The Mechanistic, Structural, and Evolutionary Origin of Lactate Dehydrogenase Substrate Specificity in the Apicomplexa

Master’s Thesis
Presented to

The Faculty of Graduate School of Arts and Sciences
Brandeis University
Department of Biochemistry
Douglas Theobald, Advisor

In Partial Fulfillment
of the Requirements for the Degree

Master of Science
in
Biochemistry

by
Changhan Xu

May 2017
Acknowledgements

Before getting to my own work, I would like to pause and express my gratitude.

First and foremost, I would like to thank Professor Theobald for letting me into his lab. I came into his lab knowing essentially nothing about the subject, and it was only because of this opportunity that I was able to gain some real insights into the real world of biochemistry. Now as I am ready to graduate, I believe I have learned way more than I expected when I first joined.

After the boss, I would like to thank my mentor Jacob. Jacob (and sometimes Doug too) has been very tolerant with my limited communication skills in English for more than two years and yet, he managed to teach me and help me with all the lab stuff.

Of course, all the other lab members have also been very helpful. Rick, Miriam, and Brian have made the Theobald lab a delightful place to work.

The SHARKK PLUG and the speakers are amazing. There is even a bass.

Outside the lab, the Biochemistry Office has been a huge help whenever I had a question with pretty much anything about any process in the department. Both Jennifer Roy and Maryanna Aldrich have been very helpful.

Outside of Waltham, I would like to thank my family. They made absolutely no contribution whatsoever to this particular thesis project. BUT, I wouldn’t be doing any of this without their full support, and I am very grateful for all they have done for me.
Abstract

The Mechanistic, Structural, and Evolutionary Origin of Lactate Dehydrogenase Substrate Specificity in the Apicomplexa

A thesis presented to the Department of Biochemistry
Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts
By Changhan Xu

Apicomplexa are unicellular eukaryotic parasites of many animals, including humans. The most notorious example is malaria, caused by the apicomplexan parasite *Plasmodium falciparum*. The apicomplexan lactate dehydrogenase enzyme (LDH) evolved independently from human and other “canonical” LDHs found in the host. Apicomplexan LDHs have distinct structures from the host LDHs, and thus they have been identified as a drug target for the parasitic diseases. *Plasmodium falciparum* LDH (*Pf*LDH) has been extensively studied due to its value for drug targeting, and yet the mechanistic and evolutionary basis of the functional and structural divergence of the entire apicomplexan LDH family is still not well-understood. A related apicomplexan parasite, *Eimeria maxima*, is the cause of an economically significant disease of poultry. To further explore the source of the functional divergence in apicomplexan LDHs, I characterized the kinetic and specificity profile for *Pf*LDH, *Eimeria maxima* LDH (*Em*LDH), and their most recent common ancestor (*Anc*LDH) calculated from extant sequences, using different substrates. *Pf*LDH diverged from *Anc*LDH by evolving high specificity toward pyruvate, having very low activities toward other bigger and hydrophobic substrates. *Em*LDH, on the other hand, largely retains the specificity profile from the *Anc*LDH ancestor: both *Em*LDH and *Anc*LDH have reduced activities toward other alternative substrates, but the magnitudes were no match to *Pf*LDH.
The most prominent difference can be seen when phenylpyruvate was used as an alternative substrate. \( PfLDH \) has very limited activity toward this hydrophobic substrate, while \( AncLDH \) and \( EmLDH \) are both able to ignore the phenyl ring and turn over phenylpyruvate at an efficiency only marginally lower than turning over the native substrate pyruvate. I have also solved the first crystal structure for \( EmLDH \) (1.8 Å). Structural comparison of \( EmLDH \), \( PfLDH \), and \( AncLDH \) shows several residue differences near the active site, and it is likely that the differences in their specificity profiles can be attributed to these residues. Preliminary data also suggests that the rate-limiting step for \( EmLDH \) is after the hydride transfer step, similar to \( PfLDH \). This work will serve as a foundation for further understanding the origin of the convergent evolution of LDH function in the Apicomplexa.
Contents

Acknowledgements ........................................ iii
Abstract ................................................. iv
List of Figures .......................................... viii
List of Tables ........................................... ix

1 Introduction ........................................... 1
  1.1 Canonical and apicomplexan dehydrogenases ................. 3
  1.2 The phylogenetic approach ................................ 9
  1.3 LDHs in Plasmodium falciparum and Eimeria maxima .......... 12
  1.4 Goal of the thesis ..................................... 15
  1.5 Experimental background ................................ 16
     1.5.1 Steady-state kinetics ................................ 16
     1.5.2 Steady-state approximation ......................... 17
     1.5.3 Pre-steady-state kinetics ......................... 17
     1.5.4 X-Ray crystallography .............................. 19

2 Material and methods .................................. 20
  2.1 Plasmid construction and mutation ......................... 20
  2.2 Protein expression .................................... 20
  2.3 Protein purification ................................... 21
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>PfLDH</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>LDH monomer</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>Gln102 vs. Trp107</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Scheme of PfLDH</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>Molecular mechanism</td>
<td>7</td>
</tr>
<tr>
<td>1.6</td>
<td>Trp107 in PfLDH</td>
<td>9</td>
</tr>
<tr>
<td>1.7</td>
<td>Relevant clade from phylogeny</td>
<td>11</td>
</tr>
<tr>
<td>1.8</td>
<td>The opposing loop</td>
<td>14</td>
</tr>
<tr>
<td>2.1</td>
<td>Substrate and Alternative Substrates</td>
<td>23</td>
</tr>
<tr>
<td>3.1</td>
<td>Activity profiles</td>
<td>31</td>
</tr>
<tr>
<td>3.2</td>
<td>Superpositions of known apicomplexan LDHs</td>
<td>34</td>
</tr>
<tr>
<td>3.3</td>
<td>Multiple turnover experiment</td>
<td>39</td>
</tr>
<tr>
<td>6.1</td>
<td>Sequence alignment</td>
<td>50</td>
</tr>
<tr>
<td>6.2</td>
<td>Numbering system</td>
<td>51</td>
</tr>
<tr>
<td>6.3</td>
<td>Numbering system continued</td>
<td>52</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Steady-state kinetics for pyruvate ........................................ 27
3.2 Residue replacements ...................................................... 36
3.3 Mutants ................................................................. 38

6.1 $Em$LDH steady-state kinetics ........................................ 53
6.2 $Pf$LDH steady-state kinetics ........................................ 53
6.3 AncLDH steady-state kinetics ........................................ 53
6.4 Table 1 generated by PHENIX. ..................................... 54
Chapter 1

Introduction

Lactate dehydrogenase (LDH) and malate dehydrogenase (MDH) are homologous enzymes that share similar structures (overall fold), mechanisms, and substrates [1–5]. LDH is involved in anaerobic metabolism, where it catalyzes the interconversion between pyruvate and lactate (regenerating NAD$^+$ and producing lactate when used in the anaerobic metabolism), while MDH is involved in the last step in the citric acid cycle, catalyzing the interconversion between oxaloacetate and malate. These two enzymes are ubiquitous in all life forms on Earth. In fact, previous work from the Theobald lab has shown that LDHs were evolved from MDHs on at least four independent occasions [6]. Apicomplexan LDHs and the “canonical” LDHs that many other organisms, including human, evolved from MDHs in two independent events. This results in their different activity profiles as well as structures. These disparities make the apicomplexan LDHs a potential drug target in fighting many metazoan diseases caused by apicomplexan parasites, including malaria (*Plasmodium falciparum*), toxoplasmosis (*Toxoplasma*), and babesiosis (*Babesia*) [6–10]. Even though these are all LDHs, their functional differences are not insignificant: their substrate specificity can be drastically different, and when it comes to their native substrates, pyruvate and oxaloacetate, their specificity
can differ by twelve orders of magnitude. All enzymes can catalyze some side-reactions, just not as efficient as its native catalysis [11]. This phenomenon is known as promiscuity, the very opposite of specificity, where the enzyme can recognize one specific substrate with much greater precision, and turn over the substrate with a much higher efficiency than other similar substrates [12]. For example, some apicomplexan LDHs are more specific at turning over their native substrate pyruvate than others. Due to its high medicinal potential values, *Plasmodium falciparum* LDH (*Pf*LDH, Figure 1.1) has been studied extensively, and it has been shown that *Pf*LDH is a highly efficient and specific LDH.
Figure 1.1: *Plasmodium falciparum* LDH (PDB ID: 1T2D) as a tetramer. Each monomer has its own active site, and no cooperativity was found [13, 14]. The topology and fold of the monomer are also shared with the “canonical” LDHs.

