

Strong epistasis constrains the evolution of strict substrate specificity in apicomplexan lactate dehydrogenases

Jacob D. Wirth¹, Jeffrey I. Boucher², Changhan Xu, Scott Classen³, and Douglas L. Theobald^{4*}

¹presently at Marengo Therapeutics, Cambridge, MA

²presently at Tessera Therapeutics, Somerville, MA

³Lawrence Berkeley National Lab, Berkeley, CA

⁴Department of Biochemistry, Brandeis University, Waltham, MA

Corresponding author

*D.L.T., Department of Biochemistry, Brandeis University, 415 South St., Waltham, MA 02453.

Telephone: 781-736-2303. E-mail: dtheobald@brandeis.edu

Abstract

The homologous enzymes lactate and malate dehydrogenase (L/MDH) are structurally similar but are specific for different substrates. LDH vs MDH specificity is canonically governed by the identity of a single “specificity residue” at position 102. However, LDH function has convergently evolved from a specific MDH at least four times, and the catalytic role of residue 102 is not conserved between different phyla. The apicomplexa are a phylum of obligate, intracellular eukaryotic parasites responsible for wide-spread disease such as *Plasmodium falciparum* (malaria), *Cryptosporidium parvum* (cryptosporidiosis), *Toxoplasma gondii* (toxoplasmosis), and *Eimeria maxima* (eimeriosis). The apicomplexan LDH evolved via a five-residue insertion that produced a novel specificity residue, W107f. The commonly accepted mechanism of LDH specificity involves charge balance and steric occlusion, but our data shows that the general mechanism of apicomplexan LDHs does not use W107f as a steric block. Only *Plasmodium* LDHs evolved substantial steric specificity, making them exceptional among Apicomplexa. Strong protein epistasis constrained this evolution, making it difficult to revert to ancestral phenotypes. Here, we use ancestral sequence reconstruction (ASR), steady-state kinetics, and x-ray crystallography to characterize apicomplexan LDHs which challenge current assumptions about the evolution of L/MDH activity. We demonstrate the unique specificity of *Plasmodium* LDHs and identify the active site residues controlling their substrate recognition. The extraordinarily high specificity of *Plasmodium* LDHs presents difficulties for small-molecule inhibitor development, and successful drugs against *Plasmodium* LDH may not be efficacious against other Apicomplexa LDHs and their diseases.

Introduction

Protein epistasis is a widely occurring phenomenon where the effect of a mutation is dependent on its genetic context, in this case defined as the presence of other residues¹⁻². Epistasis can be positive (permissive) or negative (restrictive) and in some cases cause a mutation's effect to change from beneficial to detrimental (or vice versa)³. Nonspecific epistasis may describe a many-to-many relationship that involves a large amount of a protein's sequence while more specific epistasis could involve the presence of a single permissive residue allowing for the mutation of another². The direct effect of an epistatic relationship is witnessed when mutating the equivalent position in two closely related proteins has significantly different outcomes.

Lactate and malate dehydrogenases (LDHs and MDHs) are homologous metabolic enzymes that share a protein fold⁴ and a common catalytic mechanism (Figure 1)⁵⁻⁸. LDH interconverts pyruvate and lactate in anaerobic glycolysis. In the presence of oxygen, metabolites are instead driven through the citric acid cycle where MDH interconverts oxaloacetate and malate. Their roles in central metabolism make them essential and found throughout all three domains of life⁹. The shared catalytic mechanism and ordered bi-bimolecular kinetic scheme stem from the family's highly conserved active site and similar substrates (Figure 1). The enzymatic reduction of substrate proceeds as follows: 1) NADH binds free enzyme, 2) R171 coordinates the substrate's carboxylic acid group, 3) a loop closes over the active site, binding the substrate, 4) hydride transfer occurs, 5) the active site loop opens, ejecting the reduced product, and 6) the enzyme regenerates by releasing NAD⁺. D168 is required to activate the catalytic H195, which is directly involved in the proton transfer during catalysis. Research has shown that the movement of the active site loop, known as the specificity loop, is the reaction's primary rate-limiting step¹⁰⁻¹¹.

The loop is so-named because it plays an important role in substrate recognition. LDHs and MDHs have strict specificity for their respective substrates, though occasionally substrate analogs (Figure 1c) are turned over with modest activities^{10, 12-14}. Residue 102 resides on the loop and is commonly referred to as the ‘specificity residue’ because it differentiates between the functional groups of different α -ketoacids^{11-12, 15}. MDHs have R102 which can form a salt bridge with the γ -carboxylate group of its native substrates, oxaloacetate and malate. The most common form of LDH, like those in bacteria and metazoa, contain Q102 which lacks arginine’s formal positive charge and allows for stable binding of pyruvate or lactate¹².

The enzyme family has often been a model for functional evolution studies with many descriptions, solved protein structures, and documented functional changes¹⁶⁻¹⁸. No two types of

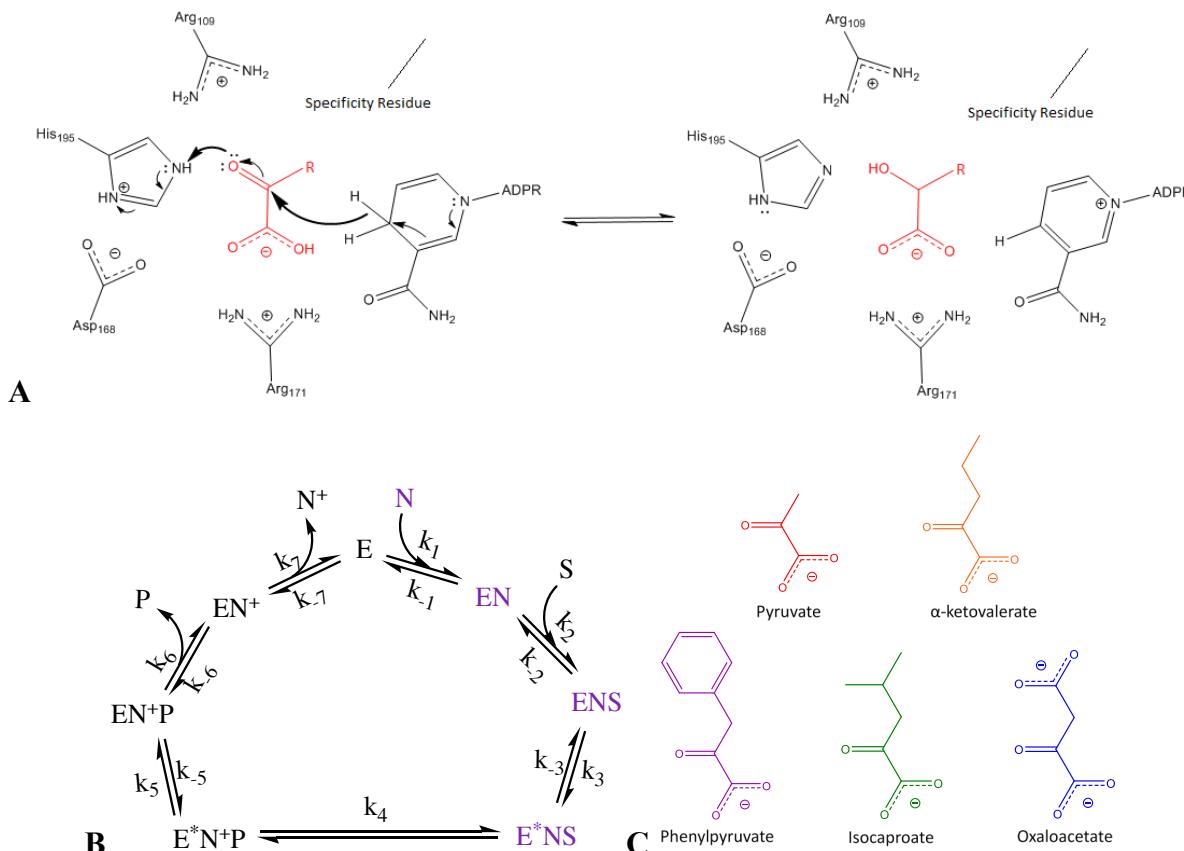


Figure 1 Mechanism of loop closure and catalysis. **(A)** and **(B)** are the molecular and kinetic mechanism of substrate turnover for LDHs and MDHs as described in the text. Specificity residues vary. **(C)** Alternate substrates can be used to study LDH and MDH specificity.

LDH naturally use the same specificity residue(s) for activity, suggesting that each group evolved a different form of substrate discrimination. In some cases it is possible to swap specificities between LDH and MDH^{12, 19}, while in others it is only marginally effective²⁰ or produces completely inactive enzymes²¹. As LDHs evolved from MDHs multiple independent times^{14, 18, 22-23}, these specificity studies support the idea that epistasis has strongly constrained the evolution of LDH activity which in turn could explain why each type of LDH required very different mutations to develop pyruvate specificity.

One independent family of LDHs belongs to the intracellular eukaryotic parasites Apicomplexa, human pathogens that cause many diseases including malaria, cryptosporidiosis, babesiosis, and toxoplasmosis. The organisms responsible for malaria belong to the *Plasmodium* genus with the most fatal form of malaria due to *Plasmodium falciparum*. *P. falciparum* proceeds through a complex life-cycle in two different hosts where *P. falciparum* sporozoites infect humans *via* a mosquito bite. LDH in *Plasmodium* is transcribed at all stages in blood²⁴ and has been measured in the micromolar range²⁵ indicating the enzyme likely performs a critical role. While in host blood, *P. falciparum* respires anaerobically to regenerate NAD⁺ due to the conditions of the human red blood cell²⁶⁻²⁷. It is likely that LDH is the only means of NAD⁺ regeneration during this time, making the enzyme essential for the pathogen's survival.

LDHs in Apicomplexa evolved nearly one billion years ago²⁸ from a horizontal gene transfer of an ancient α -proteobacterial malate dehydrogenase, independently of canonical metazoan and bacterial LDHs^{18, 29-30}. The gene transfer occurred before Apicomplexa started to speciate and so all apicomplexan LDH genes, save one genus, come from one convergent event and have similar characteristics²². For the purpose of this study, LDHs from *Cryptosporidium* will not be addressed as they have evolved separately²². Apicomplexan LDHs did not evolve by

mutating a glutamine at position 102, despite sharing the human LDH catalytic mechanism¹⁰.

