequence for CLYBL (CCDS32002.1), we removed the TargetP-predicted 22 amino acid targeting sequence and added a methionine to facilitate future cloning and expression of a C-terminally tagged protein (22). The sequence was codon-optimized for bacterial expression, synthesized with 5′ *Sal*1 and 3′ *Xho*1 restriction sites and cloned into pUC57-Kan by Genewiz. The construct was digested with *BamH*1 and *EcoR*1 (NEB) and ligated into the pET-30a(+) vector, in frame with the N terminal 6$\times$His tag, using NEB T4 DNA ligase. After sequence verification, the protein was expressed with Invitrogen One Shot BL21 Star (DE3) chemically competent *E. coli* cells grown at 37°C shaking at 225 rpm. Cultures (OD_{600} 0.4–0.7) were induced with 1 mM IPTG and grown overnight at 15°C before harvest.

The purification protocol was a modified version of previous protocols for recombinant malate synthase purification (24). We resuspended pelleted bacteria in 50 mM HEPES pH 7.9, 300 mM NaCl, 10% glycerol, 10 mM $MgCl_2$, supplemented day of use with 3 mM 2-mercaptoethanol and Roche cOmplete mini EDTA protease inhibitor tablets. Cells were lysed with a sonicator (1 s on, 1 s off, 30% intensity $\times$ 15 $\times$ 3). Lysate was cleared by spinning in a JA20 centrifuge at 19 500 rpm for 30 min at 4°C. Clear lysate was incubated with pre-washed GE Ni-Sepharose 6 fast flow resin for 1 h at 4°C, then washed with wash buffer (50 mM Hepes, 300 mM NaCl, 20 mM imidazole, pH 7.8, 10 mM $MgCl_2$). Protein was eluted with elution buffer (50 mM HEPES pH 7.8, 300 mM NaCl, 500 mM imidazole, 10 mM $MgCl_2$, 3 mM betamercaptoethanol) and dialyzed at 4°C using Thermo Pleated Snakeskin Dialysis Tubing (2 $\times$ 1.5 h) into 25 mM potassium phosphate pH 8.0 for enzymatic assays. Protein was dialyzed in 20 mM Tris pH 7.8 for ion screening. Protein concentration was determined using the Pierce BCA protein assay.

A similar method was used for gene synthesis and expression of *M. tuberculosis* malate synthase GlcB. QuikChange Lightning mutagenesis was used to modify specified residues in the CLYBL bacterial expression construct.

Enzymatic activity assays

Chemicals and reagents were purchased from Sigma, with the exception of β-methylmalate from Otava Ltd. The production of CoA-SH from ACoA or PCoA in response to glyoxylate was monitored spectrophotometrically at 412 nm (37°C) in the presence of 100 μM DTNB (23). The assay used 1–10 μg of recombinant protein per well in 96 well format. Protein was resuspended in an assay mix composed of 50 mM potassium phosphate monobasic, 10 mM $MgCl_2$, and 2 mM EDTA, pH 8.0, plus 500 μM final concentration of the CoA of interest. The reaction was mixed and monitored for 2 min before the addition of 10 mM glyoxylate (prepared fresh each time). Activity was determined by subtracting the slope pre-glyoxylate addition from the slope post-glyoxylate addition, converting the change in absorbance at 412 nm to moles of SH product (and thereby to moles of acyl-CoA) using the extinction coefficient of 14 150 $M^{-1}cm^{-1}$ for DTNB and the 0.585 cm path length specified for our 96 well plates.

K_m values were calculated by varying the concentrations of acyl-CoA and glyoxylate; divalent ion selectivity was determined by replacing 10 mM $MgCl_2$ with 10 mM $MnCl_2$, $CaCl_2$, $CoCl_2$, or $NiCl_2$, or 20 mM EDTA, in a reaction mix composed of 10 mM glyoxylate, 300 μM acyl-CoA, and 20 mM Tris buffer (pH 7.8) instead of potassium phosphate to avoid precipitation.

Mass spectrometry

To confirm the identity of the non-CoA-SH product in each reaction, we used LC/MS-MS. Recombinant protein (1 μg per 50 μl final reaction) was incubated in 20 mM Tris pH 7.8, 10 mM $MgCl_2$, 2 mM EDTA and 300 μM of either ACoA or PCoA. Reactions were initiated with 10 mM glyoxylate, mixed and incubated at 37°C, then quenched 1:1 (v/v) with HPLC-grade methanol at

the indicated time point. Malate or β-methylmalate standard curves were run in a pre-quenched reaction mix. Samples were kept in a sealed 384-well plate at 4°C until analysis (within 24 h).