1.1 Canonical and apicomplexan dehydrogenases

Despite their high structural similarities and catalytic mechanisms, both LDHs and MDHs recognize their respective substrates (pyruvate and oxaloacetate) with high specificities. They employ different mechanisms to govern their specificities. In MDHs and canonical LDHs, a “specificity residue” was found at position 102 by Wilks et al [15]. This position is
one of the two key differences among all known MDHs, canonical LDHs, and apicomplexan LDHs: all known MDHs have a positively charged arginine at the 102 to interact with the negatively charged oxaloacetate, all canonical LDHs have a neutral glutamine at the 102 position to interact with the methyl group on pyruvate, and all known apicomplexan LDHs have a lysine at the 102 position (with the exception of Cryptosporidium parvum LDH, CpLDH) with the side chain pointing away from the active site.

In addition, apicomplexan LDHs have a five-residue insertion in the specificity loop (Figure 1.2) that closes upon binding of pyruvate and seals the active site. This insertion contains the most important residue for apicomplexan LDHs that governs the substrate specificity: Trp107 [6,16]. This residue, along with the entire insertion (Figure 1.3), is absent in canonical LDHs.
CHAPTER 1. INTRODUCTION

Figure 1.2: The monomer form of PfLDH (PDB ID: 1T2D). This fold is shared across the canonical LDHs and apicomplexan LDHs. The specificity loop closes and covers the active it upon substrate binding. Oxalate was the substrate analog used to crystalize the protein, and it resides in the active site.

Previously it has been shown that charge complementarity in the active site plays an important role in both dehydrogenases in their specificity toward pyruvate (one negative charge) and oxaloacetate (two negative charges) [17]. If this is the entire story, then, the positive charge on Lys102 is expected to have similar interactions with the bigger substrate oxaloacetate, and thus the apicomplexan LDHs are expected to reduce oxaloacetate [6]. However, Lys102 is not even in the active site in crystal structures: it points away from the
Figure 1.3: Superposition of Human (PDB ID: 1L10) in light purple and *Pf*LDH (PDB ID: 1T2D) in yellow. The specificity loop closes and seals the active site upon binding of pyruvate. Oxamate/oxalate were the substrate analogs used to crystalize the two LDHs instead of pyruvate, and is shown in sphere. The space where human LDH has the specificity residue Gln102 is occupied by a Trp at position 107 in *Pf*LDH. This is also true for other apicomplexan LDHs. The Trp107 is critical for the function of apicomplexan LDHs.

 Consequently an alternative mechanism is required to control the specificity of the dehydrogenases.
Figure 1.4: Scheme of PfLDH and other LDHs [18]. It has been assumed that all the LDHs share this similar mechanisms.

Figure 1.5: Mechanism of apicomplexan LDHs [6]. Only the steps starting from the conformational change of the enzyme (loop closure) upon binding of the substrate and the chemical step (hydride transfer) are shown here.
CHAPTER 1. INTRODUCTION

Figure 1.5: Relevant catalytic residues are shown. K102 is not shown because it is not pointing into the active site in apicomplexan LDHs, which is not the case for canonical LDHs or MDHs.

The answer may be found on the “specificity loop” (Figure 1.2) covering the active site [6]. The mechanism for LDH is shown in Figure 1.4 [18]. It has long been assumed that all LDHs share this similar, ordered bi-bi mechanism (Figure 1.4) [18] due to their high structural similarities. Loop closure over the active site following pyruvate binding was previously thought to be the rate-limiting during catalysis in the majority LDH/MDH superfamily [5], and mutations on this loop affects activity and substrate specificity to a large extent [15]. However, later studies showed that the apicomplexan parasite *Toxoplasma gondii* has a LDH that has the hydride transfer step as its rate-limiting step [19]. This means that despite their structural similarities, their respective specific mechanisms, like the rate-limiting step(s), may vary. Further studies from our lab have confirmed that Trp107, which sits on the specificity loop in apicomplexan LDHs (Figure 1.6), is largely responsible for its LDH activity [6]. This residue is hypothesized to have two main characteristics that confer its role: the hydrophobicity contributes to binding through packing against the C3 methyl group of pyruvate, and the bulky ring system excludes larger substrates like oxaloacetate in the active site [6]. The mechanism employed by MDHs, on the other hand, to discriminate against the small, less negatively charged pyruvate is also electrostatic interactions (charge complementarity), as emphasized previously. In porcine cytoplasmic MDH, there are two Arginine residues (91 and 97) on the active site loop, such that the loop can only close properly if there is a substrate with two negative charges [17]. This is exactly the property of oxaloacetate.
Figure 1.6: Trp107 in the active site of PfLDH against oxalate. Behind the oxalate in the figure is NADH, shown as sticks. Trp107 clashes into the methyl group of pyruvate, helping it to recognize the small, uncharged substrate with an uncharged methyl group.

1.2 The apicomplexan LDHs are divided in groups based on phylogenetic studies

To better understand the impact on functions from differences in structures, horizontal comparisons, including alignments and structure superpositions as well as functional comparisons, have been employed for a long time. Recently, ancestral reconstruction methodology has become increasingly popular. Just as one would be confused by the recurrent laryn-
geal nerve on the Vagus nerve in our body if he does not understand its development, the understanding of the history and origins of modern enzymes are critical for understanding their current forms and functions [20]. Another reason is that horizontal approaches in many cases were unsuccessful, due to the massive mutations accumulated between modern proteins [21]. We now know that proteins can gain novel functions in the process of duplication by mutations [22], so ancestral reconstruction methods, in this respect, can shed new light on the structure/sequence-function relationships and the order of the mutations. From there we can learn more about the interactions between the mutations, as well as highlighting the mutations that truly contributed to their activity profiles [21]. Even though quantitative characterizations of resurrected ancestral proteins may not be very accurate depictions of the real ancestor, qualitative conclusions drawn from such studies are robust [23]. The Theobald Lab took on this approach to analyze the history of MDH/LDH superfamily.

Previous efforts in this lab generated a phylogeny based on extant LDH and MDH sequences using maximum likelihood [6]. Figure 1.7 includes all apicomplexan LDHs (except for Cryptosporidium parvum LDH, and they can all be traced back to their most recent common ancestor at node 356 (AncLDH), which is also one of the events where specific LDH activity was gained from the evolution of MDHs. However, these apicomplexan LDHs exhibit different specificity profiles, especially between PfLDH and EmLDH (labeled as LDH_EIMA in Figure 1.7), who took separate paths in evolution immediately after LDH activity was acquired at node 356.
Figure 1.7: A clade from a much larger phylogenetic tree. This clade includes the most recent common ancestor of \textit{PfLDH} (LDH\textsubscript{PLFA}) and \textit{EmLDH} (LDH\textsubscript{EIMA}). Branch lengths are not specified in this figure. The two MDH labels in black are most likely be the result of mislabeling from the database, as these two enzymes both have a lysine at the 102 position, has the five-residue insertion in the specificity loop, and a tryptophan is found on the specificity loop. The common ancestor of all the modern apicomplexan LDHs (except for \textit{CpLDH}) shown in this figure is the node on the left side of the figure, node 356 (AncLDH). AncLDH is also one of the independent events where pure LDH activity was first obtained. In this case, AncLDH is the most recent common ancestor to all the apicomplexan LDHs shown in this clade.
1.3 LDHs in *Plasmodium falciparum* and *Eimeria maxima*

LDH found in *Plasmodium falciparum* (*PfLDH*) has been studied extensively since it has been identified as an attractive drug target for malaria. *PfLDH* is a ∼140 kD homotetramer (Figure 1.1) with four identical subunits that are each 35 kD, with their respective active sites, and no cooperativity has been found between the subunits [13,14]. Despite only sharing moderate sequence identity with the canonical (like mammalian) LDHs, it shares the same Rossmann fold topology with the canonical LDHs [9]. There are, however, some remarkable differences that demarcate *PfLDH* as an apicomplexan LDH.