Rather, they evolved from an ancestral MDH through a unique five amino acid insertion in the substrate specificity loop that switches substrate specificity from malate/oxaloacetate to lactate/pyruvate²³. The insertion lengthens the loop but otherwise the structures of apicomplexan and metazoan LDHs are highly similar³¹.

A canonical human-like LDH structure with the substrate specificity loop in the closed conformation has the Q102 specificity residue contacting substrate³². In contrast, the *P.falciparum* LDH (LDH_PLFA) crystal structure shows W107f contacting a substrate analog in the closed conformation, which indicates that W107f is the ‘specificity residue’ Apicomplexa evolved instead of something at position 102³³. Point mutations and deletions within the substrate specificity loop have demonstrated that the essential W107f is the only crucial residue when determining enzyme activity and substrate recognition, while mutating K102 has negligible effects^{23, 34}.

As stated above, pyruvate specificity in canonical LDHs or MDHs is decided by position 102 on an active site loop (Q102 or R102, respectively)¹². We can rationalize W107f as a specificity determining residue as it satisfies the charge balance requirements of pyruvate turnover and can sterically occlude oxaloacetate from the closed conformation of the active site (Figure 2)³³. Apicomplexan LDHs lack any activity towards oxaloacetate despite there being an available positive charge at K102, leading many to question why the enzymes do not also function as MDHs^{10, 25, 35-38}. More puzzling, our group was able to make a bifunctional apicomplexan LDH through the mutation K102R, calling in to question exactly how the loop-insertion is used by the enzyme to determine specificity²³.

Noted previously, LDH_PLFA is unable to turnover large substrate analogs¹⁰, a behavior that we show is atypical for Apicomplexa. Instead, LDH_PLFA has exceptionally high specificity when compared to closely related LDHs from other Apicomplexa such as *Babesia bovis* (LDH_BABO), *Toxoplasma gondii* (LDH_TOGO), and *Eimeria maxima* (LDH_EIMA). LDH_PLFA has frequently been a target for small-molecule drug therapies due to its essential role during the parasite's life cycle and active site features distinct from human LDH^{33, 35, 39-45}. Any difficulties developing small-molecule drug therapies meant to target *Plasmodium* LDHs should be unsurprising if the organisms evolved an important but unexplored or uncharacterized mechanism of substrate specificity. Could that specificity be easily relaxed or changed? Could other LDHs easily evolve higher specificity?

Apicomplexa LDHs provide a good model system for investigating several questions in molecular evolution, such as the number of mutations required to change a protein's phenotype, how enzymes evolve strong specificities, and to what degree epistasis affects the course of

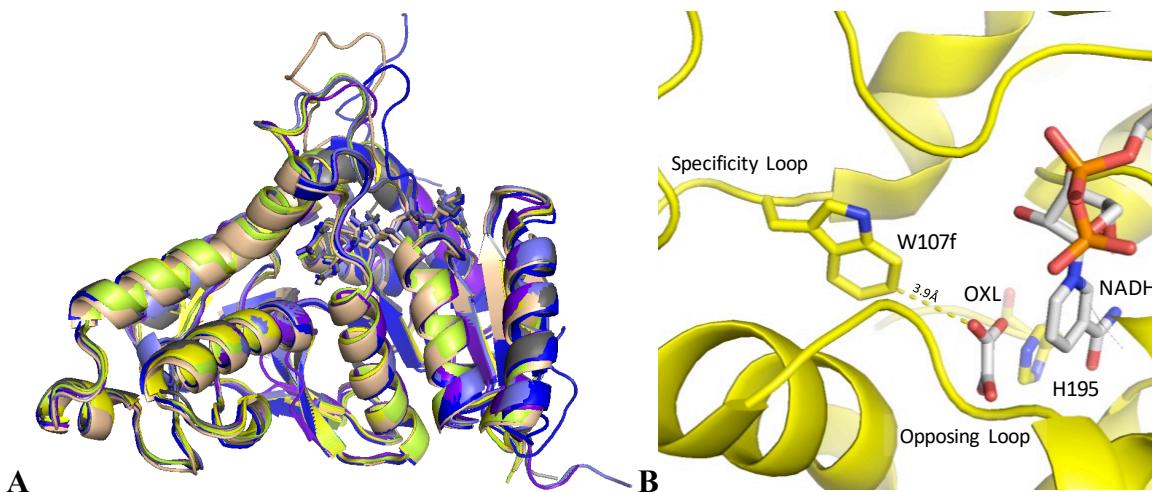


Figure 2 Apicomplexan LDH protein structure. **(A)** Six LDH structures from the Apicomplexa phylum exhibit highly similar global monomer structures: one *Eimeria*, two *Toxoplasma*, and three *Plasmodium* (PDBID: 6CT6, 1PZH, 1SOW, 1T2D, 2A92, and 1OC4, respectively). **(B)** The specificity loop with W107f is labeled, as well as a visible loop seen orientated across the active site. The pyruvate analog oxalate and NADH are shown as white sticks.

functional evolution. Here we use ancestral sequence reconstruction, steady-state enzyme kinetics, and x-ray crystallography to characterize LDHs from several Apicomplexa, including modern proteins, ancestral constructs, and evolutionary intermediates. We identify what attributes are unique in *Plasmodium* LDH active sites and their contribution to substrate specificity. We also demonstrate that the specific order of mutation has a large impact when maintaining an enzyme's activity as it evolves new specificity.

Results

Substrate Specificity in Apicomplexan LDHs

The closely related extant LDH_PLFA and LDH_EIMA were assayed for their steady-state activities towards several substrates. Both enzymes have high activity using the native substrate pyruvate yet LDH_PLFA is significantly more specific. Compared to other LDHs from Apicomplexa, LDH_PLFA has a k_{cat}/K_M higher for pyruvate and much lower when using other substrates (Figure 3, Supp. Table 1). LDH_EIMA greatly prefers pyruvate over oxaloacetate but has a more relaxed specificity overall, making it more like close homologs³⁶. To generalize the behavior of all apicomplexan LDHs, our group reconstructed the sequence of their last common ancestor as described in the methods (Figure 4)²³. LDH_EIMA and the reconstructed ancestor (AncLDH) have qualitatively identical phenotypes: enzyme activity decreases as the substrate R-group loses planarity and increases in size. In general, most apicomplexan LDHs are not as sterically specific towards their substrates as LDH_PLFA.

LDH_PLFA's inability to turnover phenylpyruvate was recorded previously¹⁰ but the mechanism behind such robust specificity was not explored. The large planar R-group is uncharged, allowing for use as a suitable substrate mimic to probe steric vs charge sensitivity in LDHs. Two additional LDH ancestors were reconstructed (AncLDH2 and AncLDH3, Figure 4) to

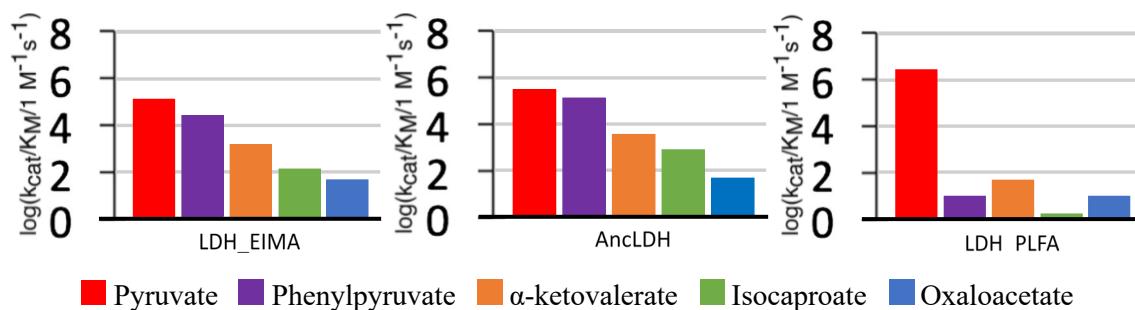


Figure 3 Apicomplexan LDH substrate specificity. Values are $\log(k_{cat}/K_M)$ towards α -ketoacid substrates in units of $M^{-1}s^{-1}$. Each bar is colored by substrate. AncLDH is the reconstructed last common ancestor of all apicomplexan LDHs.

follow the lineage towards *Plasmodium* LDHs and determine when the organisms gained the increased specificity. Along with another extant enzyme, LDH_BABO, the ancestral proteins were assayed for their pyruvate, oxaloacetate, and phenylpyruvate activities (Figure 5, Supp. Table 2). Apicomplexan MDHs are very specific for oxaloacetate, meaning the five-residue insertion is required for pyruvate turnover. It is apparent that the insertion also confers high phenylpyruvate activity which implies loss of specificity for oxaloacetate is due to charge and not size. Only LDH_PLFA and AncLDH3 were unable to use phenylpyruvate as a substrate which makes their steric discrimination a recent development in Apicomplexa.

Novel Active Site Residues in Plasmodium LDH

Comparing apicomplexan LDH crystal structures reveals a potential explanation for *Plasmodium*'s increased steric specificity. A *Toxoplasma* LDH crystal structure was available as an extant comparison and by using that, along with our solving of the LDH_EIMA structure (PDBID: 6CT6, Supp. Table 3), we were able to identify active site features unique to *Plasmodium* LDHs (Figure 6). Within the crystal structures, high steric specificity correlates with the identity of an active site loop opposite the specificity loop, between the α -G2 and α -G3 helices (the 'opposing loop', Figure 2b and 6b)^{23, 33, 37-38, 46-47}. The opposing loop in most

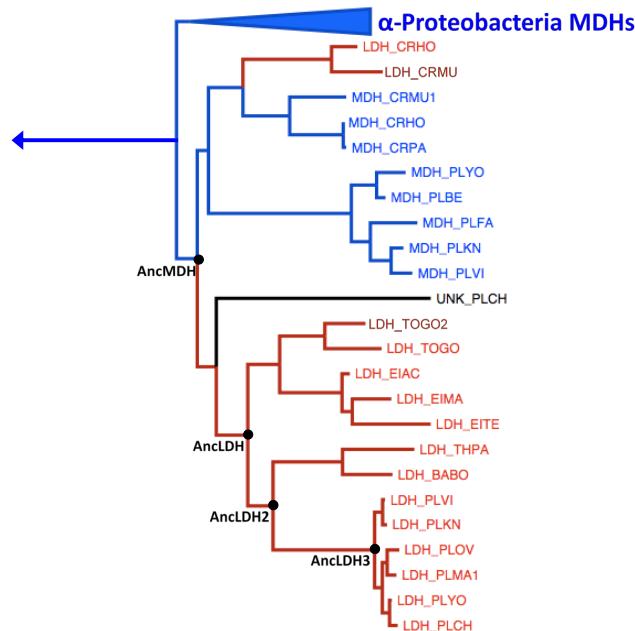


Figure 4 The apicomplexan L/MDH clade. MDHs are labeled blue and LDHs are red. The relevant reconstructed nodes are labeled in black.

apicomplexan LDHs has a GQG motif (or GNG for LDH_BABO) but in *Plasmodium* LDHs there is a single alanine in the same three-dimensional space. Position 236 and 246 identity also appears mutated in the *Plasmodium* genus: G236A and A246P. The alanine and proline are both within 4 Å of pyruvate and as such are likely involved with substrate binding interactions (Figure 6c).

