

Gradual neofunctionalization in the convergent evolution of trichomonad lactate and malate dehydrogenases

Phillip A. Steindel, Emily H. Chen, Jacob D. Wirth, and Douglas L. Theobald*

Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02453

Received 1 November 2015; Accepted 15 February 2016

DOI: 10.1002/pro.2904

Published online 17 February 2016 proteinscience.org

Abstract: Lactate and malate dehydrogenases (LDH and MDH) are homologous, core metabolic enzymes common to nearly all living organisms. LDHs have evolved convergently from MDHs at least four times, achieving altered substrate specificity by a different mechanism each time. For instance, the LDH of anaerobic trichomonad parasites recently evolved independently from an ancestral trichomonad MDH by gene duplication. LDH plays a central role in trichomonad metabolism by catalyzing the reduction of pyruvate to lactate, thereby regenerating the NAD⁺ required for glycolysis. Using ancestral reconstruction methods, we identified the biochemical and evolutionary mechanisms responsible for this convergent event. The last common ancestor of these enzymes was a highly specific MDH, similar to modern trichomonad MDHs. In contrast, the LDH lineage evolved promiscuous activity by relaxing specificity in a gradual process of neofunctionalization involving one highly detrimental substitution at the “specificity residue” (R91L) and many additional mutations of small effect. R91L has different functional consequences in LDHs and in MDHs, indicating a prominent role for epistasis. Crystal structures of modern-day and ancestral enzymes show that the evolution of substrate specificity paralleled structural changes in dimerization and α -helix orientation. The relatively small “specificity residue” of the trichomonad LDHs can accommodate a range of substrate sizes and may permit solvent to access the active site, both of which promote substrate promiscuity. The trichomonad LDHs present a multi-faceted counterpoint to the independent evolution of LDHs in other organisms and illustrate the diverse mechanisms by which protein function, structure, and stability coevolve.

Keywords: protein evolution; ancestral sequence reconstruction; malate and lactate dehydrogenase; trichomonad; crystallography; enzymology; protein stability; epistasis

Introduction

Proteins can evolve divergent functions after gene duplication.^{1–3} Functions may arise anew from an ancestor with a different function (neofunctionaliza-

tion)^{1,4} or may result by specialization of a promiscuous ancestor (subfunctionalization).^{5–7} Several evolutionary theories posit different distributions of the size of the effect of substitutions on function: the gradualist

Abbreviations: BP, BAli-Phy; Ha, hypotrichomonas Acosta; LCA, last common ancestor; LDH, lactate dehydrogenase; MB, MrBayes; MCMC, Markov chain Monte Carlo; MDH, malate dehydrogenase; Ph, Pentatrichomonas hominis; PM, PhymI; PP, posterior probability; Tv, trichomonas vaginalis

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institutes of Health (NIH); Grant numbers: R01GM096053, R01GM094468; Grant sponsor: National Institute of Health; Grant sponsor: MINOS; Grant number: R01GM105404.

*Correspondence to: Douglas L. Theobald, MS009 Brandeis University, PO Box 549110, Waltham, MA 02454-9110.
E-mail: dttheobald@brandeis.edu

view emphasizes small positive effects,^{8,9} the neutral theory relies on small positive and negative effects,^{10,11} while saltation (or macromutation¹²) expects that a few large-effect mutations account for the bulk of functional change.¹³ Epistasis can modulate these effect sizes,¹⁴ but it is unknown whether epistasis generally constrains^{15,16} or opens up¹⁷ evolutionary paths.

The malate/lactate dehydrogenase superfamily has long been a model system for structural and functional evolution.^{18–20} These proteins catalyze the reduction of 2-ketocarboxylic acids via a NADH cofactor (Fig. 1). At least four LDH families evolved independently from MDHs: the canonical LDHs, two families of apicomplexan LDHs,^{20,21} and the trichomonad LDHs.²² A key active site residue, known as the “specificity residue” (R91 in the cytosolic trichomonad MDHs) mutated to four distinct residues in the four LDH families (Fig. 1). R91 is found on the active site loop, which closes upon substrate binding and is essential for catalysis.^{23,24} Mutations of the specificity residue can switch LDH specificity to MDH specificity in some LDHs²⁵ but not others,^{22,26} and additional mutations are necessary to swap specificity in MDHs.^{27,28}

The trichomonad LDHs evolved from the trichomonad cytosolic MDHs,²² a clade phylogenetically remote from other known LDHs. Trichomonad MDHs and LDHs occupy key roles in the cytosolic metabolism of these organisms—both operate in conjunction with the hydrogenosome, a hydrogen-producing, anaerobic mitochondrion.²⁹ In particular, *Trichomonas vaginalis* (*Tv*) LDH expression and activity increase in organisms with reduced hydrogenosome function,^{30–32} for instance in response to anti-trichomoniasis drugs. *Tv* MDH is highly specific, while *Tv* LDH is promiscuous.²²

Here we use ancestral sequence reconstruction, biochemical analysis, and crystallography to dissect the evolution of structure, function, and stability of the trichomonad LDHs and MDHs. We show that, like the unusual apicomplexan LDHs, the trichomonad LDHs evolved from a highly specific MDH. However, this process was gradual, involved more mutations, involved a different biochemical mechanism, and resulted in promiscuous LDHs.

Results

Modern LDHs are promiscuous; the modern MDH is highly specific

We expressed five modern trichomonad MDHs and LDHs (Fig. 2, Supporting Information Fig. S1) and assayed them for their ability to reduce 2-ketoacids under steady-state conditions (Fig. 3, Supporting Information Tables S3–S6): two *Trichomonas vaginalis* (*Tv*) MDHs, a *Tv* LDH, *Hypotrichomonas acosta* (*Ha*) LDH, and *Pentatrichomonas hominis*