Samples were injected into a Phenomenex Luna 3 μm NH2 100 Å 150 × 2.00 mm column eluted at 0.2 ml/min. Samples were eluted over a 55 s ramp from 10 to 100% A (20 mM ammonium hydroxide, 20 mM ammonium acetate in water), followed by 2 min at 100% A, a 1 min ramp back to 10% A, and 2 min of re-equilibration at 10% A. Mobile Phase B was 75:25 v/v acetonitrile/methanol with 20 mM ammonium hydroxide (55). Data (negative ion mode) were collected with a Q Exactive mass spectrometer (Thermo Scientific) using the following parameters: sheath gas flow rate 45, auxiliary gas flow rate 15, sweep gas flow rate 0, spray voltage 3.20 kV, S-lens RF level 75.0, capillary temperature 320°C and heater temperature 200°C.

Extracted ion chromatograms were plotted for the exact masses of malate and β-methylmalate, 133.014 and 147.029, respectively (width 0.003), as identified by single ion monitoring (SIM). MS/MS fragmentation was also performed for malate (133.01 → 115.00, 71.01) and β-methylmalate (147.03 → 129.02, 85.03, 72.99) using the following conditions: isolation width 4.0 m/z, fixed first mass 50.0 m/z, normalized collision energy 45.0%.

Structural modeling

Modeling of human CLYBL 3D structure (Q8N0X4-1) was performed using the Swiss Model server (http://swissmodel.expasy.org/) (56–58). The model was built in several steps. First, we used the Basic Local Alignment Search Toll (BLAST) available via NCBI server (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify the proteins in PDB whose amino acid sequences are most closely related to the fragments 23–340 of CLYBL (the N-terminal MTS deleted). The highest scoring proteins (28–34% sequence identity) were those annotated as citrate lyase beta subunits. The highest scoring structure of a malate synthase was that from *H. volcanii* (PDB 1oyx). However, with only 17% sequence identity this template proved unsuitable for direct modeling due to uncertainty in sequence alignment. Thus, the structure of citrate lyase from *M. tuberculosis* (PDB 1u5 h) (7) that had the highest resolution (1.65 Å) was selected as a template. This yielded an initial model of CLYBL (Model1), which featured a characteristic $(\beta/\alpha)_8$ TIM barrel fold common to a large group of enzymes including malate synthases. Then we used the PDBe Fold server at the European Bioinformatics Institute http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver (59) to structurally align CLYBL Model1 with several citrate lyase betas and malate synthases, which revealed strict conservation of residues considered critical for catalysis. With the amino acid sequence alignment from PDBe Fold we have modeled CLYBL using as a template the 1.95 Å resolution structure of *H. volcanii* malate synthase ternary complex with ACoA and pyruvate (PDB 3oyz) (26). The obtained model (Model 2) had a good geometry despite only 17% sequence identity with the template. Finally, coordinates of ACoA, glyoxylate (in place of pyruvate) and Mg^{2+} ions with two coordinating H_2O molecules from 3yoz were combined with Model 2 using COOT (60,61). Only minor adjustments of rotamers for a few side chains were required. In the final model of CLYBL, all the amino acids important for catalysis were found in the correct positions for the interaction with substrates, including the octahedral coordination of Mg^{2+} essential for catalysis. All URLs were last accessed December 18, 2013. Structures were visualized in PyMOL (30).

NOTE ADDED IN PROOF

Subsequent to the submission of this work, Zarzycki and Kerfeld (62) reported the X-ray structures of two malyl-CoA lyases (MCL), bacterial proteins homologous to human CLYBL. Zarzycki and Kerfeld reported that the active form of MCL is a dimer of trimers, in which a flexible C-terminal domain complements the active site of an adjacent molecule in the trimer, providing a critical aspartate for catalysis. This residue corresponds to Asp320 of CLYBL, which we have predicted to serve this same function (see Discussion).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Erika Kovács-Bogdán for assistance with protein expression and purification; Kerry Pierce, Noam Prywes, Virendar Kaushik, Nadine Elowe and Brian Hubbard for assistance with mass spectrometry and Ralph Mazitschek, Clary Clish, Ashley Wolf, Mohit Jain and Nikhil Madhusudhan for helpful discussions. Vamsi Mootha is an Investigator of the Howard Hughes Medical Institute.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by an NSF Graduate Research Fellowship (L.S.), the National Institute of General Medical Sciences (T32GM007753 to N.N.) and the National Institutes of Health (1R01GM097136 to V.K.M.).