Effects of Single Mutations

Using site-directed mutagenesis, we tested each position of interest for its contribution to steric specificity as well as all possible combinations. Trying to grant LDH_PLFA ancestral levels of promiscuity by mutating the active site residues abolishes enzyme activity to a non-measurable range. Epistasis in protein evolution is well-documented and an extant enzyme resisting reversion

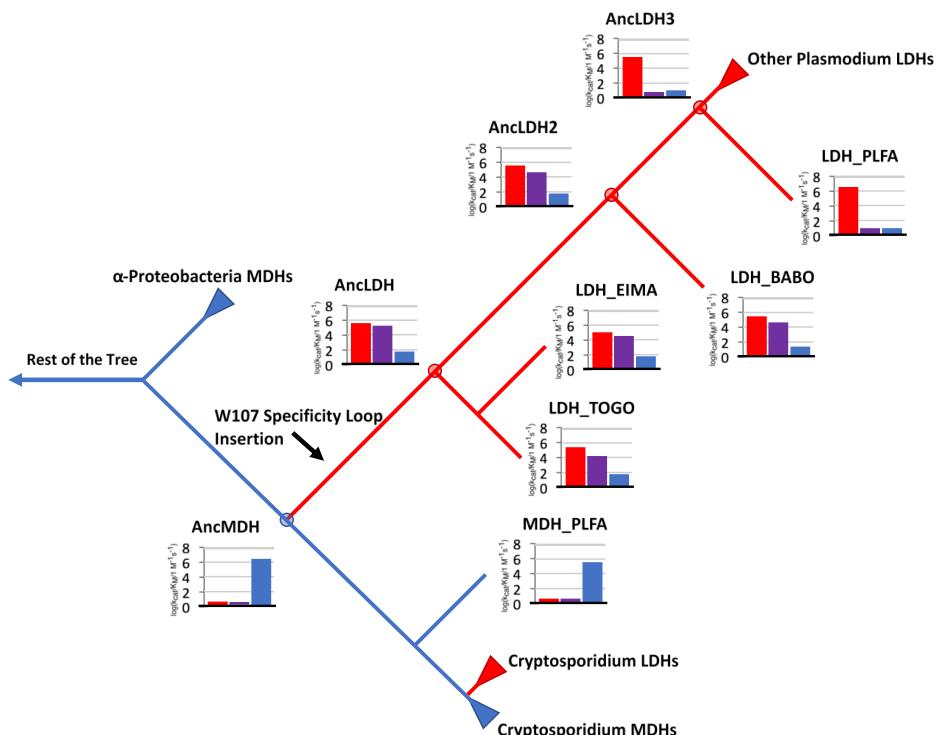


Figure 5 Evolution of Apicomplexa L/MDH Activity. The modern apicomplexan LDHs, MDHs, and ancestral reconstructions are plotted according to their phylogenetic relationship. The *Plasmodium* LDH and the *Plasmodium* ancestor have no phenylpyruvate activity. The coloring is consistent with Figures 3 and 4. Branch lengths are not to scale.

‘backward’ through time to an ancestral state is unsurprising¹⁻². The relevant mutations in our ancestral constructs are less detrimental.

AncLDH2 has effectively no preference between pyruvate and phenylpyruvate (Figure 5 and 7). There is a decrease of five orders of magnitude in ability to turnover large substrates between AncLDH2 and AncLDH3 with almost no change in pyruvate rates. Each active site mutation incorporates into AncLDH2 with a minimal effect on pyruvate activity (Figure 7a). Phenylpyruvate turnover is reduced in AncLDH2_G236A by more than three orders of magnitude (Figure 7b). While this is a considerable drop, phenylpyruvate turnover is still ~10-fold higher than in wild-type AncLDH3, so other residues must also be involved. A246P and the loop deletion have hardly any effect on the AncLDH2 construct. AncLDH3 is resistant to reversion but not as much as its modern counterpart. Pyruvate activity is lowered by a factor of 16 or 130 by mutating positions 236 or 246, respectively, with no accompanying increase in phenylpyruvate recognition.

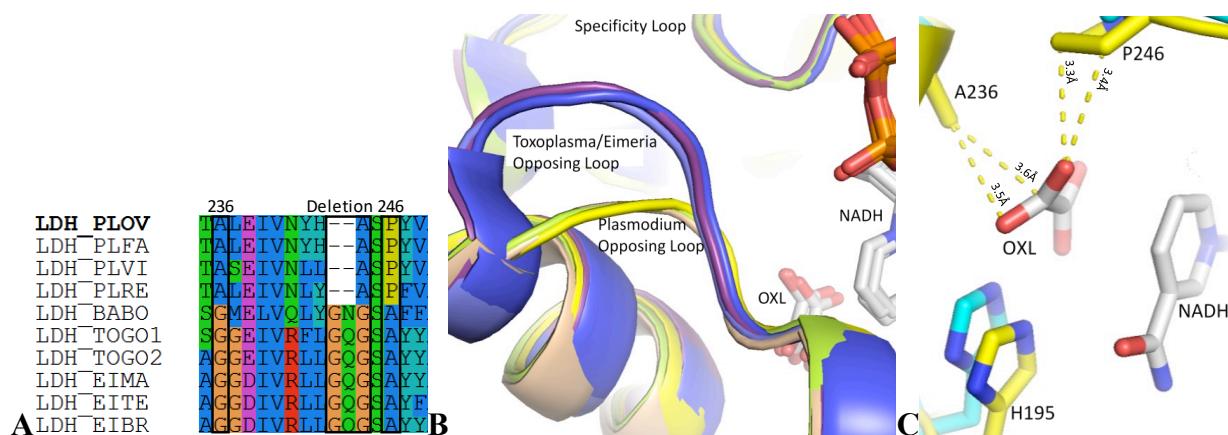


Figure 6 Comparison of apicomplexan LDH active sites. **(A)** Sites 236, 246, and opposing loop residues are generally slow-mutating positions (squared in black). Only *Plasmodium* LDHs have altered sequence identity when compared to other Apicomplexa. **(B)** *Plasmodium* LDHs have a novel deletion in an active site loop (structures from Figure 2). **(C)** Two residues are otherwise highly conserved in Apicomplexa. The unique *Plasmodium* residues A236 and P246 are within 4 Å of pyruvate in the *P. falciparum* LDH structure (Yellow; PDBID: 1T2D).

Combining Mutations and Apparent Protein Epistasis

AncLDH and AncLDH2 have identical active sites and share 93% sequence identity. The evolutionarily relevant mutations have very different effects in the two ancestors, with changes in AncLDH far more deleterious. It follows that residues within the 7% sequence difference act as permissive mutations by making AncLDH2 more tolerant of active site changes. For example, replacing the GQG motif with an alanine in AncLDH causes a 1000-fold decrease in pyruvate activity and an even greater decrease in activity towards a larger substrate. Doing the equivalent

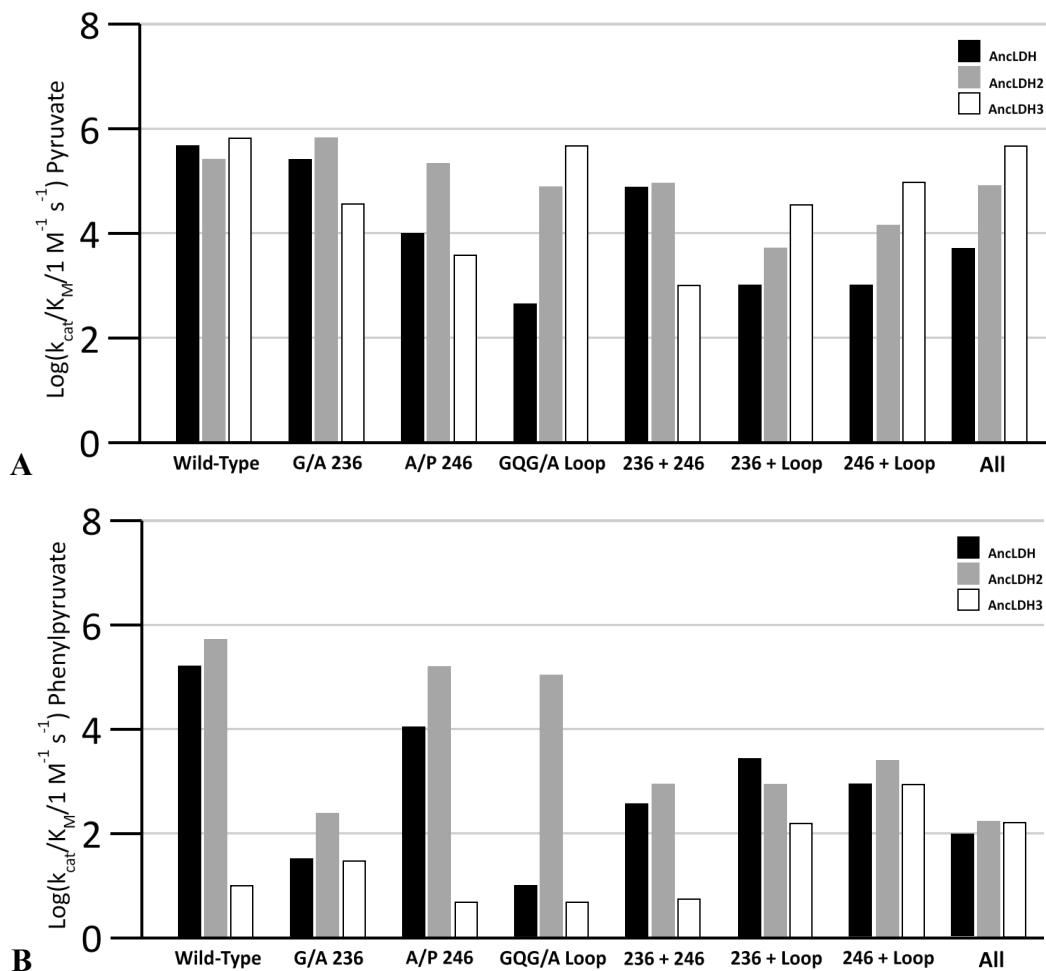


Figure 7 Kinetics of mutant ancestor enzymes **(A)** Pyruvate and **(B)** phenylpyruvate substrate activities are plotted as $\log(k_{\text{cat}}/K_M)$. Each x-axis bin corresponds to the position(s) mutated. ‘All’ refers to the construct that includes both point mutations and the loop mutation. Enzymes are labeled by color.

substitution in AncLDH2 has essentially no effect on either substrate, a clear case of protein epistasis (Figure 7, Supp. Table 5 and 6).