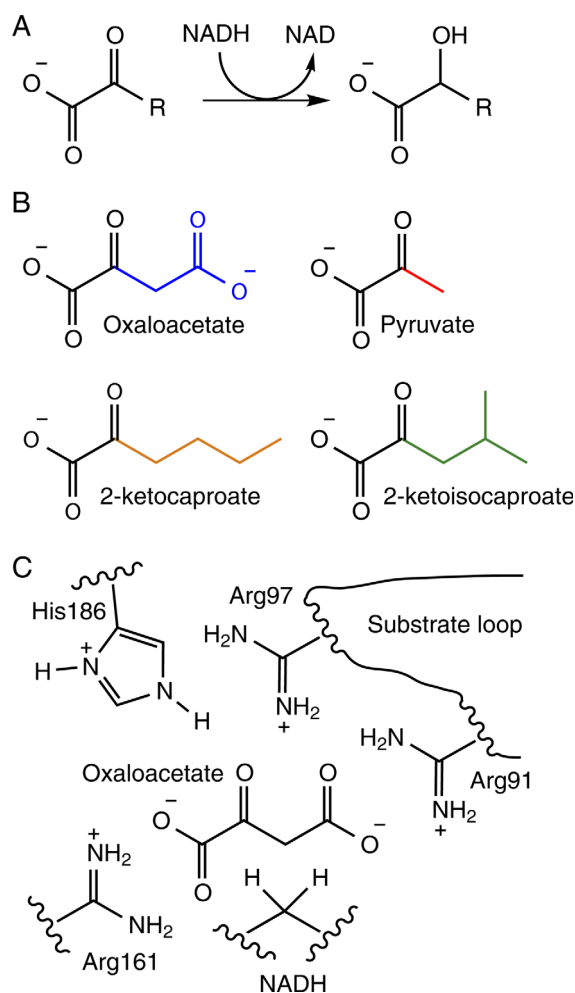


Figure 1. (A) NADH-dependent reduction of 2-ketoacids, the reaction catalyzed by L/MDHs. (B) 2-ketoacid substrates. Oxaloacetate (blue R group) is reduced in MDHs, Pyruvate (red) is reduced in LDHs, and 2-ketocaproate (yellow) and 2-ketoisocaproate (green) are other 2-ketoacids that may potentially be reduced by L/MDHs. (C) Active site of cytosolic MDHs, LDHs have a different residue at position 91: Gln in canonical LDHs, Lys in *Plasmodium* LDHs, Gly in *Cryptosporidium* LDHs, and Leu in trichomonad LDHs.

(*Ph*) LDH. The 2-ketoacids assayed were oxaloacetate (the MDH substrate), pyruvate (the LDH substrate), 2-ketocaproate, and 2-ketoisocaproate (alternate substrates and proxy indicators of promiscuity). In this work, we primarily report the steady-state kinetic constant k_{cat}/K_m as a measure of enzyme (low substrate) activity; “specificity” refers to relative activity among two or more substrates. One modern *Tv* MDH (MDH 1, henceforth referred to as simply *Tv* MDH) is highly specific for the MDH substrate oxaloacetate, with little activity towards pyruvate or alternate 2-ketoacids. In contrast, modern *Tv*, *Ph*, and *Ha* LDH and the other (likely misannotated) *Tv* MDH (MDH 2) are promiscuous LDHs, able to reduce oxaloacetate and other alternate substrates nearly as well as pyruvate.

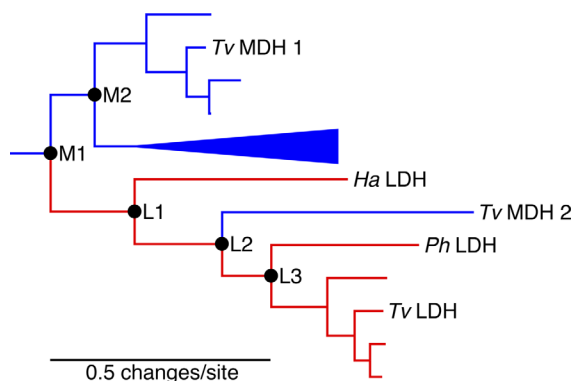


Figure 2. Phylogeny of Trichomonad LDHs and MDHs estimated by the program BALi-Phy. Full phylogeny is in Figure S2a. Exterior branches are colored by Genbank notation, where MDH is blue and LDH is red. Interior branches are colored by consensus of all descended branches. The nodes M1, L1, L2, and L3 have posterior probability greater than 99%; see Supporting Information Table S6.

Ancestral LDH and MDH reconstructions are robust

In order to recapitulate the evolution of substrate specificity in the Trichomonad LDH and MDH enzymes, we expressed and assayed key ancestral proteins in the lineages leading to the modern enzymes. We reconstructed three different sets of ancestral sequences, using three different phylogenies: one estimated with a maximum likelihood method [Phyml³³ (PM) phylogeny, Supporting Information Fig. S1(A)], one estimated with a Bayesian MCMC method [MrBayes³⁴ (MB) phylogeny, Supporting Information Fig. S1(B)], and the other estimated jointly with an alignment using a Bayesian MCMC method [BALi-Phy³⁵ (BP) phylogeny, Figs. 2, Supporting Information Fig. S1(C)]. By using three different phylogenetic analyses, we hoped to assess the robustness of our ancestral inferences to model assumptions and methodology. The topology of the trichomonad protein phylogeny is substantially similar among the three analyses.

Nodes M1, M2, L1, L2, and L3 are consistent among all three phylogenies and are well-supported (Fig. 2, Supporting Information Fig. S1, Table S2). Maximum likelihood ancestral sequences were reconstructed with PAML³⁶ for these nodes for each phylogeny. Most of the reconstructed residues have posterior probability (PP) greater than 90% [Supporting Information Fig. S2(A)]. The reconstructions for each node from the PM and MB phylogenies are almost identical, while the reconstructions from the BP phylogenies have more differences (Supporting Information Table S2). As expected, sites with low PP generally have the largest number of differences [Supporting Information Fig. S2(B,C)]. Due to their high sequence identity, we combined the PM and MB reconstructions into one set of sequences for

further analysis, using the highest posterior probability residue where there are differences (the PM/MB reconstructions); the BP reconstructions were analyzed separately.