One of the relatively well-studied regions is the specificity loop in the active site. There is a five-residue insertion in the loop adjacent to a highly-conserved arginine residue compared to canonical LDHs [9,14]. Both *PfLDH* and *EmLDH*, along with other apicomplexan LDHs from this clade, have this five-residue insertion, but not Dogfish LDH, the one used for the numbering system (see Appendix 6.1), or other “canonical LDHs”. Another difference is residue 102, as discussed earlier. In canonical LDHs, Gln102 is in the active site. In *PfLDH*, however, Lys102 is excluded from the active site, unlike the corresponding position (Arg102) in MDHs [6]. These differences render the unique activity profile of *PfLDH*. It is because of these differences, both structurally and kinetically, that *PfLDH* became a potential antimalarial drug target, and thus has been studied extensively.

Not all apicomplexan LDHs, however, has been studied to such an extent. *Eimeria maxima* is a closely-related parasite to *Plasmodium falciparum* in the Apicomplexa phylum, and it is only known to infect poultry [24]. The LDH from this organism (*EmLDH*) has not been extensively studied yet, but it is closely related to all the apicomplexan LDHs in this
CHAPTER 1. INTRODUCTION

clade. *Em*LDH is labeled as LDH_EIMA in a phylogeny calculated in this lab [6], shown in Figure 1.7. *Pf*LDH and *Em*LDH can be traced back to their most recent common ancestor at node 356 (AncLDH) in Figure 1.7. Despite *Em*LDH sharing many distinct apicomplexan LDH characteristics with *Pf*LDH, including the specificity loop insertion and the specificity residue Trp107, there are still plenty of differences among the two (sequence identity = 52%). This makes sense, because immediately after the LDH activity was first gained at the AncLDH, *Pf*LDH and *Em*LDH took separate paths to evolve, and thus the differences can be attributed to their evolutionary divergence. One of the conspicuous differences is the insertion/deletion on the opposing loop: there is a three-residue deletion (An Ala instead of a Gly-Gln-Gly) in all *Plasmodium* LDHs in this loop (Figure 1.8), and this loop is in its longer form in all other modern apicomplexan LDHs as well as the AncLDH. This loop has been studied in this lab, and it has been shown to have significance in terms of catalysis [25]. The activity profiles of *Em*LDH and *Pf*LDH also differ.
Figure 1.8: The opposing loop insertion/deletion. The structures that have the longer loop are: LDH_TOGO, LDH_TOXGO, LDH_EIMA, and the shorter loop is from the structures of LDH_PLFA, LDH_PLGB, and LDH_PLVI. AncLDH also has the longer version of this loop.

An thorough study of EmLDH and comparison of this enzyme to other known apicomplexan LDHs on the phylogeny can potentially impart new knowledge of both the modern apicomplexan LDHs and their ancestors. The information of their ancestors can, in turn, inform us about the modern apicomplexan LDHs, including PfLDH, and can thus assist the LDH-based drug design. It can hopefully answer the question of the origin of the functional divergence of apicomplexan LDHs.
1.4 Goal of the thesis

In my thesis project, the main goal is to try to understand the origin of the functional divergence of apicomplexan LDHs. I first attempted to understand the enzyme EmLDH, and then compared it to the modern PfLDH, who took a separate path of evolution immediately after AncLDH acquired pure LDH activity. I explored the kinetics and specificity profile for PfLDH, EmLDH, and AncLDH, and attempted to understand the structures’ implications on its function. To obtain comprehensive structural information, I solved the first crystal structure of EmLDH, as this enzyme has yet to be studied in detail. I first characterized the enzyme’s activity profile using steady-state kinetics (Section 2.4) with a few probing substrates. Structural information was obtained by X-ray crystallography (Section 2.6 and Section 2.7). I then tried to find the differences between EmLDH and PfLDH that are close to the active site. These residue differences include all the different residues within 6 Å of the active site, as well as the opposing loop. Mutants were then made in hope to test if these mutations matter, using steady-state kinetics. Unfortunately I have not been able to successfully express or purify the mutants yet. To get a deeper understanding of EmLDH, I also performed pre-steady-state kinetics experiments to study its rate-limiting step. With this information, future studies can focus on narrowing down the origin of the mechanistic differences between these modern apicomplexan LDHs, characterizing their ancestors, and gain more insights into the source of the divergence in apicomplexan LDHs’ functions and structures.
1.5 Experimental background

1.5.1 Steady-state kinetics

In this investigation, one of the primary means of characterizing the activity profile is through steady-state kinetics, which can not only give information about the ability for an enzyme to catalyze its own reaction, but it can also give useful information about its specificity, if other alternative substrates/substrate analogs are used as chemical probes. The parameters that falls out of a round of kinetic assay after calculations are $k_{\text{cat}}$ and $K_M$. These parameters are extracted from the Michaelis-Mention equation:

$$v = \frac{V_{\text{max}}[S]}{K_M + [S]} \quad (1.1)$$

$$= \frac{k_{\text{cat}}[E][S]}{K_M + S} \quad (1.2)$$

$$\frac{v}{[E]} = \frac{k_{\text{cat}}[S]}{K_M + [S]} \quad (1.3)$$

In Equation 1.1, $v$ is the reaction rate, $V_{\text{max}}$ is the maximum rate that can be reached by a given amount of enzyme, $[E]$ is the total enzyme put into the reaction, $[S]$ is the substrate concentration, and $K_M$ is the Michaelis constant, which is also equal to the substrate concentration when the reaction rate is at half maximum. $k_{\text{cat}}$ can later be extracted from the Michaelis-Mention equation, as shown in Equation 1.3.

Another useful parameter is the specificity constant, $k_{\text{cat}}/K_M$. It has a unit of $M^{-1}s^{-1}$, and is a second-order rate constant, which summarizes the efficiency of all the steps in catalysis. In our study, this very parameter is also interpreted as the specificity constant. The ratio of $k_{\text{cat}}/K_M$ across substrates gives direct and quantitative information about the
preference of the enzyme.

To get a full picture of the activity profile, multiple substrates were used to probe the enzyme (see Section 2.4 for details).

1.5.2 Steady-state approximation

Steady-state kinetics assumes two key players, the enzyme and the substrate. The reaction is usually second-order. For lactate/malate dehydrogenases, however, in addition to the enzyme itself and the substrate, a cofactor, nicotinamide adenine dinucleotide hydride (NADH/NAD$^+$) is also needed to provide/accept a hydride. To address this problem, one reactant (NADH) is taken out of the picture by saturating the enzyme with it. The $K_M$ of PfLDH for NADH is about $11\mu M$ [18]. By using a consistent 200 $\mu M$ NADH concentration (20 times above its $K_M$), we can be reasonably sure that the rate of the reaction is solely dependent on the substrate concentration.

1.5.3 Pre-steady-state kinetics

Steady-state kinetics experiments can give useful overall activity information about an enzyme. But in order to understand the elementary steps in catalysis, we need to force the enzyme out of its steady-state by using a non-catalytic amount of the catalyst. In a stopped-flow instrument, the first turnover event can be captured due to the short dead time of the instrument.

In a single turnover experiment, the enzyme concentration is set to be equal to the NADH concentration, while the substrate is still at saturating levels. This ensures single turnovers, and the rate of the steps including from pyruvate binding up to the hydride transfer (Figure 1.5) is measured because that is where the absorbance signal at 340nm is lost. Because of
the stochastic nature of the system, the absorption signal would decrease in an exponential fashion:

\[ A = A_0 e^{-\mu t} \]  
\[ = A_0 e^{-\frac{t}{\tau}} \]  

Here \( \mu \) in Equation 1.4 is the Poisson parameter, which is also the rate constant for the steps that the stopped-flow instrument measures, until the absorption signal is lost. \( \tau \) is the time constant, which is the reciprocal of the Poisson parameter, and is the average lifetime of the starting species, in this case, the average lifetime of the NADH. In theory, if the rate constant for this step is much larger than \( k_{cat} \) measured from the steady-state experiments, then the rate-limiting step does not lie in anywhere from binding of the substrate to the completion of the hydride transfer on the scheme. If the rate constant here is equal to \( k_{cat} \), then we would know that the rate-limiting step is either the binding of the substrate, loop closure, or the hydride transfer. The substrate can then be deuterated to narrow down if the chemical step is rate-limiting by exploiting kinetic isotope effect [18].