Many of the mutation combinations exhibit non-intuitive interactions, especially when using the larger substrate (Figure 7b). In AncLDH2, the G236A swap alone accounts for a ~3500-fold reduction in phenylpyruvate activity, turnover nearly as slow as with AncLDH3. Coupling G236A with either A246P or the opposing loop mutation, neither having a large impact on their own, reduces the effect to only a few hundred-fold loss in activity (350 and 600, respectively). The equivalent swaps behave similarly in AncLDH3, with each single mutation having amplified (or mitigated) effects when combined that are non-additive.

AncLDH3 was able to recover wild-type pyruvate activity once the entire active site was converted to that of AncLDH2 though curiously some mutant pairs have greatly diminished activity (Figure 7a). Partial rescues in primary substrate activity must be due to epistatic interactions between position 236, 246, and the opposing loop. Interestingly, the AncLDH3 triple mutant only accounted for 100-fold greater phenylpyruvate activity, and so the remaining near-four orders of magnitude difference between the mutant and AncLDH2 must be due to non-active site mutations. Likewise, AncLDH2 cannot be made as specific as AncLDH3 or the modern *Plasmodium* enzyme with active site mutations alone.

Discussion

Current Explanations of LDH Specificity Are Incomplete

Epistasis can dictate a protein's evolutionary trajectory, causing a potentially beneficial mutation in one protein to be harmful in a close relative². LDHs and MDHs were long thought to recognize their substrate solely by using residue 102 on their active site loop. The fact that some canonical LDHs can swap specificities^{12, 19}, while others cannot²⁰⁻²¹, is evidence of epistatic interactions from unknown non-active site residues that help dictate specificity. In recent years, studies have shown that even the classic L/MDH dogma of Q102 vs R102 is not universally correct by characterizing LDHs that use W107f²³ or L102¹⁴ as their primary specificity residue.

Apicomplexan LDHs use W107f which was previously thought to hydrophobically pack against the methyl group of pyruvate but sterically block the larger oxaloacetate as a substrate²³. We have shown here that steric blocking is not a large contribution in the lack of oxaloacetate turnover. Many apicomplexan LDHs can turn over substrates much larger than pyruvate meaning the inability to turnover oxaloacetate is likely only due to charge (Figures 3 and 5). W107f lacks a formal positive charge, which would make oxaloacetate binding disfavored, but the positively charged lysine at position 102 appears positioned such that it could function as a potential charge balancer. The fact that apicomplexan LDHs have poor oxaloacetate despite K102 has been studied but is still unexplained^{10, 25, 35-38}.

It is possible that the five amino acid insertion responsible for pyruvate preference displaces position 102 too far from the active site, making K102 unable to balance a negatively charged substrate. However, in Boucher *et al*, a bifunctional ancestral apicomplexan L/MDH required only a single mutation (K102R). Crystal structures of the mutant construct are in the pyruvate-bound conformation, with W107f in the active site and R102 still far from the substrate

binding pocket, which should make oxaloacetate unable to fit into the available space. The group's hypothesis was that the loop could flip either specificity residue into the active site as needed, which is still being tested.

Plasmodium LDHs Have Atypically High Specificity

LDH_PLFA is unlike its ancestral counterpart because it does not become bifunctional with the K102R mutation, though epistatic 'locking in' of function is not an uncommon occurrence^{1-2, 14, 23, 48}. LDH_PLFA is locked in to pyruvate preference, has unusually high activity, and is exceptionally specific when compared to close relatives (Figures 3 and 5). Reasons for LDH_PLFA's remarkable behavior were hypothesized but had not been empirically explored before now¹⁰. *Plasmodium* LDHs have several mutations that distinguish them from the rest of Apicomplexa and alter the shape of their active sites.

A unique deletion in an active site loop strongly correlates with the inability to use large substrate analogs for turnover (Figure 5 and 6). The two residues 236 and 246 are also unique to *Plasmodium* LDHs and within 4 Å of bound substrate in the closed conformation of the LDH_PLFA (Figure 6). We hypothesize that G236 and A246 which are conserved in other Apicomplexa are more helpful for local flexibility than the *Plasmodium* equivalents A236 and P246. Our data indicates that by evolving these attributes, particularly A236, *Plasmodium* LDHs gained an altered and novel mechanism of substrate recognition, one that includes a strong steric sensor and is absent from other LDHs in the phylum (Figure 5). *Plasmodium* LDHs are more active as well as more specific and it is possible that there is some inherent trade-off between efficiency and promiscuity.

Epistatic interactions Shape Evolutionary Trajectories

There are a few examples of protein epistasis in our results, both inside and outside the active site. The most obvious is the phenotypic differences between AncLDH mutants and AncLDH2 mutants. The proteins share 93% sequence identity and the differences are exclusively outside of the active site. Mutating the opposing loop motif, however, has drastically different effects: AncLDH was mostly inactivated while AncLDH2 was largely unaffected (Figure 7). The two wild-type enzymes behave similarly meaning some combination within the 7% difference (twenty-two residues) allowed the ancestral LDH to accept the historic opposing loop changes as it evolved towards AncLDH3. The mutation G236A, on the other hand, causes an identical change in phenotype for both ancestors and is the largest source of steric discrimination without affecting pyruvate activity (Figure 7).

The reverse mutation does not confer higher promiscuity on AncLDH3. Instead, A236G has little effect on the enzyme. No combination of active site mutation brings AncLDH3 to ancestral levels promiscuity (Figure 7). The dozens of residues that separate AncLDH3 from the other two ancestral constructs constrain specificity despite all being distant from the active site. Only a few AncLDH3 mutants were crippled with respect to pyruvate activity while changes in LDH_PLFA were highly detrimental. While not generally uncommon, the behavior surprised us with LDH_PLFA considering how adaptable the enzyme's primary specificity loop is to mutation^{23, 34}.

Numerous epistatic interactions make it hard to identify route the ancestor LDH must have taken during its evolution to arrive at the modern-day sequence. We know that an increase in steric specificity evolved within the *Plasmodium* lineage and that a change should minimally affect the enzyme's primary role in glycolysis. Given the data, some mutants are more efficient LDHs than

others by orders of magnitude implying some evolutionary paths are much less desirable than others. For instance, if AncLDH2's active site is untouched while the other 90 residues are changed, pyruvate activity is unaffected (Figure 7). There is then only a single active site alteration left available that keeps pyruvate activity at WT levels: G236A. Alternatively, the AncLDH2 active site can mutate through multiple paths with pyruvate activity staying within 10-fold (Figure 7). Some unknown combination of remaining residues increases pyruvate activity and specificity up to AncLDH3 levels.

Specificity Mechanisms, Conservation, and Inhibitor Design

P. falciparum LDH is an active small-molecule drug target because it is essential for the parasite and has an unusual active site architecture^{27, 33, 35, 39-45}. The chief difference between *Plasmodium*-like LDHs and host LDH is the specificity loop, its insertion strongly differentiating it from canonical LDHs and making it a target for selective antibody inhibitors as well⁴⁹. Efforts to selectively target the apicomplexan specificity loop itself may be ineffective, however, due to the lack of functional constraint. Recent work has outlined that the evolution of the loop insertion was largely noncontingent, with its unique length and sequence of little importance, implying a large potential for the development of resistance towards loop-targeting drugs³⁴.

High shared sequence identity within apicomplexan LDHs has led to the belief that successful small-molecule inhibitors selectively targeting the active site (but not necessarily the loop) would likely be effective at treating other apicomplexan-caused diseases^{33, 35-36, 38-39, 43-44}. However, the evolution of a unique substrate recognition mechanism would mean inhibitor interactions studied in a *Plasmodium* LDH model may not be conserved across Apicomplexa. Just

as likely, inhibitors robust towards Apicomplexa such as *Toxoplasma* and *Eimeria* may be rendered ineffective for *Plasmodium* due to different specificity mechanisms.

P. falciparum LDH's strong specificity makes rational inhibitor design difficult without further studying its source. The straightforward approach reported here revealed that the enzyme is unable to reduce large substrate analogs using NADH but gave little insight into effects on binding potential. AncLDH3 has a K_M value for phenylpyruvate 10-fold weaker than AncLDH2 but a k_{cat} 3600-fold slower (Figure 7). If the rate limiting step in the ancestors is also closure of the specificity loop, large substrates are possibly rejected due to the inability for the loop close and occlude water proficiently. Active site substitutions in AncLDH2 that lower activity towards large substrates without affecting pyruvate rates accomplish it primarily through reducing k_{cat} while K_M is relatively unaffected.

Slow loop closure could allow substrate to dissociate more easily before catalysis, causing a sufficiently large increase in k_{off} that the observed K_M remains the same⁵⁰. This suggests that it will be highly difficult to find effective inhibitors or other substrate analogs that selectively target the unique *Plasmodium* LDH active site. The data presented here is insufficient to elucidate true binding affinity, however. Whether a change in k_{off} or k_{on} was the key factor for *Plasmodium* when evolving such strong substrate preference is unknown and those hypotheses must be tested in the future.

Methods

Phylogenetics

Protein sequences used for phylogenetic tree construction were obtained using either the RefSeq database⁵¹ with the BLASTP search algorithm⁵² using the chosen query sequences. Four query sequences were used to ensure sufficient coverage for the superfamily: UniProtIDs P11708, Q76NM3, C6KT25, and Q73G44. Redundant sequences, constructs, and PDB sequences were removed. Sequences were curated based on length and sequence identity in order to reduce computation time. The dataset was trimmed down to 60% sequence identity with apicomplexan sequences reintroduced to ensure full coverage of the nodes of interest. The final dataset had 277 taxa. A sequence alignment was generated with MUSCLE and a ML tree was inferred using PhyML⁵³ using the LG substitution matrix⁵⁴ and estimating the gamma parameter using 12 categories.