Promiscuous ancestral LDHs, but a specific MDH LCA

The ancestral proteins M1, M2, and L1 from both reconstructions and L2 and L3 from the BP reconstruction were expressed and assayed in the same manner as the modern enzymes. The alternate reconstructions result in ancestral variants with somewhat different efficiencies, but the relative specificities are similar (Supporting Information Fig. S3). As we do not know which estimated ancestors are most representative of the “true” ancestors, we averaged $\log(k_{cat}/K_M)$ s for M1, M2, and L1 for a final estimation of ancestral efficiencies (Fig. 3, Supporting Information Tables S3–S6). The average and standard deviation should be considered a simple descriptive statistic, a conservative measure of resurrection disparity (i.e., any non-flat weighting scheme would give a smaller variance and bias towards one of the reconstructions).

M1, the last common ancestor (LCA) of trichomonad MDHs and LDHs, is a highly specific MDH, similar to *Tv* MDH1. M2, the LCA of the MDHs, is also highly specific. In contrast, L1, the LDH LCA, has a large increase in $\log(RS)$ for pyruvate, 2-ketocaproate, and 2-ketoisocaproate compared to M1, largely due to a substantial decrease in oxaloacetate activity. In this enzyme, pyruvate efficiency is barely increased relative to M1 and is essentially equal to that of oxaloacetate. In contrast, all other LDHs assayed, both ancestral and modern, are promiscuous enzymes (Fig. 3), with greater activity for all substrates overall but also greater pyruvate specificity (relative activity).

Structures of ancestral and modern enzymes

We solved the X-ray crystal structures of the ancestral MDH M1 (1.6 Å resolution, PDB entry 4UUP), modern *Tv* MDH (3.0 Å, 4UUO), and four variants of modern *Tv* LDH [apo WT (1.4 Å, 4UUM), apo L91R (1.3 Å, 4UUL), WT bound to NADH (1.9 Å, 4UUN), and WT bound to NADH and oxamate (2.0 Å, 5A1T) (Supporting Information Table S7). All enzymes are dimeric (Supporting Information Fig. S4), like other cytosolic MDHs. This dimeric LDH form is unique, as all other known LDHs are tetramers.²⁰ The LDH/NADH/oxamate structure has a similar dimeric form as the other structures; however, it additionally forms a tetramer in the crystal which resembles other tetrameric LDHs [Supporting Information Fig. S4(C)]. MDH and LDH structures are generally found in two different states (open or closed), depending on the conformation of the

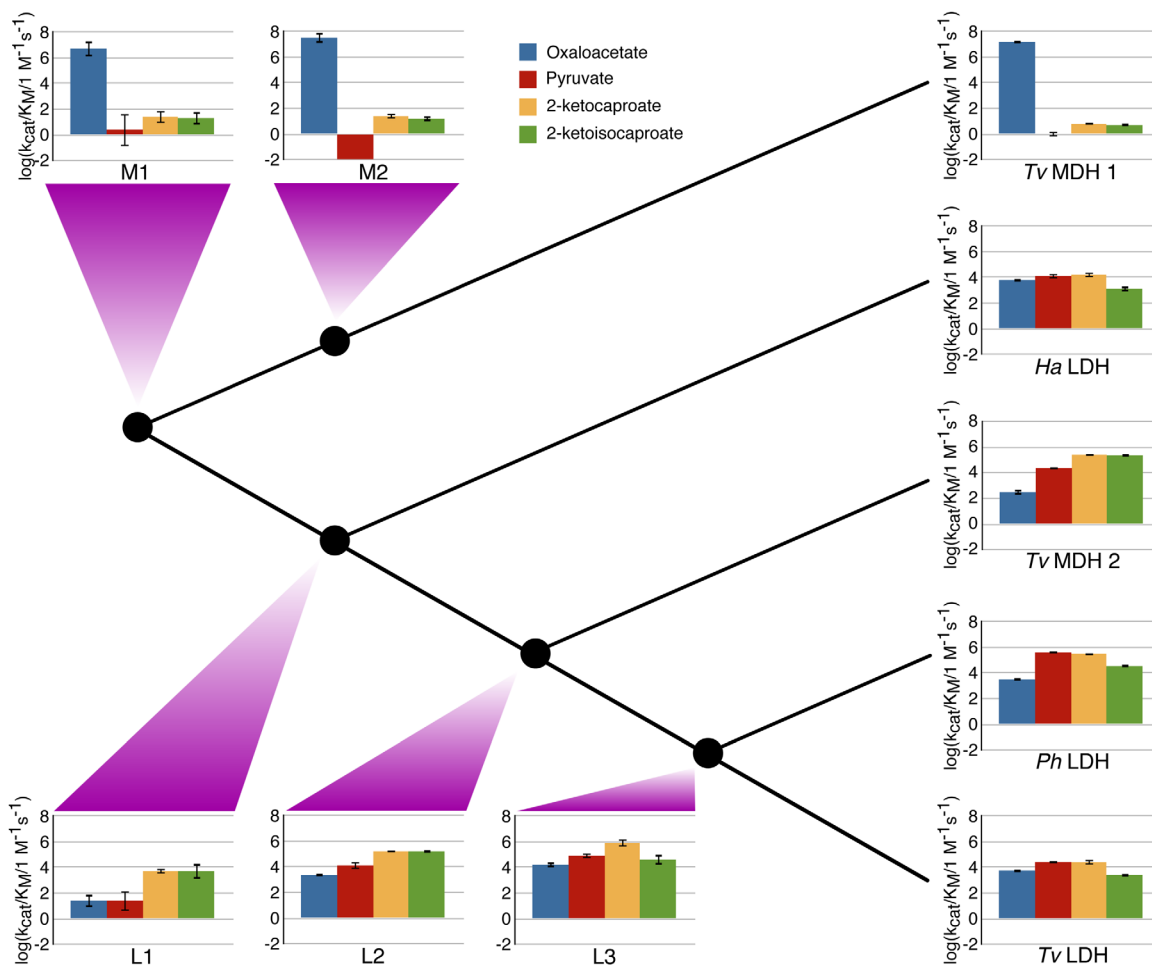


Figure 3. $\log(k_{cat}/K_M)$, relative to $1 \text{ M}^{-1} \text{ s}^{-1}$, for the NADH-dependent reduction of 2-ketoacid substrates by ancestral and modern trichomonad L/MDHs. Substrates tested were oxaloacetate (blue), pyruvate (red), 2-ketocaproate (yellow), and 2-ketoisocaproate (green). Enzymes are arranged according to an approximate phylogeny (see Fig. 2 and Supporting Information Fig. S2). M1, M2, and L1 activities are averages of the activities of the PM/MB and BP reconstructions (see Supporting Information Fig. S3). Errors for each enzyme/substrate pair are the largest of the standard error of averaging $\log(k_{cat}/K_M)$ from two separate experiments or the individual standard errors from curve fitting either set of experimental data.

active site loop. The M1 structure adopts two distinct closed states, the LDH/oxamate/NADH structure is in one closed state, and *Tv* MDH and other *Tv* LDH structures are in the open state.