REFERENCES

1. Banerjee, R., Gherasim, C. and Padovani, D. (2009) The tinker, tailor, soldier in intracellular B12 trafficking. *Curr. Opin. Chem. Biol.*, **13**, 484–491.
2. Gherasim, C., Lofgren, M. and Banerjee, R. (2013) Navigating the B12 road: assimilation, delivery and disorders of cobalamin. *J. Biol. Chem.*, **288**, 13186–13193.
3. Lin, X., Lu, D., Gao, Y., Tao, S., Yang, X., Feng, J., Tan, A., Zhang, H., Hu, Y., Qin, X. *et al.* (2012) Genome-wide association study identifies novel loci associated with serum level of vitamin B12 in Chinese men. *Hum. Mol. Genet.*, **21**, 2610–2617.
4. Grarup, N., Sulem, P., Sandholt, C.H., Thorleifsson, G., Ahluwalia, T.S., Steinthorsdottir, V., Bjarnason, H., Gudbjartsson, D.F., Magnusson, O.T. and Sparsø, T. (2013) Genetic architecture of vitamin B12 and folate levels uncovered applying deeply sequenced large datasets. *PLoS Genet.*, **9**, e1003530.
5. Subramanian, S. and Sivaraman, C. (1984) Bacterial citrate lyase. *J. Bioscience.*, **6**, 379–401.
6. Bott, M. and Dimroth, P. (1994) Klebsiella pneumoniae genes for citrate lyase and citrate lyase ligase: localization, sequencing, and expression. *Mol. Microbiol.*, **14**, 347–356.
7. Goulding, C., Bowers, P., Segelke, B., Lekin, T., Kim, C., Terwilliger, T. and Eisenberg, D. (2007) The structure and computational analysis of *Mycobacterium tuberculosis* protein CitE suggest a novel enzymatic function. *J. Mol. Biol.*, **365**, 275–283.