Ancestral Sequence Reconstruction

Sequences at internal branch points in the phylogenies were reconstructed using the codeml function in the PAML program suite⁵⁵. Posterior amino acid probabilities were calculated using the LG substitution matrix⁵⁴ given the tree generated by PhyML⁵³. The reconstructions estimated background frequencies of amino acids from the alignment. To assist with proper gene expression, the N- and C-termini were modified manually to match the closest modern sequence, which was determined by branch length.

Plasmid Construction and Mutagenesis

For all proteins, codon optimized genes were synthesized by GenScript (Piscataway, NJ) and sub-cloned into pET-24a, between the NdeI and XhoI restriction sites, without the N-terminal T7-tag but using the C-terminal 6xHistidine-tag. All point mutations were made using the QuikChange Lightning kit from Agilent (Santa Clara, CA) and all relevant mutagenic primers were synthesized by IDT (Coralville, IA). Larger mutations (Indels >1 amino acid) were performed by GenScript directly using the previously synthesized construct as a template. Sequences were confirmed by Sanger Sequencing at Genewiz (Cambridge, Massachusetts). Sequence numbering for all mutations is taken directly from the sequence alignment.

Protein Expression and Purification

All materials were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Plasmids were transformed into BL21-DE3 (pLysS) *E. coli* cells (Invitrogen, Grand Island, NY) for expression. Cells were grown in a shaker-incubator at 37 °C, 225 RPM agitation in 2xYT media supplemented with 30 mM KH₂PO₄, pH 7.8, 0.1% (w/v) glucose and cell growth was monitored at OD₆₀₀. Once an OD₆₀₀ of 0.5-0.8 was reached, protein expression was induced by adding 0.5 mM isopropyl β-D-1-thiogalactopyranoside to each culture and they were further incubated for 4 hours at 37 °C with 225 RPM agitation. Cells were pelleted by centrifugation and stored at -80 °C.

Pellets were thawed on ice and resuspended in 20 mL of HisTrap Binding Buffer (50 mM NaH₂PO₄, pH 7.4, 300 mM NaCl, 10 mM imidazole) and 2 μL Pierce Universal Nuclease (Thermo Scientific, Rockford, IL). Once resuspended, lysate was sonicated on ice at 35% amplitude in pulses of 30 seconds ON and 20 seconds OFF for 2 minutes. Insoluble cell debris was separated by centrifugation at 18,000xg for 20 minutes. The supernatant was syringe-filtered to 0.22 μm and

then purified using a 5 mL HisTrap FF nickel affinity column (GE Healthcare, Piscataway, NJ). Fractions were eluted with an imidazole buffer (50 mM NaH₂PO₄, pH 7.4, 300 mM NaCl, 500 mM imidazole) on an AKTA Prime Plus (GE Healthcare, Piscataway, NJ). Fractions were analyzed by SDS-PAGE, pooled, and concentrated using Amicon Ultracel-10K centrifugation filter units (Millipore, Billerica, MA). Proteins were buffer-exchange into 50 mM Tris, pH 7.4, 100 mM NaCl, 0.1 mM EDTA, 0.01% azide by a PD-10 column (GE Healthcare, Piscataway, NJ). Enzyme concentrations were determined by the sample's 280 nm absorbance, using extinction coefficients and molecular weights calculated by ExPASy's ProtParam tool.

X-ray Crystallography

Conditions were optimized based on promising hits identified from screens with Crystal Screen and Crystal Screen 2 (Hampton Research, Aliso Viejo, CA). Proteins were crystallized by hanging-drop vapor diffusion (2 μ l protein stock and 2 μ l well solution) in VDX greased plates and siliconized cover slides (Hampton) and equilibrated against 1 mL of the well solution at room temperature. Crystals were cryo-protected by being soaked in a 15% (w/v) dextrose solution for 3 minutes, transferred to a 30% (w/v) dextrose solution, and then quickly flash-frozen in liquid N₂.

Diffraction datasets were collected at the SIBYLS beamline (Lawrence Berkeley National Laboratory, Berkeley, CA). All data was collected at 100 K, indexed, integrated, and called with XDS/XSCALE⁵⁶, and all reflections with a CC(1/2) above 10% were used. Molecular replacement and refinement using these reflections were done with the PHENIX software suite⁵⁷. Modeling was performed using Coot⁵⁸ and model quality was validated with MolProbity⁵⁹⁻⁶⁰. Superpositions were generated using THESEUS⁶¹ and images were rendered with PyMOL⁶². In solving the structure 6CT6, the active site was not fully closed, and accordingly has a disordered loop with

high B-factors. The electron density in this region is poor and insufficient to support modeling the structure of the entire loop. Consequently, a few residues are missing from the solved structures.

Steady-State Kinetics Assays

Enzymes were assayed at 25 °C by recording the change in absorbance at 340 nm for 300 seconds on a Cary 100 Bio (Agilent, Santa Clara, CA) in 50 mM Tris, pH 7.5, 50 mM KCl^{10, 14, 23, 34}. Enzyme concentrations varied as needed and ranged from 1 nM to 10 μM. The NADH concentration was a constant 200 μM, except where noted, well above saturating for most MDHs⁶³ and LDHs⁶⁴, while substrate concentrations were varied. All experiments monitored the oxidation of NADH unless stated otherwise. Data was fit together with Kaleideograph as a chi-squared estimate of either the Michaelis-Menten equation or a Substrate Inhibition model:

$$\frac{V_{max}}{E_{tot}} = \frac{k_{cat}[S]}{K_M + [S]} \quad \text{or} \quad \frac{V_{max}}{E_{tot}} = \frac{k_{cat}[S]}{(K_M + [S] + \frac{[S]^2}{K_i})}$$

The error for each parameter is reported as the standard error of the mean (SEM) from triplicate measurements as estimated by the fitting software.

Oxaloacetate spontaneously decarboxylates to pyruvate at 25 °C under neutral aqueous conditions at a significant rate⁶⁵. Oxaloacetate preparations are known to contain pyruvate contamination (a few percent in purchases from Sigma-Aldrich, dependent on the batch) which must be kept in mind when assaying promiscuous L/MDHs. All oxaloacetate stocks were made fresh directly before each experiment and stored on ice throughout. Enzymes that behave as ‘poor’ LDHs are relatively unaffected by the contaminant. Enzymes that have high pyruvate activity or are especially promiscuous can appear to have high artifactual oxaloacetate activity that is due to pyruvate turnover^{10, 12, 23, 66}. Therefore, oxaloacetate activity for primary LDHs was assayed at high

enzyme concentrations (micromolar amounts) which results in a non-linear biphasic decrease in A_{340} . The initial burst corresponds to pyruvate contamination quickly being turned over by the enzyme and the slow, second linear portion was used as the apparent oxaloacetate reduction^{12, 66}.

Supplemental Information

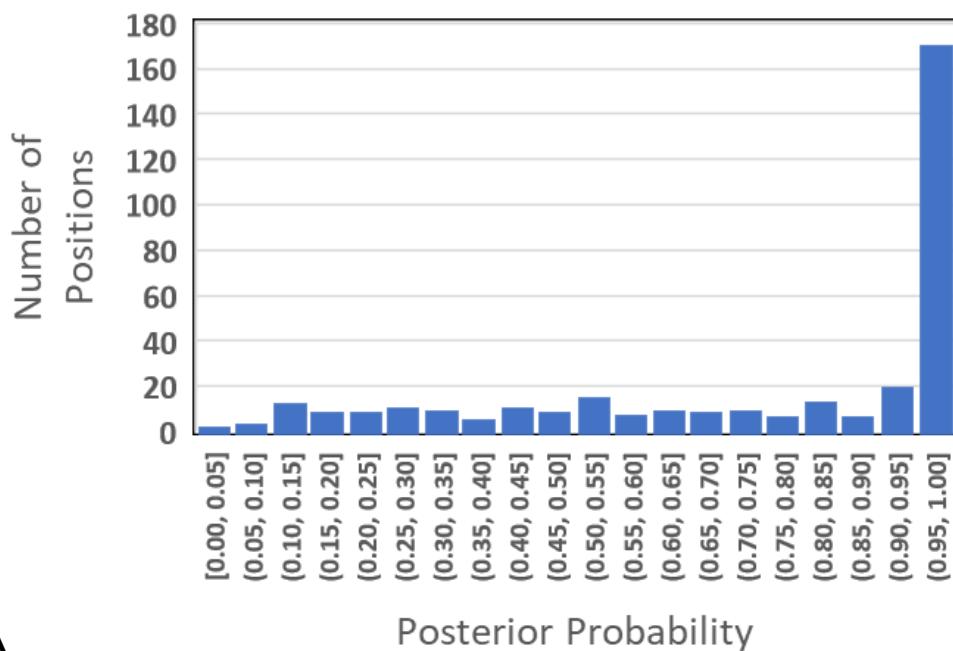
Supp. Table 1 Alternative Substrate Kinetic Parameters

Enzyme	LDH_PLFA			LDH_EIMA			Anc_LDH		
Substrate	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)
Pyruvate	91 ±3	72 ±9	1.3x10 ⁶	7.4 ±0.2	57 ±4	1.3x10 ⁵	45 ±3	100 ±30	4.2x10 ⁵
Phenylpyruvate	0.03 ±0.01	1500 ±200	2.0x10 ¹	16 ±0.4	560 ±60	2.8x10 ⁴	52 ±2	440 ±80	1.2x10 ⁵
α-ketovalerate	0.10 ±0.01	2400 ±300	5.7x10 ¹	3.1 ±0.1	3200 ±400	9.7x10 ²	9.5 ±0.4	3000 ±500	3.1x10 ³
Isocaproate	ND	ND	ND	0.42 ±0.01	2700 ±400	1.6x10 ²	2.0 ±0.1	2500 ±500	8.0x10 ²
Oxaloacetate	ND	ND	1.0x10 ¹	ND	ND	3.8x10 ¹	0.37 ±0.01	6500 ±500	5.6x10 ¹