In the M1 structure, only one subunit has a fully closed substrate loop (Supporting Information Figs. S4 and S6). There is also strong yet ambiguous electron density in the active site; a phosphate fits the density best and was therefore modeled (Supporting Information Fig. S6). The M1 active site loop is found in a similar conformation as other MDHs, with R91 contacting the phosphate modeled in the active site [Fig. 4(A)]. In contrast, the LDH active site loop is not fully closed [Fig. 4(B)], as L91 in the LDH structure is prevented from contacting the substrate analogue oxamate by residues on the “ $\alpha 1/2G$ ” helix,³⁷ A229 and W230. Closure of the active site loop still appears to occlude the active site from solvent, however.

In order to compare the structures, we calculated superpositions in which only one subunit of

each dimer was superposed, letting the other subunit move freely (Fig. 5). The superimposed subunits have an average RMSD of 1.9 \AA [Fig. 5(A)]. Overall the structures are highly similar, except for the active site loop, the C terminus, and a portion of the $\alpha 1/2G$ helix [labeled with number 4 in Fig. 5(A)]. Most of the variance of the active site loop is due to the differences between the closed and open structures; otherwise the substrate loops of *Tv* LDH (wild-type and L91R) and MDH are similar (Supporting Information Fig. S5).

The superposed subunits have lower RMSD than the subunits allowed to move freely [Fig. 5(B)], indicating that the relative orientation of subunits in the dimer is different among the various proteins. Most of this difference is due to the open LDH, as removing it reduces the RMSD [2.8 \AA with versus 1.9 \AA without; Fig. 5(C,D)]. This difference may simply be due to the conformational change between open and closed; however the open MDH structure

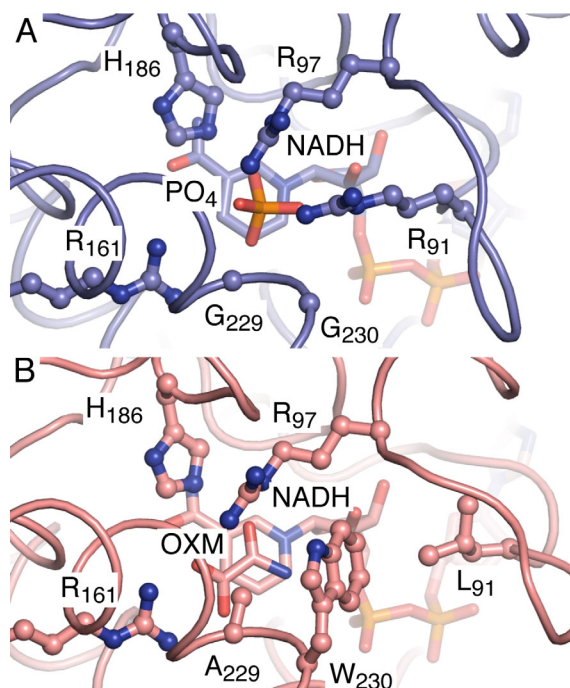


Figure 4. Comparison of active sites of (A) M1 (slate) and (B) closed *Tv* LDH (salmon).

superpositions better with closed M1 than does the open LDH structure. Therefore, the open dimerization orientation has likely been conserved in the MDHs but has changed at some point in the evolution of the LDHs.

There are 131 substitutions between M1 and *Tv* LDH, and 84 substitutions between M1 and *Tv* MDH (out of 333 total residues in M1). The dimer interface has a large concentration of substitutions. Of the approximately 40 residues that compose the interface, 22 residues differ between M1 and *Tv* LDH [Fig. 6(A), Supporting Information Table S8]. In contrast, only 9 residues at the interface differ between M1 and *Tv* MDH, mostly at the periphery of the interface [Fig. 6(B), Supporting Information Table S8]. In particular, three substitutions in *Tv* LDH dimer disrupt salt bridges found in the center of M1 and *Tv* MDH (Supporting Information Fig. S7). The NADH binding site also has a higher incidence of substitutions between M1 and *Tv* LDH — 10 substitutions out of ~23 residues total [Fig. 6(C), Supporting Information Table S8]. Several of these residues are also in the vicinity of the active site or near the dimer interface.

There are 30 total substitutions separating M1 and L1. Fifteen of these substitutions occur within 14 Å (~4 turns of an α -helix) of the active site phosphate in the closed subunit (Supporting Information Table S8). In M1, a 114 residues out of 333 total are within 14 Å of the active site phosphate (considering the entire substrate loop as within that radius), indicating that residue substitutions were about twice

as common in proximity to the active site versus distal regions.

Stability of ancestral and modern enzymes

Thermostability is strongly correlated with protein structure and function,³⁸ and therefore we examined how thermostability coevolved in the trichomonad LDHs and MDHs. We measured the thermostability of PM/MB M1, M2, and L1; BP L2 and L3; *Tv* LDH; *Ha* LDH, and *Tv* MDH (Table I). All proteins assayed except M2 undergo a single unfolding transition during thermal denaturation (Supporting Information Fig. S8). M1 and L2 unfold reversibly, while L3, *Tv* MDH, and *Tv* LDH unfold irreversibly. Unfolding of L1 is only partially reversible, while M2 has two unfolding transitions, the first reversible and the second irreversible (Supporting Information Fig. S8). All the MDHs have similar stabilities, whether modern or ancestral. In contrast, all ancestral LDH proteins are considerably more stable than their modern LDH descendants, with increases in melting temperature ranging from 5° to 20°C (Table I).