8. Elshourbagy, N.A., Near, J.C., Kmetz, P.J., Wells, T.N., Groot, P.H., Saxty, B.A., Hughes, S.A., Franklin, M. and Gloger, I.S. (1992) Cloning and expression of a human ATP-citrate lyase cDNA. *Eur. J. Biochem.*, **204**, 491–499.
9. Consortium, T.1.G. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **490**, 56–65.
10. Bick, A.G., Calvo, S.E. and Mootha, V.K. (2012) Evolutionary diversity of the mitochondrial calcium uniporter. *Science*, **336**, 886–886.
11. Yoshida, A., Huang, I.Y. and Ikawa, M. (1984) Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *P. Natl Acad. Sci. USA*, **81**, 258–261.
12. Hempel, J., Kaiser, R. and Jörnvall, H. (1984) Human liver mitochondrial aldehyde dehydrogenase: a C-terminal segment positions and defines the structure corresponding to the one reported to differ in the Oriental enzyme variant. *FEBS Lett.*, **173**, 367–373.
13. Beutler, E. (1994) G6PD deficiency. *Blood*, **84**, 3613–3636.
14. Pagliarini, D., Calvo, S., Chang, B., Sheth, S., Vafai, S., Ong, S. and Walford, G. (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell*, **134**, 112–123.
15. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J. *et al.* (2011) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
16. Lattin, J.E., Schroder, K., Su, A.I., Walker, J.R., Zhang, J., Wiltshire, T., Saijo, K., Glass, C.K., Hume, D.A., Kellie, S. *et al.* (2008) Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res.*, **4**, 5.
17. Geiger, T., Velic, A., Macek, B., Lundberg, E., Kampf, C., Nagaraj, N., Uhlen, M., Cox, J. and Mann, M. (2013) Initial quantitative proteomic map of 28 Mouse Tissues Using the SILAC Mouse. *Mol. Cell. Proteomics*, **12**, 1709–1722.
18. Mootha, V.K., Bunkenborg, J., Olsen, J.V., Hjerrild, M., Wisniewski, J.R., Stahl, E., Bolouri, M.S., Ray, H.N., Sihag, S. and Kamal, M. (2003) Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell*, **115**, 629–640.
19. Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T. *et al.* (2011) eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.*, **40**, D284–D289.
20. Torres, R., Chim, N., Sankaran, B., Pujol, C., Bliska, J.B. and Goulding, C.W. (2012) Structural insights into RipC, a putative citrate lyase β subunit from a *Yersinia pestis* virulence operon. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **68**, 2–7.
21. Erb, T.J., Frerichs-Revermann, L., Fuchs, G. and Alber, B.E. (2010) The apparent malate synthase activity of *Rhodobacter sphaeroides* is due to two paralogous enzymes, (3S)-malyl-coenzyme A (CoA)/β-methylmalyl-CoA lyase and (3S)-malyl-CoA thioesterase. *J. Bacteriol.*, **192**, 1249–1258.
22. Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.*, **2**, 953–971.
23. Cox, R.B. and Quayle, J.R. (1976) Synthesis and hydrolysis of malyl-coenzyme A by Pseudomonas AM1: an apparent malate synthase activity. *J. Gen. Microbiol.*, **95**, 121–133.
24. Anstrom, D.M., Kallio, K. and Remington, S.J. (2003) Structure of the *Escherichia coli* malate synthase G:pyruvate:acetyl-coenzyme A abortive ternary complex at 1.95 Å resolution. *Protein Sci.*, **12**, 1822–1832.
25. Quartararo, C.E. and Blanchard, J.S. (2011) Kinetic and chemical mechanism of malate synthase from *Mycobacterium tuberculosis*. *Biochemistry*, **50**, 6879–6887.
26. Bracken, C.D., Neighbor, A.M., Lamlenn, K.K., Thomas, G.C., Schubert, H.L., Whitby, F.G. and Howard, B.R. (2011) Crystal structures of a halophilic archaeal malate synthase from *Haloferax volcanii* and comparisons with isoforms A and G. *BMC Struct. Biol.*, **11**, 23.
27. Wierenga, R.K. (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.*, **492**, 193–198.
28. Nagano, N., Orengo, C.A. and Thornton, J.M. (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.*, **321**, 741–765.
29. Wise, E.L. and Rayment, I. (2004) Understanding the importance of protein structure to nature's routes for divergent evolution in TIM barrel enzymes. *Acc. Chem. Res.*, **37**, 149–158.
30. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
31. Lohman, J.R., Olson, A.C. and Remington, S.J. (2008) Atomic resolution structures of Escherichia coli and Bacillus anthracis malate synthase A: comparison with isoform G and implications for structure-based drug discovery. *Protein Sci.*, **17**, 1935–1945.
32. Bowden, L.L. and Lord, J.M.J. (1978) Purification and comparative properties of microsomal and glyoxysomal malate synthase from castor bean endosperm. *Plant Physiol.*, **61**, 259–265.
33. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
34. Kornberg, H.L. and Krebs, H.A. (1957) Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature*, **179**, 988–991.
35. Meister, M., Saum, S., Alber, B.E. and Fuchs, G. (2005) L-Malyl-coenzyme A/β-methylmalyl-coenzyme A lyase is involved in acetate assimilation of the isocitrate lyase-negative *Bacterium Rhodobacter capsulatus*. *J. Bacteriol.*, **187**, 1415–1425.