Supp. Table 2 Steric Specificity Kinetics in Apicomplexan LDHs

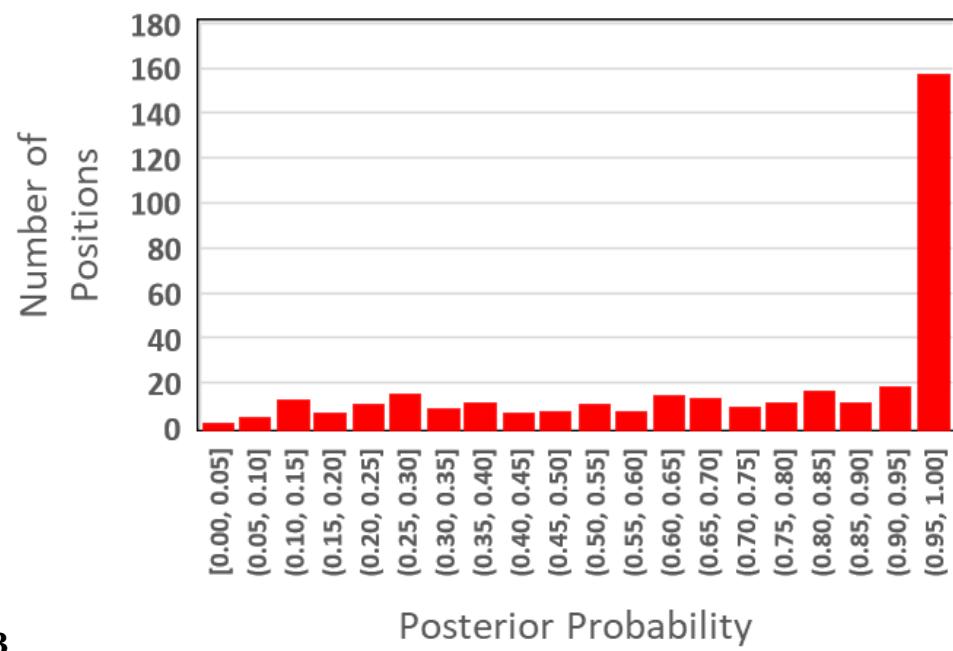
Substrate	Pyruvate			Phenylpyruvate			Oxaloacetate		
Enzyme	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)
MDH_PLFA	ND	ND	ND	ND	ND	ND	64 ±4	120 ±30	5.6x10 ⁵
LDH_PLFA	91 ±3	72 ±9	1.3x10 ⁶	0.03 ±0.01	1500 ±200	2.0x10 ¹	ND	ND	1.0x10 ¹
LDH_EIMA	7.4 ±0.2	57 ±4	1.3x10 ⁵	16 ±0.4	560 ±60	2.8x10 ⁴	ND	ND	3.8x10 ¹
LDH_TOGO	73 ±2	180 ±20	4.2x10 ⁵	24 ±2	1700 ±300	1.4x10 ⁴	1.0 ±0.1	26000 ±6300	3.7x10 ¹
LDH_BABO	40 ±1	82 ±8	4.9x10 ⁵	18 ±1	680 ±100	2.6x10 ⁴	0.61 ±0.08	12000 ±3000	5.3x10 ¹
AncMDH	0.03 ±0.01	16000 ±2300	1.9x10 ⁰	ND	ND	ND	14 ±1	0.80 ±0.3	1.8x10 ⁷
AncLDH	45 ±3	100 ±30	4.2x10 ⁵	52 ±2	440 ±80	1.2x10 ⁵	0.37 ±0.01	6500 ±500	5.6x10 ¹
AncLDH2	6.4 ±0.3	29 ±5	2.2x10 ⁵	74 ±3	110 ±20	6.7x10 ⁵	0.13 ±0.01	520 ±200	2.5x10 ²
AncLDH3	84 ±1	140 ±6	6.2x10 ⁵	0.02 ±0.01	1300 ±300	1.4x10 ¹	3.1 ±0.6	104910 ±2836	3.0x10 ¹

AncMDH



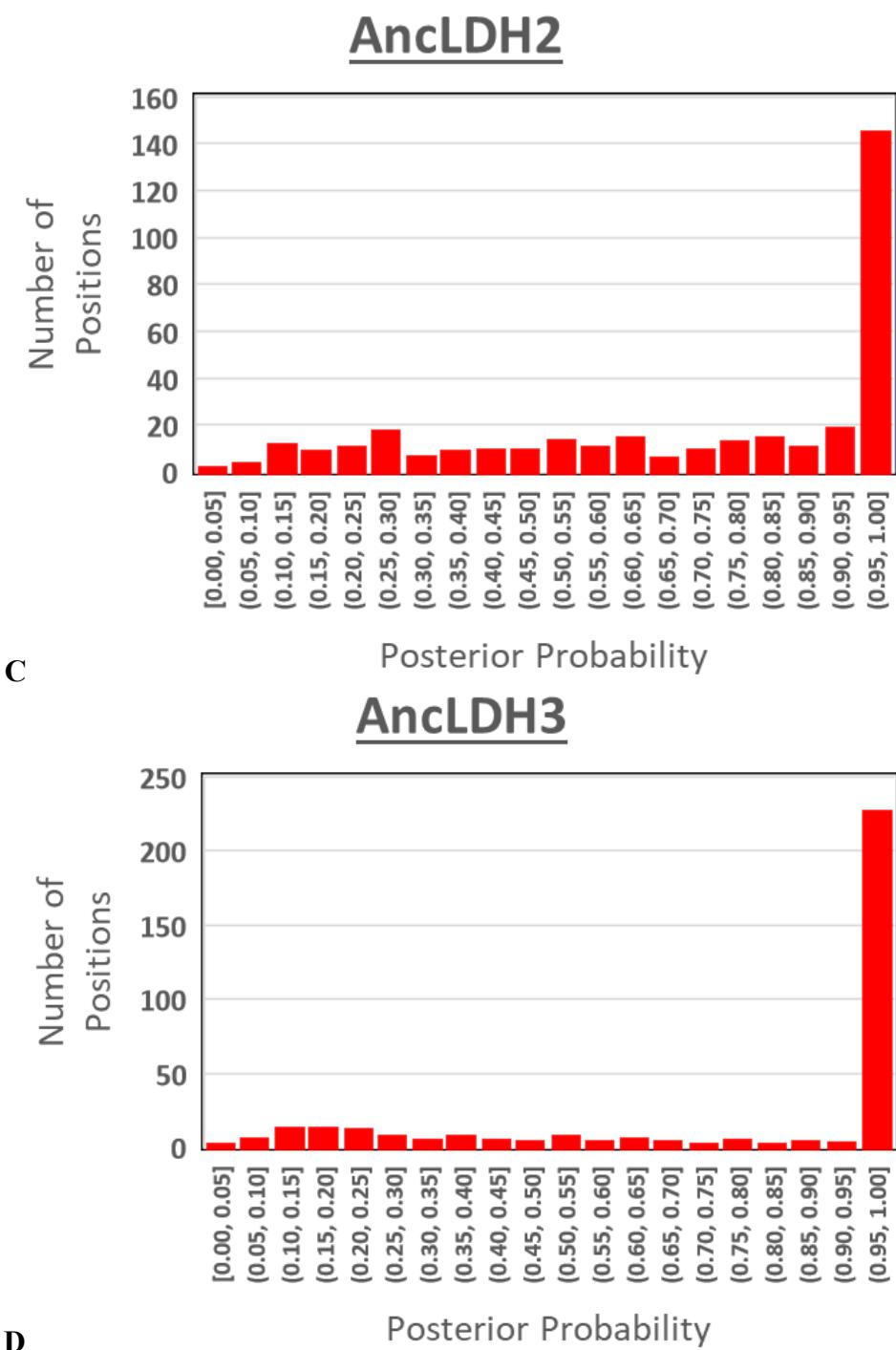
A

AncLDH



B

Posterior Probability



Supp. Figure 1 Posterior probabilities histograms. Each amino acid site in the ancestral reconstructions are binned according to the posterior probability of the predicted residue. **(A)** AncMDH, **(B)** AncLDH, **(C)** AncLDH2, and **(D)** AncLDH3 are colored by primary substrate preference.

Supp. Table 3 Crystallographic Statistics for LDH_EIMA

Data Collection		LDH_EIMA (6CT6)
Space group		P 31 2 1
Cell dimensions		
	<i>a, b, c</i> (Å)	83.4 83.4 227.1
	α, β, γ (°)	90 90 120
Resolution (high res. bin) (Å)		44.64 - 1.705 (1.729 - 1.705)
I/σI (high res. bin)		6.43 (0.34)
R_{meas}		0.134 (3.66)
CC(1/2)		100 (66.6)
Resolution at CC(1/2) = 0.50 (Å)		1.05
Completeness (%)		88.68 (98.5)
Total observations		170863
Unique observations		100328 (9532)
Redundancy		1.84
Wilson B-factor		30.76
Collected at beamline		ALS 12.3.1
Model Refinement		
Resolution (Å)		45.7 – 1.05
$R_{\text{work}}/R_{\text{free}}$		0.2016 / 0.2183
$R_{\text{work}}/R_{\text{free}}$ at CC(1/2) = 0.5 res. cutoff		0.2016 / 0.2183
Mol. Per ASU (protein + ligand)		2
Number of atoms		5438
	protein	4958
	ligand/ions	74
	waters	406
Average B-factor		14.0
r.m.s. deviations		
	Bond lengths (Å)	0.009
	Bond angles (Å)	0.79
Ramachandran favored (%)		97.04
Ramachandran outliers (%)		0.16
All-atom clashscore		0.2

Supp. Table 4 Sequence Identity of Modern and Ancestor Proteins

Enzyme	MDH PLFA	LDH PLFA	LDH EIMA	LDH TOGO	LDH BABO	Anc MDH	Anc LDH	Anc LDH2	Anc LDH3
MDH_PLFA	-	40%	42%	40%	38%	51%	48%	46%	40%
LDH_PLFA	40%	-	52%	49%	51%	65%	63%	66%	91%
LDH_EIMA	42%	52%	-	59%	50%	56%	65%	61%	53%
LDH_TOGO	40%	49%	59%	-	48%	55%	64%	55%	50%
LDH_BABO	38%	51%	50%	48%	-	55%	62%	67%	51%
AncMDH	51%	65%	56%	55%	55%	-	82%	78%	57%
AncLDH	48%	63%	65%	64%	62%	82%	-	93%	66%
AncLDH2	46%	66%	61%	55%	67%	78%	93%	-	70%
AncLDH3	40%	91%	53%	50%	51%	57%	66%	70%	-