The “specificity residue” is insufficient for evolution of LDH activity

Tv LDH residue 91 is commonly regarded as the “specificity residue” in the MDH/LDH family (corresponding to residue 102 in conventional dogfish LDH numbering³⁹).^{19,20} All known MDHs have an Arg for the specificity residue [see Fig. 1(C)], while canonical LDHs have a Gln. In contrast, in modern trichomonad LDHs the specificity residue is nearly always a Leu (Supporting Information Table S1). It is thought that the specificity residue contributes to substrate discrimination largely by balancing charge in the active site: the positively charged Arg in MDHs interacts with and negates the negatively charged C4 carboxylate of the oxaloacetate/malate substrate, whereas the neutral Gln in canonical LDHs interacts with the C3 methyl of the pyruvate/lactate substrate.

To ascertain the importance of the specificity residue in the evolution of the Trichomonad proteins, we mutated Leu91 to Arg in both PM/MB L1 and *Tv* LDH, and we also mutated R91 to Leu in both PM/MB M1 and *Tv* MDH [Fig. 7(A)]. In the MDHs, R91L has a large negative effect on both M1 and *Tv* MDH oxaloacetate activity, with only minor effects on pyruvate, 2-ketocaproate, or 2-ketoisocaproate activity. However, the reverse mutation in LDHs, L91R, does not have the reverse effect—oxaloacetate affinity increases marginally and pyruvate affinity decreases only slightly. In sum, in both the modern and ancestral enzymes, swapping the specificity residue is insufficient to interconvert MDHs and LDHs [Fig. 7(A)].

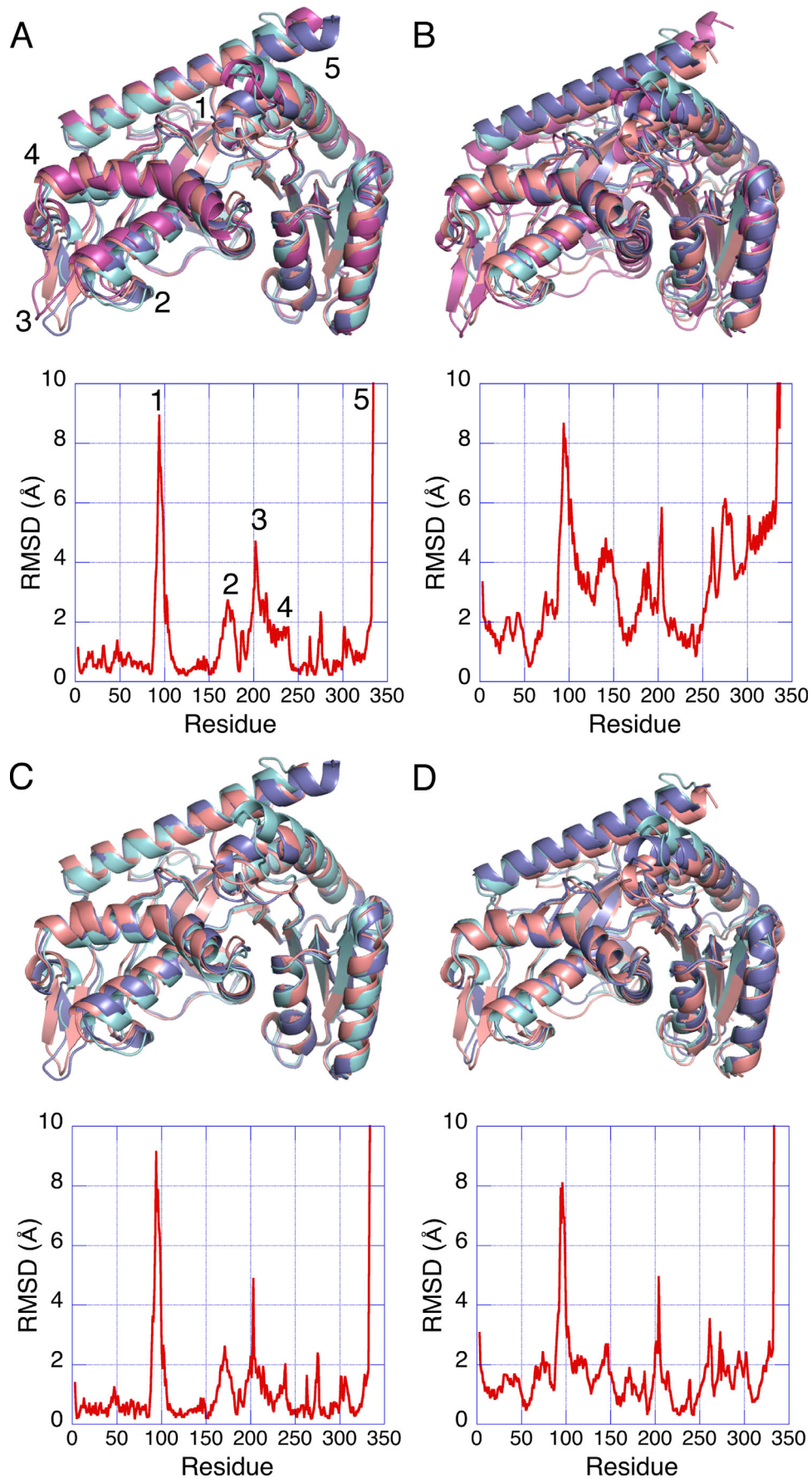


Figure 5. Superpositions of trichomonad MDH and LDH structures (top) and RMSD by residue (bottom). (A, B) Superpositions of the open (magenta) and closed (salmon) *Tv* LDH, *Tv* MDH 1 (cyan), and M1 (slate) dimers, superimposing chain A only and leaving chain B free. Chain A is displayed in (A) and chain B is displayed in (B). Numbers in the superposition in (A) correspond to numbered peaks in the RMSD versus residue plot. (C, D) Superpositions (A, B) with open *Tv* LDH removed and RMSDs recalculated.