In a multiple turnover experiment, there could be up to three phases: a initial burst phase (this may not present if the rate-limiting step lies between pyruvate binding and hydride transfer), a steady-state phase (a few turnover events), and eventually a phases where all the substrate/cofactor is burnt out. Under this condition, the magnitude of the burst phase, if there is one, would be magnified by the non-catalytic amount enzyme. Excluding the last phase, if there is a burst, the shape of the absorption curve should look like the sum of a single exponential function and a linear function:
CHAPTER 1. INTRODUCTION

\[ A = A_0 e^{-kt} + m \cdot t + C \] (1.6)

The presence of burst phase suggests that the first reaction is faster than the steady-state rate. Depending on the system, the instrument could be measuring different part of the mechanism. In our case, the burst phase is indicative of a faster rate of the steps from pyruvate binding to the hydride transfer step (Figure 1.5).

1.5.4 X-Ray crystallography

One of the richest pieces of information about a protein is its structure, and one of the most popular way of obtaining such data is through X-Ray crystallography. By shooting a single crystal lattice made of protein and solvent with an X-Ray beam and Fourier transformations, an electron-density map could be produced. In order to perform such calculation, two pieces of prior information are required: the wave phase and wave intensity. The wave intensity can be directly determined through the experiment. The phase problem is solved by molecular replacement using a known, similar structure. In our case, since there is no known structure of *Em*LDH, a homology model was used instead. The homology models are built based on the sequence of our protein and the knowledge of known structures who have similar sequences. See Section 2.7 for details.
Chapter 2

Material and methods

2.1 Plasmid construction and mutation

Protein sequences, after adding 6 histidine tags on the C-terminus, were codon-optimized for expression in *Escherichia coli*. The coding sequence of *Eimeria maxima* LDH was subcloned into pET-24a. *Plasmodium falciparum* LDH was subcloned into pET-11b. In all cases, the N-terminal T7-histags were bypassed, and 6 histidines were added to the C-terminal instead. All gene syntheses and subcloning was performed by Genscript (Piscataway, NJ).

2.2 Protein expression

Plasmids were transformed into BL21 DE (pLysS) *E. coli* cells (Invitrogen, Grand Island, NY). Cells were then grown at 37°C with 225 RPM agitation in 2xYT media, supplemented with 30mM potassium phosphate, pH7.8 and 0.1% (wt/vol) glucose. The cells were induced with 0.5mM IPTG after OD$_{600}$ were between 0.5 and 0.8, and the expressed for 4 hours. The cells were separated from the media by centrifugation at 10,000×*g* for 15 minutes and stored
CHAPTER 2. MATERIAL AND METHODS

at -80°C.

2.3 Protein purification

Lysis buffer (50mM NaH$_2$PO$_4$, pH8.0, 300mM NaCl, 10mM Imidazole) was added to the cell pellets, and the pellets were thawed on ice, releasing lysozyme produced by the pLysS plasmid within BL-21. 2 µL of Pierce Universal Nuclease (Thermo Scientific, Rockford, IL) per 1.5L culture was also added. Once thawed the pellets were homogeneously resuspended, and the lysate was sonicated on ice at 35% amplitude (30s ON, 20s OFF, 2 minutes in total, Fisher Scientific Sonic Dismembrator Model 500). Cell debris was then separated by centrifugation at 28,000×g for 20 minutes.

The supernatant was 0.22µm filtered and was then applied to affinity chromatography on a 5ml HisTrap FF nickel column (GE Healthcare, Piscataway, NJ), and eluted via imidazole gradient from 10mM to 500mM on an AKTA Prime (GE Healthcare, Piscataway, NJ). Fractions were analyzed by SDS-PAGE, pooled, and then concentrated using Amicon Ultracel-10K centrifugal filters (Millipore, Billerica, MA). The proteins were then buffer exchanged into 50mM Tris, pH7.4, 100mM NaCl, 0.1mM EDTA and 0.01% azide by PD10 columns (GE Healthcare, Piscataway, NJ). Enzyme concentrations were determined by absorbance at 280nm, using extinction coefficients and molecular weights calculated by ExPASy’s ProtParam tool (http://web.expasy.org/protparam/).

2.4 Steady-state kinetic assays

Enzymatic reduction of substrates was monitored at 25°C by following the decrease in absorbance at 340nm due to NADH oxidation on a Cary 100Bio (Agilent, Santa Clara, CA)
in 50mM Tris pH7.5, 50mM KCl. The substrates used in the assays are shown in Figure 2.1. These include pyruvate, oxaloacetate, \( \alpha \)-ketovalerate (2-oxopentanoic acid), isocaproatate (4-methyl-2-oxopentanoate), and phenylpyruvate (2-oxo-3-phenylpropanoate). Isocaproatate was purchased from Santa Cruz Biotechnology (Dallas, TX), and all the other substrates as well as NADH was purchased from Sigma Aldrich (St. Louis, MO).

Kinetic parameters were estimated by chi-squared fitting to either the Michaelis-Menten equation (Equation 1.3) or substrate inhibition equation (Equation 2.1) [6] in most cases.

\[
\frac{v}{[E]_T} = \frac{k_{cat}[S]}{K_M + [S] + [S]^2/K_i}
\]  

(2.1)

A “phi-fit” equation (Equation 2.2) is occasionally used when the enzyme cannot reach the substrate saturation point (unmeasurably high \( K_M \)) [6]:

\[
\frac{v}{[E]_T} = \frac{k_{cat}}{K_M} \cdot [S] \cdot \frac{1}{1 + S/K_M}
\]  

(2.2)

\[
\frac{v}{[E]_T} = \frac{k_{cat}}{K_M} \cdot [S] \cdot 1
\]  

(2.3)

At concentrations much lower than \( K_M \) ([S] \ll K_M), Equation 2.3 holds, and the resulting graph is approximately linear, and the slope can be taken as the specificity constant.

All data fits were performed using the KaleidaGraph software.

Because aqueous oxaloacetate spontaneously decarboxylates to pyruvate at 25°С and neutral pH at a rate of 3 \times 10^{-5} \text{s}^{-1} [26], oxaloacetate stocks contain pyruvate contamination (approximately 1-3% from Sigma-Aldrich, depending on batch), and must be used with caution. All solubilized oxaloacetate stocks were made fresh before assay, kept on ice throughout the assay, and the stocks had to be remade for every two hours. For enzymes with low activity towards pyruvate relative to oxaloacetate, the contamination has negligible
Figure 2.1: Substrate and alternative substrates used in the kinetic assays. The differences between the substrates are highlighted in red.

effects. For modern LDHs, however, artifactual oxaloacetate activity might be recorded due to the contamination. To resolve the issue, high enzyme concentrations were used to burst through the pyruvate contamination before the enzymes achieve steady-states with oxaloacetate. The result is a biphasic $\Delta A_{340}$, with bursts in early phases, followed by a linear phase. The linear phases were used to characterize the oxaloacetate activity.

2.5 Multiple turnover experiment

70$\mu M$ of enzyme was premixed with 400$\mu M$ of NADH. It is then injected into the chamber with pyruvate, so that in the chamber the pyruvate concentration was 10,000$\mu M$, while the enzyme concentration was 35$\mu M$ and the NADH concentration was 200$\mu M$. The NADH absorption at 340$nm$ was monitored and recorded (Applied Photophysics SX20 Stopped Flow Spectrometer).
2.6 Protein crystallization

Crystallization trials were performed using hanging-drop vapor-diffusion at room temperature. The crystallization conditions were taken from Crystal Screen and Crystal Screen 2 from Hampton Research (Aliso Viejo, CA). Crystals of EmLDH in complex with oxamate and NADH were grown at room temperature by hanging-drop vapor-diffusion with drops of 4µL of 1:1 precipitating buffer:protein. The crystal rendered the best solution was grown at from 0.1M HEPES, pH7.5, and 2.0M ammonium formate.

The crystals were cryoprotected with 30% (wt/vol) dextrose/reservoir solution [6]. The crystals were harvested from the drop, soaked in the 15% solution, immediately transferred to the 30% solution, and flash-frozen immediately in liquid N₂.