Supp. Table 5 Single Mutant Ancestral Apicomplexan LDH Kinetics

Substrate	Pyruvate			Phenylpyruvate			Oxaloacetate		
Enzyme	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)	k _{cat} (s ⁻¹)	K _M (μM ⁻¹)	k _{cat} /K _M (M ⁻¹ s ⁻¹)
AncLDH G236A	100 ±4	470 ±70	2.3x10 ⁵	0.10 ±0.01	3500 ±600	3.5x10 ¹	0.31 ±0.02	7500 ±1000	4.2x10 ¹
AncLDH2 G236A	37 ±1	54 ±8	6.7x10 ⁵	0.30 ±0.01	220 ±200	1.8x10 ²	0.20 ±0.01	5400 ±700	3.9x10 ¹
AncLDH3 A236G	23 ±1	600 ±100	3.8x10 ⁴	0.09 ±0.01	2100 ±700	4.1 x10 ¹	ND	ND	4.9x10 ¹
AncLDH A246P	15 ±1	2400 ±400	6.0x10 ³	20 ±1	1700 ±200	1.2x10 ⁴	0.68 ±0.02	12000 ±2000	5.6x10 ¹
AncLDH2 A246P	15 ±0.5	230 ±20	6.2x10 ⁴	48 ±2	500 ±60	9.6x10 ⁴	0.20 ±0.01	2600 ±400	6.7x10 ¹
AncLDH3 P246A	7.3 ±0.3	1600 ±200	4.6x10 ³	0.04 ±0.01	6600 ±3000	6.0x10 ⁰	5.5 ±3	200000 ±100000	2.8x10 ¹
AncLDH Deletion	14 ±1	25000 ±4000	5.7x10 ²	0.10 ±0.01	2300 ±300	1.7x10 ¹	0.31 ±0.02	10000 ±2000	3.0x10 ¹
AncLDH2 Deletion	39 ±1	730 ±70	5.3x10 ⁴	75 ±2	1300 ±200	5.8x10 ⁴	0.50 ±0.03	8900 ±900	5.5x10 ¹
AncLDH3 Insert	71 ±2	130 ±10	5.4x10 ⁵	0.10 ±0.02	15000 ±6000	6.5x10 ⁰	5.5 ±3	200000 ±100000	2.8x10 ¹
AncLDH Combined	20 ±1	3300 ±600	5.9x10 ³	0.20 ±0.01	2400 ±200	9.4x10 ¹	1.2 ±0.2	45000 ±10000	2.8x10 ¹
AncLDH2 Combined	5.1 ±1	130 ±20	3.9x10 ⁴	0.60 ±0.04	2100 ±600	2.9x10 ²	0.05 ±0.03	9600 ±2000	5.2x10 ¹
AncLDH3 Combined	16 ±0.4	63 ±5	2.5x10 ⁵	0.24 ±0.01	1100 ±100	2.1x10 ²	2.9 ±0.6	96000 ±30000	3.0x10 ¹

Supp. Table 6 Double Mutant Ancestral Apicomplexan LDH Kinetics

Substrate	Pyruvate			Phenylpyruvate			Oxaloacetate		
Enzyme	k_{cat} (s ⁻¹)	K_M (μM^{-1})	k_{cat}/K_M (M ⁻¹ s ⁻¹)	k_{cat} (s ⁻¹)	K_M (μM^{-1})	k_{cat}/K_M (M ⁻¹ s ⁻¹)	k_{cat} (s ⁻¹)	K_M (μM^{-1})	k_{cat}/K_M (M ⁻¹ s ⁻¹)
AncLDH G236A/A246P	57 ± 1	820 ± 80	7.0×10^4	0.90 ± 0.1	1300 ± 300	6.4×10^2	0.32 ± 0.02	5000 ± 800	6.4×10^1
AncLDH2 G236A/A246P	6.3 ± 4	83 ± 20	7.7×10^4	2.1 ± 0.1	1700 ± 300	2.0×10^3	0.10 ± 0.01	1800 ± 300	7.0×10^1
AncLDH3 A236G/P246A	4.1 ± 0.2	1800 ± 300	2.2×10^3	0.03 ± 0.01	2200 ± 400	1.1×10^1	3.0 ± 0.8	94000 ± 40000	3.1×10^1
AncLDH G236A/Del	63 ± 10	26000 ± 8000	2.4×10^3	16 ± 2	3200 ± 300	5.1×10^3	0.80 ± 0.1	31000 ± 5000	2.6×10^1
AncLDH2 G236A/Del	12 ± 0.03	1300 ± 100	9.0×10^3	1.9 ± 0.1	1800 ± 300	1.2×10^3	0.50 ± 0.01	2000 ± 200	2.4×10^2
AncLDH3 A236G/Ins	38 ± 8	870 ± 70	4.3×10^4	0.20 ± 0.01	870 ± 200	2.3×10^2	ND	ND	2.4×10^1
AncLDH A246P/Del	15 ± 0.8	5400 ± 1000	2.8×10^3	6.0 ± 0.3	3000 ± 500	1.9×10^3	0.50 ± 0.03	25000 ± 3000	1.9×10^1
AncLDH2 A246P/Del	10 ± 0.3	460 ± 60	2.2×10^4	30 ± 6	3200 ± 900	9.3×10^3	0.60 ± 0.1	13000 ± 3000	4.3×10^1
AncLDH3 P246A/Ins	22 ± 0.4	140 ± 10	1.5×10^5	0.01 ± 0.01	2300 ± 400	3.7×10^0	1.9 ± 0.3	57000 ± 30000	3.4×10^1

References

1. Harms, M. J.; Thornton, J. W., Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol* **2010**, *20* (3), 360-6.
2. Starr, T. N.; Thornton, J. W., Epistasis in protein evolution. *Protein Sci* **2016**, *25* (7), 1204-18.
3. Weinreich, D. M.; Watson, R. A.; Chao, L., Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **2005**, *59* (6), 1165-74.
4. Rossmann, M. G.; Liljas, A.; Brnden, C.-I.; Banaszak, L. J., *Evolutionary and structural relationships among dehydrogenases*. 3rd ed.; Academic Press: 1975.
5. Birktoft, J. J.; Banaszak, L. J., The presence of a histidine-aspartic acid pair in the active site of 2-hydroxyacid dehydrogenases. X-ray refinement of cytoplasmic malate dehydrogenase. *The Journal of Biological Chemistry* **1983**, *258*, 472-482.
6. Clarke, A. R.; Wigley, D. B.; Chia, W. N.; Barstow, D.; Atkinson, T.; Holbrook, J. J., Site-directed mutagenesis reveals role of mobile arginine residue in lactate dehydrogenase catalysis. *Nature* **1986**, *324* (6098), 699-702.
7. Clarke, A. R.; Wilks, H. M.; Barstow, D. A.; Atkinson, T.; Chia, W. N.; Holbrook, J. J., An investigation of the contribution made by the carboxylate group of an active site histidine-aspartate couple to binding and catalysis in lactate dehydrogenase. *Biochemistry* **1988**, *27* (5), 1617-1622.
8. Waldman, A. D. B.; Hart, K. W.; Clarke, A. R.; Wigley, D. B.; Barstow, D. A.; Atkinson, T.; Chia, W. N.; Holbrook, J. J., The use of a genetically engineered tryptophan to identify the movement of a domain of *B. stearothermophilus* lactate dehydrogenase with the process which limits the steady-state turnover of the enzyme. *Biochemical and Biophysical Research Communications* **1988**, *150* (2), 752-759.
9. Voet, D.; Voet, J. G., *Biochemistry*. 3rd ed.; J. Wiley & Sons: New York, 2004.
10. Shoemark, D. K.; Cliff, M. J.; Sessions, R. B.; Clarke, A. R., Enzymatic properties of the lactate dehydrogenase enzyme from *Plasmodium falciparum*. *Febs Journal* **2007**, *274* (11), 2738-2748.
11. Wilks, H. M.; Halsall, D. J.; Atkinson, T.; Chia, W. N.; Clarke, A. R.; Holbrook, J. J., Designs for a broad substrate specificity keto acid dehydrogenase. *Biochemistry* **1990**, *29* (37), 8587-91.
12. Wilks, H. M.; Hart, K. W.; Feeney, R.; Dunn, C. R.; Muirhead, H.; Chia, W. N.; Barstow, D. A.; Atkinson, T.; Clarke, A. R.; Holbrook, J. J., A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* **1988**, *242* (4885), 1541-4.
13. Chapman, A. D. M.; Cortes, A.; Daform, T. R.; Clarke, A. R.; Brady, R. L., Structural Basis of Substrate Specificity in Malate Dehydrogenases: Crystal Structure of a Ternary Complex of Porcine Cytoplasmic Malate Dehydrogenase, a-Ketomalonate and TetrahydNAD. *Journal of Molecular Biology* **1999**, *285*, 703-712.
14. Steindel, P. A.; Chen, E. H.; Wirth, J. D.; Theobald, D. L., Gradual neofunctionalization in the convergent evolution of trichomonad lactate and malate dehydrogenases. *Protein Sci* **2016**, *25* (7), 1319-31.
15. Clarke, A. R.; Holbrook, J. J., The rational construction of new enzymes: design and synthesis of a dehydrogenase which accepts an unnatural substrate. *Sci Prog* **1989**, *73* (289 Pt 1), 117-31.