36. Šmejkalová, H., Erb, T.J. and Fuchs, G. (2010) Methanol assimilation in methylobacterium extorquens AM1: demonstration of all enzymes and their regulation. *PLoS One*, **5**, e13001.
37. Davis, W.L., Goodman, D.B.P., Crawford, L.A., Cooper, O.J. and Matthews, J.L. (1990) Hibernation activates glyoxylate cycle and gluconeogenesis in black bear brown adipose tissue. *Biochim. Biophys. Acta*, **1051**, 276–278.
38. Davis, W.L. and Goodman, D.B.P. (1992) Evidence for the glyoxylate cycle in human liver. Anat. Rec. *(*Hoboken*)*, **234**, 461–468.
39. Davis, W.L., Jones, R.G., Farmer, G.R., Dickerson, T., Cortinas, E., Cooper, O., Crawford, L. and Goodman, D.B.P. (1990) Identification of glyoxylate cycle enzymes in chick liver—the effect of vitamin D3: cytochemistry and biochemistry. Anat. Rec. *(*Hoboken*)*, **227**, 271–284.
40. Davis, W.L., Matthews, J.L. and Goodman, D. (1989) Glyoxylate cycle in the rat liver: effect of vitamin D3 treatment. *FASEB J.*, **3**, 1651–1655.
41. Goodman, D.B., Davis, W.L. and Jones, R.G. (1980) Glyoxylate cycle in toad urinary bladder: possible stimulation by aldosterone. *P. Natl Acad. Sci. USA*, **77**, 1521–1525.
42. Jones, J.D., Burnett, P. and Zollman, P. (1999) The glyoxylate cycle: does it function in the dormant or active bear? *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.*, **124**, 177–179.
43. Holmes, R.P. (1993) The absence of glyoxylate cycle enzymes in rodent and embryonic chick liver. *Biochim. Biophys. Acta*, **1158**, 47–51.
44. Ogier de Baulny, H. and Saudubray, J.M. (2002) Branched-chain organic acidurias. *Semin. Neonatol.*, **7**, 65–74.
45. Ando, T., Rasmussen, K., Wright, J.M. and Nyhan, W.L. (1972) Isolation and identification of methylcitrate, a major metabolic product of propionate in patients with propionic acidemia. *J. Biol. Chem.*, **247**, 2200–2204.
46. Danpure, C.J. (2011) Primary hyperoxaluria. In Scriver, C.R., Beaudet, A.L., Vallenet, D., Sly, W.S., Childs, B., Kinzler, K.W. and Vogelstein, B. (eds), *The Metabolic and Molecular Bases of Inherited Disease*. McGraw Hill, Vol. 15, pp. 1–141.
47. Baker, P.R.S., Cramer, S.D., Kennedy, M., Assimos, D.G. and Holmes, R.P. (2004) Glycolate and glyoxylate metabolism in HepG2 cells. *Am. J. Physiol. Cell Physiol.*, **287**, C1359–C1365.
48. Liang, C.C. (1962) Studies on experimental thiamine deficiency. Trends of keto acid formation and detection of glyoxylic acid. *Biochem. J.*, **82**, 429.
49. Hamilton, G.A., Buckthal, D.J., Mortensen, R.M. and Zerby, K.W. (1979) Reactions of cysteamine and other amine metabolites with glyoxylate and oxygen catalyzed by mammalian D-amino acid oxidase. *P. Natl Acad. Sci. USA*, **76**, 2625–2629.
50. Takayama, T., Fujita, K., Suzuki, K., Sakaguchi, M., Fujie, M., Nagai, E., Watanabe, S., Ichiyama, A. and Ogawa, Y. (2003) Control of oxalate formation from L-hydroxyproline in liver mitochondria. *J. Am. Soc. Nephrol.*, **14**, 939–946.
51. Mdluli, K., Booth, M.P.S., Brady, R.L. and Rumsby, G. (2005) A preliminary account of the properties of recombinant human glyoxylate reductase (GRHPR), LDHA and LDHB with glyoxylate, and their potential roles in its metabolism. *Biochim. Biophys. Acta*, **1753**, 209–216.
52. Cellini, B., Montioli, R., Paiardini, A., Lorenzetto, A. and Voltattorni, C.B. (2009) Molecular insight into the synergism between the minor allele of human liver peroxisomal alanine: glyoxylate aminotransferase and the F152I mutation. *J. Biol. Chem.*, **284**, 8349–8358.
53. Brown, M.S. and Goldstein, J.L. (1997) The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell*, **89**, 331–340.
54. Krieger, I.V., Freundlich, J.S., Gawandi, V.B., Roberts, J.P., Gawandi, V.B., Sun, Q., Owen, J.L., Fraile, M.T., Huss, S.I., Lavandera, J.-L. *et al.* (2012) Structure-guided discovery of phenyl-diketo acids as potent inhibitors of *M. tuberculosis* malate synthase. Chem. Biol., **19**, 1556–1567.
55. Jiang, D., Zhao, L., Clish, C.B. and Clapham, D.E. (2013) Letm1, the mitochondrial Ca2+/H+ antiporter, is essential for normal glucose

metabolism and alters brain function in Wolf–Hirschhorn syndrome. *Proc. Natl Acad. Sci. USA*, **110**, E2249–E2254.
56. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
57. Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381–3385.
58. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.
59. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.
60. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
61. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
62. Zarzycki, J. and Kerfeld, C.A. (2013) The crystal structures of the trifunctional *Chloroflexus aurantiacus* and bi-functional *Rhodobacter sphaeroides* malyl-CoA lyases and comparison with CitE-like superfamily enzymes and malate synthases. *BMC Struct. Biol.*, **13**, 28.