2.7 Structure determination and analysis

All crystal diffraction datasets were collected at SIBYLS beamline (12.3.1, Lawrence Berkeley National Laboratory, Berkeley, CA). The datasets were indexed and merged with XDS [27]. After initial failures in molecular replacements, the space group was manually changed to P3₁ 2 1, and the dataset was reindexed accordingly. Structures were solved by molecular replacement using Phaser-MR (full featured) in PHENIX [28]. Homology models were generated from sequence by the Phyre2 server [29]. The solutions were then further refined using phenix.refine in PHENIX, and manual improvements on the model were done in Coot [30]. The qualities of the models were validated with MolProbity [31,32] in PHENIX.

In the solved structure of EmLDH, the active site is not fully closed, and thus has a more disordered loop (with high B-factors). The electron density in the region is poor and insufficient to support the structure of the entire loop. Consequently, 4 residues are missing
from the loop in the structure. See Table 6.4 in Appendix 6.3 for more statistics regarding the crystal structure.

The superpositions were generated using THESEUS [33]. The structure and superposition images were rendered with PyMOL. The 6Å shell around the active site was created in PyMOL, with the oxamate/oxalate as the center of the active site.

2.7.1 Sequence alignment

The structural analysis in this project is performed on EmLDH (no PDB ID yet), PfLDH (PDB ID: 1T2D), and AncLDH (PDB ID: 4PLG). To be consistent with the field, I attempted to approach the system from a sequence alignment of all these sequence with the standard dogfish LDH (PDB ID: 3LDH) [9]. However, numerous inconsistencies between the sequence alignment and the structural superposition and the seemingly random numbering system used in 3LDH itself made the task extremely confusing. So alternatively, I first superposed the structures of 1LDG, which has already adapted the standard dogfish LDH numbering system [34], with 1T2D. I then found the residues that I wished to emphasize on PfLDH from 1T2D, and then extracted the number for that specific residue from 1LDG. Because these two structures are essentially the same thing, they superposed almost perfectly. The numbers on these residues of interest are then horizontally propagated to EmLDH and AncLDH to give the information in Table 3.2.

To avoid future confusions if one were to use the result from this text, I made a new sequence alignment that took into account of the seemingly random deletions and insertions of the numbers in the original dogfish LDH, and also integrated structural alignment information into this sequence alignment. An initial alignment that was based on structural superposition was generated by THESEUS using 1LDH, 3LDH, and EmLDH. The align-
ment was then manually corrected according to the residue numbers in 1LDG and 3LDH. The numbers in \textit{Em}LDH was then manually corrected according to the superposition. Chain A in 1LDG, chain A in 3LDH, and chain B in \textit{Em}LDH were used for superposition.
Chapter 3

Results

3.1 Kinetics

The kinetics of $PfLDH$, $EmLDH$, and AncLDH were assayed at steady-states, turning over pyruvate to lactate, at an excess NADH concentration. The result is summarized in Table 3.1.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (M$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$EmLDH$</td>
<td>7.39</td>
<td>57.16</td>
<td>$1.29 \times 10^5 \pm 1.12 \times 10^4$</td>
</tr>
<tr>
<td>$PfLDH$</td>
<td>91.05</td>
<td>72.20</td>
<td>$1.26 \times 10^6 \pm 1.60 \times 10^5$</td>
</tr>
<tr>
<td>AncLDH</td>
<td>44.76</td>
<td>106.38</td>
<td>$4.21 \times 10^5 \pm 6.80 \times 10^4$</td>
</tr>
</tbody>
</table>

Table 3.1: Steady-state kinetics of $PfLDH$, $EmLDH$, and AncLDH turning over pyruvate.

$k_{cat}$ is the turnover number, or the maximum theoretical number of reactions catalyzed per active site per unit time when the enzyme is saturated. It can reflect the maximum velocity an enzyme can reach. The AncLDH has a $k_{cat}$ of 44.76 s$^{-1}$, but as the enzymes
evolve, PfLDH gained higher maximum velocity, while EmLDH lost some of the maximum velocity.

$K_M$, on the other hand, can reflect the affinity of an enzyme toward a substrate. Less drastic changes were found in $K_M$ between the modern LDHs and the ancestral LDH. Both EmLDH and PfLDH have lower $K_M$, meaning they both gained higher affinity toward pyruvate. However, the changes here are limited to no larger than two fold.

The enzymatic efficiencies, measured in $k_{cat}/K_M$, between the three LDHs, vary only within about one order of magnitude, with the highest efficiency found in PfLDH and the lowest efficiency found in EmLDH. AncLDH has an enzymatic efficiency in between the two, suggesting functional divergence after AncLDH. Nevertheless, all three LDHs have decent LDH function.

### 3.2 Substrate specificity profiles

EmLDH, PfLDH, and AncLDH were also assayed using four other substrates (see Section 2.4) to probe their specificity. While pyruvate being the native substrate of all LDHs, the activities of LDHs are not limited to pyruvate, but just at a relatively lower efficiency toward other substrates. This allows us to assess and probe the active site with various substrates with different properties (see Figure 2.1).

All four alternative substrates are $\alpha$-ketoacids, with varying tail ends, highlighted in red in Figure 2.1. The alternative substrate that differs the most from the native substrate is oxaloacetate, due to its extra charge as well as its size. I used three additional substrates that are not involved in either the citric cycle pathway or anaerobic respiration pathway as shape sensors. They have subtle structural characteristics that put them in between pyruvate and oxaloacetate. While all have one negative charge at neutral pH, isocaproate mimics the
shape of oxaloacetate in terms of its branched structure on the opposite end of the molecule. However, a key difference, in addition to the lack of one charge, should be noted. While oxaloacetate has a planar $\pi$ system (the carboxylate) with a $sp^2$ hybridized carbon at the center, isocaproate has a $sp^3$ hybridized carbon at the equivalent position. As a result, the skeleton of isocaproate cannot adopt a planar conformation. $\alpha$-ketovalerate is a much longer, and thus more hydrophobic, version of pyruvate, which has two more carbons added to the tail. Phenylpyruvate mimics the $\pi$ system of oxaloacetate well. This substrate, however, differs in size with oxaloacetate. Together with its lack of the extra charge, phenylpyruvate is a highly hydrophobic molecule, with very limited solubility in aqueous solutions. The specificity profile constructed from the entire set of the five substrates can be informative on the mechanism the enzymes employ to achieve substrate specificity.

The data obtained at the steady-states of the three enzymes on these substrates is summarized in Figure 3.1. All three enzymes catalyze the reduction of their native substrate, pyruvate, with appreciable efficiencies, with the highest efficiency belong to PfLDH. Their activity similarities can be extended to oxaloacetate as well: all three are poor catalysts for the reduction of oxaloacetate. When it comes to substrates that are bigger and with only one negative charge, however, the specificity profile starts to differ. PfLDH still has very low activities, which suggests it has extremely high specificity toward pyruvate. The same conclusion, however, cannot be drawn for either the common ancestor or EmLDH, because despite their lack of activity toward oxaloacetate, they are able to turn over big, more hydrophobic substrates much better than PfLDH. It seems like even though isocaproate has a similar shape to oxaloacetate, as soon as the charge is disposed of, AncLDH and EmLDH are able to gain some activities back. PfLDH, on the other hand, barely has any detectable activity toward isocaproate. Because its native substrate is pyruvate, the absence of the
extra negative charge cannot be the reason of this further reduction in activity compared to oxaloacetate. This leaves the explanation only to the shape and the size of the molecule. All three LDHs have an increased activity toward α-ketovalerate, but none has an efficiency that reach $10^4 \text{ M}^{-1}\text{s}^{-1}$. The barriers for all three enzymes turning over isocaproate and α-ketovalerate is a combination of reduced theoretic maximum velocity that the enzyme can reach (lowered $k_{cat}$ value) and reduced affinity (higher $K_M$ value).