16. Golding, G. B.; Dean, A. M., The structural basis of molecular adaptation. *Mol Biol Evol* **1998**, *15* (4), 355-69.
17. Goward, C. R.; Nicholls, D. J., Malate dehydrogenase: a model for structure, evolution, and catalysis. *Protein Sci* **1994**, *3* (10), 1883-8.
18. Madern, D., Molecular Evolution within the L-Malate and L-lactate dehydrogenase super-family. *Journal of Molecular Evolution* **2002**, *54* (6), 825-840.
19. Ohta, T.; Yokota, K.; Minowa, T.; Iwata, S., Mechanism of allosteric transition of bacterial L-lactate dehydrogenase. *Faraday discussions* **1992**, (93), 153-62.
20. Cendrin, F.; Chroboczek, J.; Zaccai, G.; Eisenberg, H.; Mevarech, M., Cloning, Sequencing, and Expression in *Escherichia coli* of the Gene Coding for Malate Dehydrogenase of the Extremely Halophilic Archaeabacterium *Haloarcula marismortui*. *Biochemistry* **1993**, *32*, 4308-4313.
21. Nicholls, D. J.; Miller, J.; Scawen, M. D.; Clarke, A. R.; Holbrook, J. J.; Atkinson, T.; Goward, C. R., The importance of arginine 102 for the substrate specificity of *Escherichia coli* malate dehydrogenase. *Biochem Biophys Res Commun* **1992**, *189* (2), 1057-62.
22. Madern, D.; Cai, X. M.; Abrahamsen, M. S.; Zhu, G., Evolution of *Cryptosporidium parvum* lactate dehydrogenase from malate dehydrogenase by a very recent event of gene duplication. *Molecular Biology and Evolution* **2004**, *21* (3), 489-497.
23. Boucher, J. I.; Jacobowitz, J. R.; Beckett, B. C.; Classen, S.; Theobald, D. L., An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife* **2014**, *3*.
24. Bzik, D. J.; Fox, B. A.; Gonyer, K., Expression of *Plasmodium falciparum* lactate dehydrogenase in *Escherichia coli*. *Molecular and Biochemical Parasitology* **1993**, *59* (1), 155-166.
25. Brown, W. M.; Yowell, C. A.; Hoard, A.; Vander Jagt, T. A.; Hunsaker, L. A.; Deck, L. M.; Royer, R. E.; Piper, R. C.; Dame, J. B.; Makler, M. T.; Vander Jagt, D. L., Comparative Structural Analysis and Kinetic Properties of Lactate Dehydrogenases from the Four Species of Human Malarial Parasites. *Biochemistry* **2004**, *43*, 6219-6229.
26. Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R.; Carucci, D. J., A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **2002**, *419*, 520.
27. Royer, R. E.; Deck, L. M.; Campos, N. M.; Hunsaker, L. A.; Vander Jagt, D. L., Biologically-Active Derivatives of Gossypol - Synthesis and Antimalarial Activities of Peri-Acylated Gossylic Nitriles. *Journal of Medicinal Chemistry* **1986**, *29* (9), 1799-1801.
28. Douzery, E. J.; Snell, E. A.; Bapteste, E.; Delsuc, F.; Philippe, H., The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (43), 15386-15391.
29. Zhu, G.; Keithly, J. S., Alpha-proteobacterial relationship of apicomplexan lactate and malate dehydrogenases. *Journal of Eukaryotic Microbiology* **2002**, *49* (3), 255-261.
30. Madern, D.; Ebel, C.; Mevarech, M.; Richard, S. B.; Pfister, C.; Zaccai, G., Insights into the molecular relationships between malate and lactate dehydrogenases: structural and biochemical properties of monomeric and dimeric intermediates of a mutant of tetrameric L-[LDH-like] malate dehydrogenase from the halophilic archaeon *Haloarcula marismortui*. *Biochemistry* **2000**, *39* (5), 1001-10.

31. Dunn, C. R.; Banfield, M. J.; Barker, J. J.; Higham, C. W.; Moreton, K. M.; Turgut-Balik, D.; Brady, R. L.; Holbrook, J. J., The structure of lactate dehydrogenase from *Plasmodium falciparum* reveals a new target for anti-malarial design. *Nature Structural Biology* **1996**, *3*, 912-915.
32. Read, J. A.; Winter, V. J.; Eszes, C. M.; Sessions, R. B.; Brady, R. L., Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase. *Proteins* **2001**, *43* (2), 175-85.
33. Cameron, A.; Read, J.; Tranter, R.; Winter, V. J.; Sessions, R. B.; Brady, R. L.; Vivas, L.; Easton, A.; Kendrick, H.; Croft, S. L.; Barros, D.; Lavandera, J. L.; Martin, J. J.; Risco, F.; Garcia-Ochoa, S.; Gamo, F. J.; Sanz, L.; Leon, L.; Ruiz, J. R.; Gabarro, R.; Mallo, A.; Gomez de las Heras, F., Identification and activity of a series of azole-based compounds with lactate dehydrogenase-directed anti-malarial activity. *J Biol Chem* **2004**, *279* (30), 31429-39.
34. Wirth, J. D.; Boucher, J. I.; Jacobowitz, J. R.; Classen, S.; Theobald, D. L., Functional and Structural Resilience of the Active Site Loop in the Evolution of *Plasmodium* Lactate Dehydrogenase. *Biochemistry* **2018**, *57* (45), 6434-6442.
35. Gomez, M. S.; Piper, R. C.; Hunsaker, L. A.; Royer, R. E.; Deck, L. M.; Makler, M. T.; Vander Jagt, D. L., Substrate and cofactor specificity and selective inhibition of lactate dehydrogenase from the malarial parasite *P. falciparum*. *Molecular and Biochemical Parasitology* **1997**, *90*, 235-246.
36. Dando, C.; Schroeder, E. R.; Hunsaker, L. A.; Deck, L. M.; Royer, R. E.; Zhou, X.; Parmley, S. F.; Vander Jagt, D. L., The kinetic properties and sensitivities to inhibitors of lactate dehydrogenases (LDH1 and LDH2) from *Toxoplasma gondii*: comparisons with pLDH from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **2001**, *118* (23-32).
37. Winter, V. J.; Cameron, A.; Tranter, R.; Sessions, R. B.; Brady, R. L., Crystal structure of *Plasmodium berghei* lactate dehydrogenase indicates the unique structural differences of these enzymes are shared across the *Plasmodium* genus. *Molecular and Biochemical Parasitology* **2003**, *131* (1), 1-10.
38. Kavanagh, K. L.; Elling, R. A.; Wilson, D. K., Structure of *Toxoplasma gondii* LDH1: Active-site differences from human lactate dehydrogenases and the structural basis for efficient APAD(+) use. *Biochemistry* **2004**, *43*, 879-889.
39. Read, J. A.; Wilkinson, K. W.; Tranter, R.; Sessions, R. B.; Brady, R. L., Chloroquine binds in the cofactor binding site of *Plasmodium falciparum* lactate dehydrogenase. *Journal of Biological Chemistry* **1999**, *274* (15), 10213-10218.
40. Conners, R.; Schambach, F.; Read, J.; Cameron, A.; Sessions, R. B.; Vivas, L.; Easton, A.; Croft, S. L.; Brady, R. L., Mapping the binding site for gossypol-like inhibitors of *Plasmodium falciparum* lactate dehydrogenase. *Mol Biochem Parasitol* **2005**, *142* (2), 137-48.
41. Brady, R. L.; Cameron, A., Structure-based approaches to the development of novel anti-malarials. *Current drug targets* **2004**, *5* (2), 137-49.
42. Vivas, L.; Easton, A.; Kendrick, H.; Cameron, A.; Lavandera, J. L.; Barros, D.; de las Heras, F. G.; Brady, R. L.; Croft, S. L., *Plasmodium falciparum*: stage specific effects of a selective inhibitor of lactate dehydrogenase. *Experimental parasitology* **2005**, *111* (2), 105-14.
43. Choi, S.-r.; Pradhan, A.; Hammond, N. L.; Chittiboyina, A. G.; Tekwani, B. L.; Avery, M. A., Design, Synthesis, and Biological Evaluation of *Plasmodium falciparum* Lactate Dehydrogenase Inhibitors. *Journal of Medicinal Chemistry* **2007**, *50* (16), 3841-3850.
44. Granchi, C.; Bertini, S.; Macchia, M.; Minutolo, F., Inhibitors of lactate dehydrogenase isoforms and their therapeutic potentials. *Current medicinal chemistry* **2010**, *17* (7), 672-97.

45. Saxena, S.; Durgam, L.; Guruprasad, L., Multiple e-pharmacophore modelling pooled with high-throughput virtual screening, docking and molecular dynamics simulations to discover potential inhibitors of *Plasmodium falciparum* lactate dehydrogenase (PfLDH). *Journal of biomolecular structure & dynamics* **2018**, 1-17.

46. Nural, B.; Erdemir, A.; Mutlu, O.; Yakarsonmez, S.; Danis, O.; Topuzogullari, M.; Turgut-Balik, D., Biochemical and in silico Characterization of Recombinant L-Lactate Dehydrogenase of *Theileria annulata*. *Molecular Biotechnology* **2016**, 58 (4), 256-267.

47. Chaikuad, A.; Fairweather, V.; Conners, R.; Joseph-Horne, T.; Turgut-Balik, D.; Brady, R. L., Structure of lactate dehydrogenase from *Plasmodium vivax*: complexes with NADH and APADH. *Biochemistry* **2005**, 44 (49), 16221-8.

48. Sailer, Z. R.; Harms, M. J., High-order epistasis shapes evolutionary trajectories. *PLoS Comput Biol* **2017**, 13 (5), e1005541.

49. Kaushal, N. A.; Kaushal, D. C., Production and characterization of monoclonal antibodies against substrate specific loop region of *Plasmodium falciparum* lactate dehydrogenase. *Immunological investigations* **2014**, 43 (6), 556-71.

50. Fersht, A., Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding. W.H. Freeman: New York :, 1999.

51. O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **2016**, 44 (D1), D733-45.

52. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J Mol Biol* **1990**, 215 (3), 403-10.

53. Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **2010**, 59 (3), 307-21.

54. Le, S. Q.; Gascuel, O., An improved general amino acid replacement matrix. *Mol Biol Evol* **2008**, 25 (7), 1307-20.

55. Yang, Z., PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **2007**, 24 (8), 1586-91.

56. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010**, 66 (Pt 2), 125-32.

57. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010**, 66 (Pt 2), 213-21.

58. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica* **2010**, 66 (Pt 4), 486-501.

59. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, I. I. W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity:

all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **2007**, *35* (suppl_2), W375-W383.

60. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 1), 12-21.
61. Theobald, D. L.; Wuttke, D. S., Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput Biol* **2008**, *4* (2), e43.
62. Schrodinger, L., The {PyMOL} Molecular Graphics System, Version~1.3r1. 2010.
63. Banaszak, L. J.; Bradshaw, R. A., 6 Malate Dehydrogenases. In *The Enzymes*, Boyer, P. D., Ed. Academic Press: 1975; Vol. 11, pp 369-396.
64. Holbrook, J. J.; Liljas, A.; Steindel, S. J.; Rossmann, M. G., 4 Lactate Dehydrogenase. In *The Enzymes*, Boyer, P. D., Ed. Academic Press: 1975; Vol. 11, pp 191-292.
65. Wolfenden, R.; Lewis, C. A., Jr.; Yuan, Y., Kinetic challenges facing oxalate, malonate, acetoacetate, and oxaloacetate decarboxylases. *J Am Chem Soc* **2011**, *133* (15), 5683-5.
66. Parker, D. M.; Holbrook, J. J., The oxaloacetate reductase activity of vertebrate lactate dehydrogenase. *Int J Biochem* **1981**, *13* (10), 1101-5.