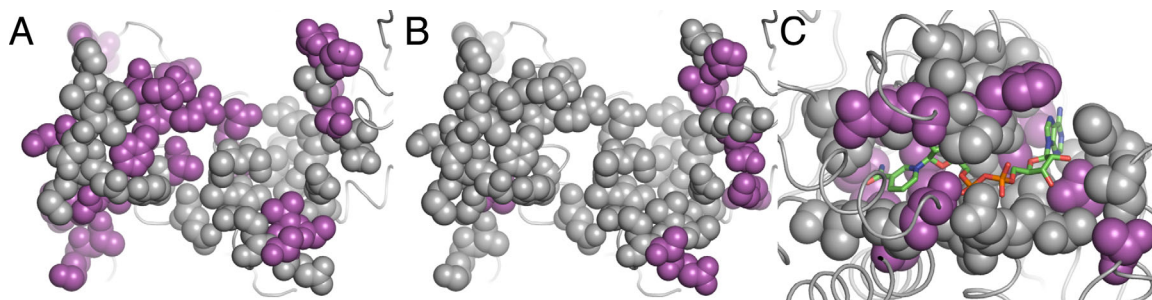


Figure 6. (A) Substitutions between M1 and inhibitor-bound or apo *Tv* LDH (magenta spheres) at either of their dimer interfaces (mapped onto M1 subunit A, grey spheres). (B) Substitutions between M1 and *Tv* MDH (magenta spheres) at either of their dimer interfaces (mapped onto M1 subunit A, gray spheres). (C) Differences between M1 and inhibitor-bound *Tv* LDH (magenta spheres) at the M1 NADH binding site (gray spheres; NADH as green, blue, red, and white sticks).

Other substitutions between M1 and L1, in addition to the specificity residue, are therefore necessary to account for the change in specificity. Substitutions physically near the M1 active site [Fig. 7(B)] are likely candidates for influencing specificity and activity. Using the closed-state M1 structure as a guide, we swapped residues proximal to the active site that differ between M1 and L1 [Fig. 7(B)]. We made three constructs, each with an increasing number of substitutions found within a given distance of the active site: (1) a set of all substitutions up to ~ 7 Å from the active site phosphate (R91L, F89V, T155S, G230W, I233L, and S239T, the “7 Å swap”), (2) another set up to ~ 10 Å away (the 7 Å swap residues plus M95Q, R156M, N188E, and A235M, the “10 Å swap”), and (3) the “10 Å swap” plus three additional residues flanking the substrate loop (G87A, E102S, and G106, the “10 Å/loop swap”). The 7 Å swap increases pyruvate activity, the 10 Å swap decreases all activities relative to the 7 Å swap, and the 10 Å/loop swap increases all activities, with a specificity profile similar to that of L1 [Fig. 7(B)].

Discussion

Promiscuous LDHs evolved from a highly specific MDH via a crippled intermediate

In the trichomonads, a group of promiscuous LDHs and specific MDHs have a highly specific MDH as their last common ancestor (Fig. 3). In the M1 LCA, activity for pyruvate is extremely low, while the ancestral L3 and the modern LDHs all have appreciable pyruvate activity (k_{cat}/K_M of 10^3 – 10^4). Hence, the evolution of novel LDH enzymes in the trichomonads is best explained as neofunctionalization,² rather than other models of gene duplication like escape from adaptive conflict⁴⁰ or subfunctionalization.⁵ Our group has previously shown that the evolution of apicomplexan LDHs follows a similar model.⁴¹ These and other results⁴² confirm that neofunctionalization is a viable model for evolution of novel activities, at least in certain protein families.

Unlike the LCA of the apicomplexan LDHs,⁴¹ the LCA of the trichomonad LDHs (L1) has an altered specificity profile with pyruvate activity at an efficiency below most other LDH enzymes.⁴³ It is possible that this low activity may be an artifact of the ancestral reconstruction method. There are two general ways that one could make an enzyme with an artifactually low pyruvate activity: (1) destabilize the protein or (2) impair its catalytic ability. However, our experimental results largely rule out these possibilities.

The L1 protein is not destabilized, since it expresses well in *E. coli*, is well-folded (see circular dichroism data in Supporting Information Fig. S8), the thermal denaturation T_m is about 16–18°C higher than modern trichomonad LDHs, and remarkably the L1 thermal denaturation is reversible (Supporting Information Fig. S8). It has been suggested, based on purely computational analysis, that the increased thermostability of ancestral proteins may itself be an artifact,⁴⁴ but this is also unlikely for the trichomonad ancestors. First, L1 is thermostable only relative to the modern LDHs, but not compared to the modern MDHs, which have very similar thermostability. Second, the ancestral trichomonad MDHs do not have increased thermostability relative to the modern MDHs, which directly contradicts the claim that ancestral reconstruction methodology artifactually increases thermostability.

Table I. Thermodynamic Parameters from CD Melting Experiments

Protein	T_m (K) ^a	ΔH_m (kcal/mol) ^a
M1	330.7 ± 0.4	140 ± 10
M2	332.3 ± 0.2	140 ± 20
<i>Tv</i> MDH 1	333.4 ± 0.3	152 ± 7
L1	336.9 ± 0.8	70 ± 10
L2	338 ± 2	100 ± 10
L3	326.6 ± 0.6	68 ± 6
<i>Tv</i> LDH	318 ± 1	200 ± 100
<i>Ha</i> LDH	320.7 ± 0.3	180 ± 40

^a Errors are standard errors from averaging (three replications, except for *Ha* LDH which had four).