The result of phenylpyruvate was a surprise, as I was not expecting any of the LDHs with the calibre to turn it over at such amazing speed. *Pf*LDH has the lowest activity toward phenylpyruvate. It appears that the big phenyl ring does not bother *Em*LDH and AncLDH much, if at all, because their catalytic efficiency toward phenylpyruvate are almost at the same level as toward pyruvate. The evidence suggests that the major barrier for *Em*LDH and AncLDH for turning over phenylpyruvate is the lowered binding affinity, reflected through an elevated $K_M$. However, $K_M$ for this substrate is still much lower than isocaproate and α-ketovalerate, suggesting the significance of the π system on the substrate. The major barrier for *Pf*LDH to turn over phenylpyruvate is a combination of maximum velocity and its affinity. The hypothesis for the decent catalytic efficiency toward phenylpyruvate seen in *Em*LDH and AncLDH is that Trp107, the specificity residue for apicomplexan LDHs in this clade, who normally contacts the methyl group on pyruvate to achieve specificity toward the small molecule, could have π stacking interactions with the phenyl ring at the equivalent position on phenylpyruvate, which compensates for the lowered binding affinity caused by the size of the molecule. However, this hypothesis would not suffice for *Pf*LDH, who uses the same specificity residue for substrate recognition. So if this hypothesis is true, *Pf*LDH must have gained other new mechanisms for substrate recognition.
Figure 3.1: Results from steady-state kinetic assays. The catalytic efficiencies were measured in $k_{cat}/K_M$. The data was plotted on a logarithmic scale. Note: *Pf*LDH has been assayed with isocaproate, but no measurable rate could be obtained. The structure of each substrate can be seen in Figure 2.1. The original data can be found in Appendix 6.2.

Because the main player in the system is LDH, which has very low activity toward oxaloacetate, in some oxaloacetate kinetic assays I could not obtain an accurate measurement of either $k_{cat}$ or $K_M$. *Em*LDH exhibits a behavior such that the rate keeps increasing as the substrate concentration. So instead of a regular rectangular hyperbola, a straight line is observed in the titration curve, with an invariant slope throughout the titration. In other words, even at abnormally high substrate concentration, the enzyme is still not half-saturated. In this case, the low substrate version of the Michaelis-Menten equation was used (Equation 2.3) [6]. This equation provides $k_{cat}/K_M$ as one parameter with a reasonable
error from the data without having to know the exact value of $k_{\text{cat}}$ and $K_M$ respectively. No reliable data can be obtained for PfLDH when isocaproate was used as the probe, as the efficiency/affinity is extremely low. At very high enzyme concentration ($0.57\mu M$) and high substrate concentration ($10mM$), the rate of the reaction is around the background noise detected by the instrument.

3.3 PfLDH diverged from the ancestral LDH

One of the advantages of analyzing enzymes using ancestral sequence reconstruction is that we can get a sense of the source of divergence that we see in modern enzymes. In our case, I was able probe both the modern enzymes and their most recent common ancestor with the five substrates. Figure 3.1 shows a closer activity profiles shared between the ancestor and EmLDH.

In terms of the native substrate pyruvate, EmLDH has the lowest efficiency among the three enzyme tested. It is about an order of magnitude lower than PfLDH, and AncLDH is in between the two. For bigger, and still hydrophobic, substrates, EmLDH exhibits much more tolerance than PfLDH, which essentially refused to catalyze the branched isocaproate at all, and only has minor activity toward the unbranched $\alpha$-ketovalerate ($k_{\text{cat}}/K_M$ less than $100M^{-1}s^{-1}$). There is a general trend, for most of the LDHs tested in this lab, that the straight-chained $\alpha$-ketovalerate is a better substrate than the branched isocaproate. The mechanism is not clear at this moment, but it is likely that this can be attributed to the shape of the active site, or the higher degree of freedom of rotation on the tail of $\alpha$-ketovalerate. Similar to EmLDH, the ancestral LDH still has appreciable activities toward large, uncharged substrates ($k_{\text{cat}}/K_M$ close to $10^3M^{-1}s^{-1}$). This result, combined with the structural analyses in Section 3.4, implies that EmLDH inherited much of the specificity
profile from this ancestor, while PfLDH evolved high efficiency and specificity that the ancestor did not possess.

3.4 Structural information from X-Ray crystallography

To obtain structural information, I obtained the first X-ray crystal structure of EmLDH (Section 2.6 and Section 2.7). Figure 3.2 is a superposition, generated using THESEUS, of all the known structures of apicomplexan LDHs in the clade shown in Figure 1.7 with known structures. As expected, the overall topology/fold of this enzyme is the same as other LDHs, as shown in Figure 3.2, with the EmLDH shown in blue. Detailed statistics about this solution is given in Appendix 6.3.

In the structure that I solved and refined, there are four residues on the specificity loop lacked the support of electron density due to its high B-factor in the non-fully closed state. However, the important Trp107 is resolved. It is very close to the Trp107 seen in PfLDH in space. The slight displacement is likely due to the fact that the loop is not as fully closed as the structure 1T2D. As expected from the sequence alignment, EmLDH also has the longer opposing loop, as opposed to the other study subject in this project, PfLDH. The opposing loop aligns well with other known apicomplexan LDH structures in this clade that have the GQG residues.

Using PyMOL, I generated the asymmetric units around the solved dimer structure, and it can be seen clearly that the enzyme is in fact a tetramer, as expected. The superposition was then used to find out the differences among these structurally very similar enzymes.
Figure 3.2: Superposition, generated from THESEUS, of EmLDH (in blue), PfLDH (PDB ID: 1T2D, in yellow), and other apicomplexan LDHs (1OC4, 2A92, 1PZH, and 1SOW) with known structures. The inhibitors, either oxamate or oxalate, and the cofactor NADH are shown as sticks in the active site. The monomers were taken and superposed in THESEUS. The overall topology of EmLDH, as expected, is very similar to the other known apicomplexan LDHs, according to the superposition. Details about this structure is given in Appendix 6.3.

3.5 Structural comparison

After the crystal structure was solved, I compared the structure of EmLDH with PfLDH. As the two proteins have only 52% sequence identity, exhaustive studies on all divergent
residues is not realistic. It would be very useful to adopt a phylogenetic approach to study the origin of the two enzyme's two different behaviors, especially if the origin is from epistatic interactions. But before diving into the more complicated analyses and looking for epistasis, it is reasonable to do a simpler study first. As a result, I focused on the direct comparison at this stage, and found the differences within the 6Å shell within the active site (defined here as the reacting $C_0$ on the oxamate or oxalate) of EmLDH, PfLDH, and AncLDH.

Within 6Å of the active site, there are 6 differences between PfLDH and EmLDH, and all of these 6 different residues are confined between PfLDH and EmLDH (Table 3.2). In other words, within the 6Å shell, EmLDH and AncLDH are identical. This is especially interesting, given the fact that PfLDH and AncLDH share 63% sequence identity, while EmLDH and AncLDH share a marginally higher, 66%, sequence identity. This implies that despite the similar amounts of mutations were accumulated in PfLDH and EmLDH since AncLDH, all the mutation in EmLDH are away from the active site. But this is not the case for PfLDH.
Table 3.2: Different residues that are close to the active site among *Em*LDH, *Pf*LDH, and AncLDH. These are picked out from a superposition of the three proteins’ crystal structures. These are divided into three groups: one group (6Å Shell) contains all the mutations that are within 6Å of the active site. The second group, the “Specificity Loop”, contains amino acid differences on the specificity loop, while the third group, the “Opposing Loop”, includes the residue differences on the opposing loop. Both AncLDH and *Em*LDH have a longer opposing loop than *Pf*LDH. It is likely that the origin of *Pf*LDH’s functional divergence can be found in this list. The gray boxes emphasizes on the residues that are different between *Em*LDH and *Pf*LDH that are considered as important.