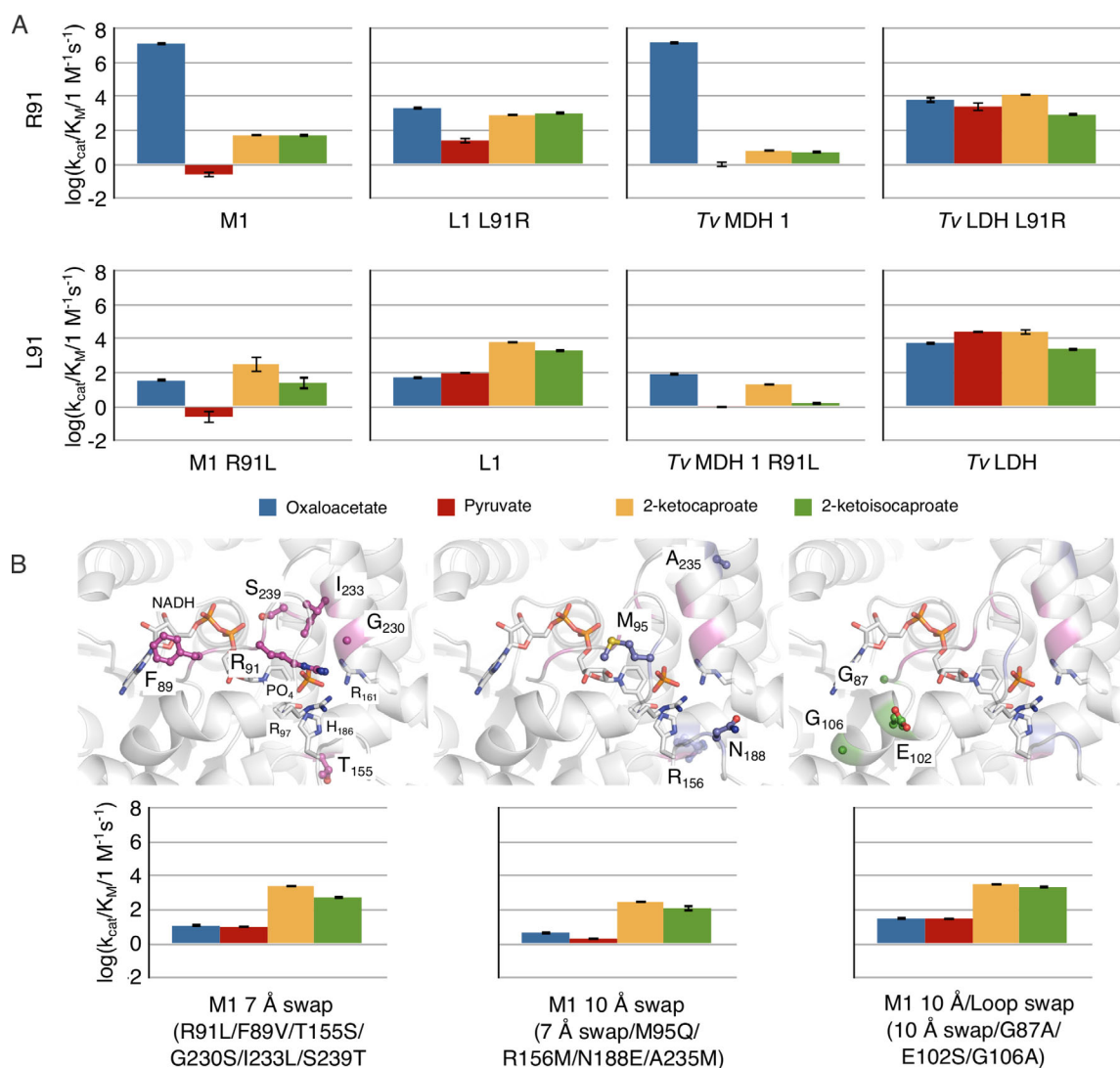


Figure 7. $\log(k_{cat}/K_M)$, relative to $1M^{-1} s^{-1}$, for the NADH-dependent reduction of 2-ketoacid substrates of ancestral and modern enzyme mutants. Errors are calculated as in Figure 3. (A) Results of the L/R swap at position 91 in ancestral and modern trichomonad enzymes. M1 and L1 (PM/MB reconstructions) data are also in Figure S3; *Tv* MDH and LDH data are also in Figure 3. (B) Results of other mutations to M1. Locations of the residues mutated in the M1 structure are indicated above the kinetics results, along with the key active site residues H186 and R161, NADH, and phosphate (white).

More importantly, the ancestral L1 enzyme is a good catalyst, reducing 2-ketocaproate and 2-ketoisocaproate substrates with high efficiency (k_{cat}/K_M of $\sim 10^4$). Furthermore, a single mutation (L91R) in L1 confers physiologically relevant activity for oxaloacetate [k_{cat}/K_M of $>10^3$, similar to the modern promiscuous *Tv* LDH, Fig. 7(A)]. The L91 site has a very high reconstruction confidence ($>99\%$ PP), as expected from visual inspection of the sequences (most of the descendants have L91; see Supporting Information Table S1). Finally, both alternative ancestral reconstructions of L1 have a similar kinetic profile (Supporting Information Fig. S3, high activity towards 2-ketocaproate and 2-ketoisocaproate, lower activity towards pyruvate and oxaloacetate), and the branch supports in this region of the phylogeny are high ($>99.8\%$ PP for all nodes

within 2 from ancestral node L1, even accounting for alignment uncertainty with Bali-Phy; Supporting Information Table S2). All of these factors indicate that L1 has been reconstructed reliably.

Assuming, then, that the reconstructed activity is representative of the actual ancestor, L1 had three possible ancestral enzymatic roles. One may have been as a very inefficient LDH that nevertheless had sufficient activity for selection to act upon. A second possibility is that L1 may have served as non-LDH 2-keto acid reductase, to reduce 2-ketocaproate, 2-ketoisocaproate, or a similar small metabolite, since activity for these substrates is representative of “normal” enzyme efficiencies. A third possibility is that the enzyme was in fact non-functional, as in the classical neofunctionalization model¹ (also known as mutation during nonfunctionality²). The

absence of LDHs in some trichomonads²⁹ supports this possibility—a nonfunctional LDH could have been more easily pseudogenized.

Pyruvate activity increased at the expense of oxaloacetate activity

In the trichomonad LDH evolutionary lineage, activities for all tested substrates gradually increased from the LDH last common ancestor L1. This increase was not uniform, as oxaloacetate appears to have “topped out” at an efficiency of $\sim 10^4 M^{-1} s^{-1}$. Meanwhile, pyruvate efficiency increased more than other substrates, resulting in an overall increase in pyruvate specificity (e.g., *Ph* LDH). This pattern of evolution is somewhat reminiscent of that observed with escape from adaptive conflict (EAC), wherein one activity increases at the expense of another as both cannot be maximized simultaneously. From our data, however, it is difficult to discern if pyruvate and oxaloacetate reduction are truly “conflicting” functions in the trichomonad MDH protein scaffold—the apparent inverse correlation between pyruvate and oxaloacetate activity in the MDHs may not be due to physical constraints (it could be adaptive). In any case, this pattern of Trichomonad LDH evolution contrasts sharply with apicomplexan LDH evolution, where a highly active bifunctional enzyme was likely an evolutionary intermediate, and EAC is ruled out.⁴¹

Specificity is determined by correlated substitutions of varying size throughout the protein

The “specificity residue” R91 is completely conserved in the trichomonad MDHs, and L91 is highly conserved in the LDHs. However, the R91L mutation, while having a large effect on specificity, does not fully account for the evolutionary change in specificity and activity in the trichomonad LDHs [Fig. 7(A)]. In MDHs, R91L greatly decreases oxaloacetate activity without affecting pyruvate activity, whereas in LDHs, L91R has very little effect on any substrate, resulting in only a marginal increase in oxaloacetate activity. Hence, the specificity residue at position 91 is influenced by epistatic interactions with other residues in the proteins, since L91 and R91 have different effects on activity in the MDHs versus the LDHs.