<table>
<thead>
<tr>
<th>Region</th>
<th>LDH_PLFA</th>
<th>LDH_EIMA</th>
<th>AncLDH</th>
</tr>
</thead>
<tbody>
<tr>
<td>6Å Shell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T235</td>
<td>A235</td>
<td>A235</td>
<td></td>
</tr>
<tr>
<td>A236</td>
<td>G236</td>
<td>G236</td>
<td></td>
</tr>
<tr>
<td>L237</td>
<td>G237</td>
<td>G237</td>
<td></td>
</tr>
<tr>
<td>P246</td>
<td>A246</td>
<td>A246</td>
<td></td>
</tr>
<tr>
<td>V142</td>
<td>L142</td>
<td>L142</td>
<td></td>
</tr>
<tr>
<td>L163</td>
<td>M163</td>
<td>M163</td>
<td></td>
</tr>
<tr>
<td>Specificity Loop</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V96</td>
<td>I96</td>
<td>I96</td>
<td></td>
</tr>
<tr>
<td>F100</td>
<td>I100</td>
<td>L100</td>
<td></td>
</tr>
<tr>
<td>A103a</td>
<td>I103a</td>
<td>A103a</td>
<td></td>
</tr>
<tr>
<td>N108</td>
<td>S108</td>
<td>S108</td>
<td></td>
</tr>
<tr>
<td>Opposing Loop</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H243</td>
<td>L243a</td>
<td>L243a</td>
<td></td>
</tr>
<tr>
<td>G243b</td>
<td>G243b</td>
<td>G243b</td>
<td></td>
</tr>
<tr>
<td>Q243c</td>
<td>Q243c</td>
<td>Q243c</td>
<td></td>
</tr>
<tr>
<td>A244</td>
<td>G244</td>
<td>G244</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2: *Em*LDH and AncLDH have almost identical environment near the active site, except for the residues on the specificity loop, which was shown to be insignificant as long as the specificity residue Trp107 is left alone [16].

The important residues around the active sites, however, could extend a little further beyond the 6Å shell, because neither the specificity nor the opposing loop resides strictly within the 6Å shell around the active site, and yet both can potentially affect catalysis. Four residues on the specificity loop are different between *Em*LDH and *Pf*LDH, while one three-residue insertion/deletion on the opposing loop was also found. In terms of the insertion/deletion on the opposing loop, it also seems like *Em*LDH and AncLDH share these same residues on this portion, and *Pf*LDH mutated the three residues into an alanine at this position, along with a mutation that precedes the loop residue(s) (Table 3.2). The numbering of all the residues are based on and adapted from canonical dogfish LDH (Section 2.7.1). The sequence alignment used to achieve this is shown in Appendix 6.1.

This list, combined with specificity assays, suggest that in this system, it is likely that residues near the active site and them alone are responsible for preserving functions in the course of evolution, and if the enzyme wished to alter its function, it needed to loosen the restraints on residues that are close to the active site. It is thus likely that the gray portion of Table 3.2 includes all the residues that can explain *Pf*LDH’s functional divergence.

### 3.6 Change it all

Among the different residues that I picked in Table 3.2, I further reduced the number. Previous studies in this lab have shown that mutations on the specificity loop do not have serious impact on the enzyme’s activity, as long as the important Trp107 is there [16]. Based
on this result, it is likely unnecessary to probe the functions of these residues in this region any further. These residues include the first four residues listed in the second group. This leaves all the residues that are within the 6Å shell and on the opposing loop.

To better explore how these residues affect the enzymes functions, I made four mutants in hope that by assessing their activity, I can get a qualitative idea if the specificity profile of these enzymes are dictated by these residues near the active site, or are they influenced by epistasis. What was done to produce each mutant is listed in Table 3.3.

Because it is still in early stages of exploring the properties and behaviors of this enzyme, I swapped all the amino acid residue differences within a close rage (6Å) at once as an attempt to get a qualitative answer of whether these mutations matter. Knowledge derived from the phylogeny can be integrated into later experiments, once the qualitative question is answered.

<table>
<thead>
<tr>
<th>Mutants</th>
<th>6A Shell Differences</th>
<th>Opposing Loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfLDH.6Aswap</td>
<td>Swapped to EmLDH Residues</td>
<td>No Change</td>
</tr>
<tr>
<td>PfLDH.6A++swap</td>
<td>Swapped to EmLDH Residues</td>
<td>Inserted</td>
</tr>
<tr>
<td>EmLDH.6Aswap</td>
<td>Swapped to PfLDH Residues</td>
<td>No change</td>
</tr>
<tr>
<td>EmLDH.6A++swap</td>
<td>Swapped to PfLDH Residues</td>
<td>Deleted</td>
</tr>
</tbody>
</table>

Table 3.3: Mutants were created based on these residue differences. There are four mutants in total. In all four mutants the different residues between PfLDH and EmLDH are replaced by the corresponding residue on the other enzyme, in their own, respective background, as shown in the first column after their name. In the second column it specified what I did to the opposing loop. For specific residue replacements, see Table 3.2.

All four constructs were cloned into plasmids that is optimized for expression in E.coli. However, they were yet to be successfully expressed or characterized.
3.7 Multiple turnover experiment

To gain more insight into the mechanism of EmLDH to compare it with PfLDH, I did a multiple turnover experiment, following the decrease in 340 nm absorption signal at high enzyme concentration. A burst phase was seen.

Figure 3.3: 35 µM of EmLDH was mixed with 200 µM of NADH and 10,000 µM of pyruvate. The decrease in absorption signal at 340 nm was followed. A burst phase and a steady-state were seen, suggesting the product release or NADH binding must be partially rate-limiting.

However, I was not able to fit the burst phase with a single exponential decay well after I subtracted the linear function, so only qualitative conclusions can be drawn from this.
preliminary data. The presence of a burst phase suggests that the rate of the first turnover event is faster than the steady-state rate. In other words, the rate-limiting step likely lies, at least partially, at the product release. The loop opening right before pyruvate release is more likely to be rate-limiting, but further mechanistic experiments are needed to elucidate the rate-limiting step and the micro rate constants.
Chapter 4

Discussion

4.1 Apicomplexan LDHs’ evolutionary divergence

PfLDH has the highest $k_{\text{cat}}$ towards pyruvate among other apicomplexan LDHs, including Toxoplasma gondii LDH 2 [6], Eimeria maxima LDH, and Babesia bovis LDH that are also from this clade. However, this is not a very large difference. EmLDH has a $k_{\text{cat}}$ that is about six fold lower than AncLDH, but its higher affinity helps to compensate for the overall efficiency ($k_{\text{cat}}/K_M$). EmLDH and PfLDH diverged immediately after AncLDH, one achieved slightly higher efficiency (PfLDH) while the other one lost some efficiency (EmLDH). PfLDH and EmLDH also share the characteristic that the rate-limiting step lies at product release or NADH binding. It is more likely that product release or loop opening are rate-limiting, as NADH binding is typically fast [18,35]. Further mechanistic experiments are needed to substantiate this hypothesis.

Specificity profiles constructed from other alternative substrates for PfLDH, EmLDH, and AncLDH, however, suggest that PfLDH acquired new ability to recognize pyruvate with much greater precision. The specificity profile of PfLDH shows very high specificity toward
pyruvate with only minimal activity toward oxaloacetate and other larger, more hydrophobic substrates. It can thus be concluded that \textit{Pf}LDH is sensitive to charge, shape, and size of a substrate, and uses a combination of these to recognize the substrate. This ability was not present in AncLDH. \textit{Em}LDH, in comparison, largely retained the specificity profile from AncLDH. The high catalytic efficiency for phenylpyruvate suggests that AncLDH and \textit{Em}LDH are insensitive to the size of the substrate - as long as the extra charge is not present and the substrate fits a certain shape, the enzymes would be able to catalyze the reaction. However, because \textit{Pf}LDH uses the same Trp107 to contact the methyl group on pyruvate, its low efficiency towards phenylpyruvate implies novel mechanism(s) that prevents substrates with a $\pi$ system for being recognized. The precise mechanism needs to be elucidated with further experiments.

### 4.2 Source of the divergence

In order to discover the source for \textit{Pf}LDH’s functional divergence, structural analysis is necessary. From structural comparison, \textit{Em}LDH and AncLDH have almost identical environments near the active site, except for the specificity loop. The residues of the specificity loop are not important, as long as the Trp107 is present in the proper orientation [16]. This implies reduced evolutionary restrictions on the loop: most mutations in this loop will allow the enzyme to retain its ancestor’s functions/specificity profile. But in places where the restraints are more rigid and close to the active site, \textit{Em}LDH retains a very similar environment and specificity profile as the AncLDH. In these very same regions, \textit{Pf}LDH has a few distinct residues, and this could be the potential source of its functional divergence. It can thus be rationalized that the reason why \textit{Pf}LDH was able to acquire increased substrate recognition capability is that it broke the evolutionary restraints after LDH activity was ac-
quired at AncLDH. The causal relationship proposed in this thesis should be substantiated in future experiments. In any event, with a comprehensive characterization of EmLDH, the study of the functional and structural divergence in this apicomplexan LDH clade, which hitherto has been largely confined between AncLDH and PfLDH and the intermediates between them, can be expanded and even simplified, because now it is possible to pick out the residues that are more likely to be essential for gaining new functions or less likely to make a difference.
Chapter 5

Future Directions and Experiments

The first thing that should be done to continue this project is to try to express and purify the mutants that I made in Section 3.6. Their activity assays can provide insights into whether these residues around the active site matter, and if they do matter, to what extent do they affect the specificity profile.

Moreover, to answer the question posed by the title, it will also be helpful to characterize some other modern apicomplexan LDHs from this clade for the same reason why EmLDH was investigated. The ancestors between the AncLDH and the modern enzymes should also be assessed, so that we could know at which point, with what mutation(s), did PfLDH gain its new ability for substrate recognition.

Only preliminary pre-steady-state kinetics data was collected for EmLDH, so precise information regarding the rate-limiting step or micro rate constants remains unknown. Future experiments including single turnover experiments, multiple turnover experiments, and even single turnover experiments with kinetic isotope effect, should be performed to rule out some steps as the rate-limiting step(s) and draw quantitative conclusions for micro rate constants. These results can then be compared with PfLDH and see if the micro rate constant or
the rate-limiting step could explain the distinct specificity profile, or the higher $k_{cat}$ seen in PfLDH.
Bibliography


Chapter 6

Appendix

6.1 Sequence alignments

Figure 6.1: Sequence alignment of *PfLDH*, *EmLDH*, and their most recent common ancestor, AncLDH. This sequence alignment can inform us on the differences on the sequences, but it fails to convey the actual, spacial relationships between residues on different structures. As a result, another sequence alignment was done to generate the numbering system used in this text.
Figure 6.2: Systematic numbering generated from the standard dogfish LDH. It has been adapted to \(Pf\)LDH by Dunn et al. [9], and I adapted the system onto \(Em\)LDH. 1LDH.A is the A chain in 1LDH (\(Pf\)LDH), 3LDH.A is the A chain of 3LDH (dogfish LDH), and EIMA is the B chain of \(Em\)LDH. See Section 2.7.1 for details.
<table>
<thead>
<tr>
<th>File</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3LDH_A.pdb</td>
<td>LGVHSCLVIGWIGQHGDSVPSVWSGMWDA---K-LHKDVDSAYEVIKLKGITYSWALVVLGVHSCLVIGWIGQHGDSVPSVWSGMWDA---K-LHKDVDSAYEVIKLKGITYSWALV</td>
</tr>
<tr>
<td>1LDG_A.pdb</td>
<td>LNVCPRDVNAHIVGAGHGNKMVLKRYITVGGIPLQEFINKN-LISDAE-LEAIFDRTVNTALLNVCPRDVNAHIVGAGHGNKMVLKRYITVGGIPLQEFINKN-LISDAE-LEAIFDRTVNTAL</td>
</tr>
<tr>
<td>LDH_EIMA.pdb</td>
<td>LHVSPHDVQGMVIGVHDNMLPLMYITINGIPIQEFINKG-LINKED-INNIYNKTKQAGGLHVSPHDVQGMVIGVHDNMLPLMYITINGIPIQEFINKG-LINKED-INNIYNKTKQAG</td>
</tr>
</tbody>
</table>
6.2 Steady-state kinetic data

Note: “N/D” is in some of the cells in the tables in this section. It means no reliable data could be acquired. See Section 3.2 for more explanations.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat} (s^{-1})$</th>
<th>$K_M (\mu M)$</th>
<th>$k_{cat} / K_M (M^{-1}s^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyruvate</td>
<td>7.39</td>
<td>57.16</td>
<td>$129370 \pm 11187$</td>
</tr>
<tr>
<td>Phenylpyruvate</td>
<td>15.62</td>
<td>557.30</td>
<td>$28030 \pm 2998$</td>
</tr>
<tr>
<td>Oxaloacetate</td>
<td>N/D</td>
<td>N/D</td>
<td>38 ± 6</td>
</tr>
<tr>
<td>Isocaproate</td>
<td>0.42</td>
<td>2676.80</td>
<td>$157 \pm 23$</td>
</tr>
<tr>
<td>$\alpha$-ketovalerate</td>
<td>3.07</td>
<td>3178.20</td>
<td>$965 \pm 107$</td>
</tr>
</tbody>
</table>

Table 6.1: Steady-state kinetic assay parameters for $Em$LDH.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat} (s^{-1})$</th>
<th>$K_M (\mu M)$</th>
<th>$k_{cat} / K_M (M^{-1}s^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyruvate</td>
<td>91.05</td>
<td>72.16</td>
<td>$1262000 \pm 160000$</td>
</tr>
<tr>
<td>Phenylpyruvate</td>
<td>0.03</td>
<td>1505.30</td>
<td>$22 \pm 3$</td>
</tr>
<tr>
<td>Oxaloacetate</td>
<td>N/D</td>
<td>N/D</td>
<td>$10.2 \pm 0.4$</td>
</tr>
<tr>
<td>Isocaproate</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
</tr>
<tr>
<td>$\alpha$-ketovalerate</td>
<td>0.14</td>
<td>2358.60</td>
<td>$57 \pm 8$</td>
</tr>
</tbody>
</table>

Table 6.2: Steady-state kinetic assay parameters for $Pf$LDH.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat} (s^{-1})$</th>
<th>$K_M (\mu M)$</th>
<th>$k_{cat} / K_M (M^{-1}s^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyruvate</td>
<td>44.76</td>
<td>106.38</td>
<td>$420820 \pm 68000$</td>
</tr>
<tr>
<td>Phenylpyruvate</td>
<td>51.63</td>
<td>436.37</td>
<td>$118373 \pm 9528$</td>
</tr>
<tr>
<td>Oxaloacetate</td>
<td>0.37</td>
<td>6488.90</td>
<td>$56 \pm 3$</td>
</tr>
<tr>
<td>Isocaproate</td>
<td>1.98</td>
<td>2483.80</td>
<td>$799 \pm 137$</td>
</tr>
<tr>
<td>$\alpha$ -ketovalerate</td>
<td>9.45</td>
<td>3020.90</td>
<td>$3130 \pm 524$</td>
</tr>
</tbody>
</table>

Table 6.3: Steady-state kinetic assay parameters for the common ancestor, or AncLDH. See Figure 1.7.
### 6.3 Crystallographic statistics

<table>
<thead>
<tr>
<th>Name</th>
<th>EIMA_LDH*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range</td>
<td>44.64 - 1.705 (1.766 - 1.705)</td>
</tr>
<tr>
<td>Space group</td>
<td>P 3 1 2 1</td>
</tr>
<tr>
<td>Unit cell</td>
<td>83.4 83.4 227.1 90 90 120</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>100300 (9532)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>0.92</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>30.84</td>
</tr>
<tr>
<td>Reflections used in refinement</td>
<td>92773 (3869)</td>
</tr>
<tr>
<td>Reflections used for R-free</td>
<td>1854 (76)</td>
</tr>
<tr>
<td>R-work</td>
<td>0.2023 (0.4612)</td>
</tr>
<tr>
<td>R-free</td>
<td>0.2214 (0.4333)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>5438</td>
</tr>
<tr>
<td>macromolecules</td>
<td>4958</td>
</tr>
<tr>
<td>ligands</td>
<td>74</td>
</tr>
<tr>
<td>Protein residues</td>
<td>652</td>
</tr>
<tr>
<td>RMS(bonds)</td>
<td>0.003</td>
</tr>
<tr>
<td>RMS(angles)</td>
<td>0.70</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>97</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>2.8</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.16</td>
</tr>
<tr>
<td>Rotamer outliers (%)</td>
<td>0.55</td>
</tr>
<tr>
<td>Clash score</td>
<td>0.2</td>
</tr>
<tr>
<td>Average B-factor</td>
<td>37.43</td>
</tr>
<tr>
<td>macromolecules</td>
<td>36.93</td>
</tr>
<tr>
<td>ligands</td>
<td>44.86</td>
</tr>
<tr>
<td>solvent</td>
<td>42.18</td>
</tr>
<tr>
<td>Statistics for the highest-resolution shell are shown in parentheses.</td>
<td></td>
</tr>
</tbody>
</table>

* Friedel mates were averaged when calculating reflection statistics.

Table 6.4: Table 1 generated by PHENIX.