The structures of closed M1 and *Tv* L1 suggest a possible source of this epistasis. In the LDH, other mutations adjacent to the active site prevent L91 from directly contacting the substrate. Therefore, the identity of this residue may have a moderate influence on substrate specificity. Meanwhile, R91 in M1 is sterically constrained at the γ -carbon by the NADH, the substrate, and waters mediating NADH binding (Fig. 8). Leucine is branched at the γ -carbon, which apparently prohibits a sterically

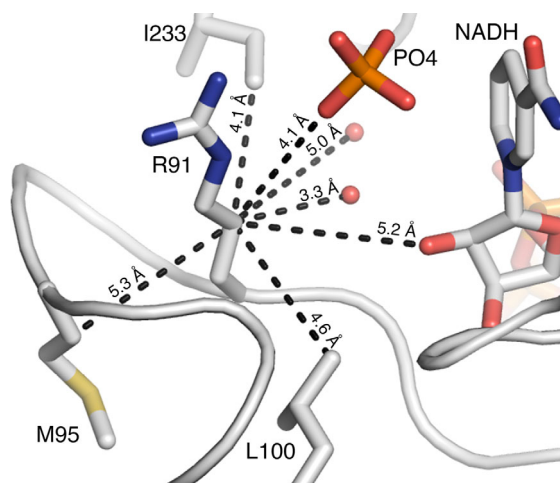


Figure 8. M1 arginine 91, with distances between the γ -carbon and other nearby nonhydrogen atoms.

allowed leucine rotamer when the substrate loop is in the closed conformation. Therefore, R91L could impede loop closure for all substrates, resulting in a much less active enzyme.

Substitutions outside the active site and active-site loop affect activity [Fig. 7(B)], indicating specificity in the LDHs involves epistasis and is determined by more than just the charge of the specificity residue at position 91. Even so, the five additional mutations involved in the “M1 7 Å swap” impart approximately the same phenotype as L1. While the residues enclosing the active site modulate the effect of residue 91, they have only slight epistatic interactions with the rest of the protein. Thus, in the trichomonad M/LDHs, epistatic effects involved in specificity appear largely restricted to a 7–10 Å shell around the active site.

Specificity for pyruvate has increased several distinct times over the course of LDH evolution (e.g., from L1 to L2, from L2 to L3, from L3 to *Ph* LDH, and from L1 to *Ha* LDH, Fig. 3). During each of these transitions, substitutions occurred in different regions of the protein that may be functionally relevant, such as at the NADH binding site and dimerization interface. Different combinations of substitutions in distinct regions may be responsible for each change in specificity, and some may have affected activity by changing broad properties like dimerization orientation and stability. These substitutions may also have had an effect on conformational changes, as open and closed LDH have quite different conformations. This is in contrast to the MDHs which apparently only differ in their substrate loops on closing.

Interestingly, the epistatic constraints on the specificity residue apparently have not changed appreciably during evolution (in contrast to the “epistatic ratchet” model of functional change⁴⁵), as R91L has the same effect in the modern LDH and

MDH as in their ancient counterparts [L1 and M1 respectively, Fig. 7(A)]. It is therefore likely that the closed L1 and *Tv* MDH structures resemble closed *Tv* LDH and M1, respectively. We do not know, however, what accounts for the change in specificity between L1 and the modern enzymes. We expect only a subset of the substitutions we have outlined here are important; discerning the key substitutions from all the possibilities is a challenging problem that we are currently pursuing.

Materials and Methods

Phylogeny and ancestral reconstruction

Initial trichomonad protein sequences were obtained from previous work on the MDH/LDH superfamily;⁴¹ additional trichomonad sequences were obtained with BLAST and PSI-BLAST⁴⁶ using these sequences as queries. Other MDH sequences were manually selected from the set of eukaryotic cytosolic MDHs and closely related eubacterial MDHs. Genes were named as annotated in the NCBI protein database. Two trichomonad MDH sequences (GI 123491614 and 123501568) were eliminated as pseudogenes due to missing catalytic residues. Initial alignments were made with MUSCLE.⁴⁷ The same alignment was used to estimate both phylogenies and ancestral sequences for the Phyml and MrBayes reconstructions, while the BALi-Phy reconstruction used the alignment estimated with the phylogeny. Phyml 3.0³³ was run with LG exchangeabilities and empirical frequencies, starting with a BIONJ phylogeny.⁴⁸ Two MrBayes⁴⁹ chains were run with a mixed evolutionary model for 1113700 cycles, with samples taken every 100 cycles. Four BALi-Phy 2.1.1³⁵ chains were run with LG exchangeabilities and RS07 indel model for 100000 cycles. All phylogenies were inferred with a discrete gamma distribution for variation of evolutionary rates over sites: Phyml and BALi-Phy with four categories and MrBayes with 12 categories (PROTTEST,⁵⁰ using the AIC criteria, found LG + Gamma to be the best model). The first 25% of the MrBayes chains and 10% of the BALi-Phy chains were discarded as burn-in. The final ASDSFs for both the MrBayes and BALi-Phy chains were 0.032, which is above the ideal value (0.01) but still likely indicative of convergence. Marginal probabilities for residues at ancestral nodes were calculated with the codeml function of PAML v4.³⁶ The gamma shape parameter and amino acid frequencies were re-estimated, along with branch lengths for the Phyml and MrBayes phylogenies. The most likely ancestral sequences at nodes M1, M2, L1, L2, and L3 were then extracted. Gaps were inferred via parsimony.

Expression and purification

For all proteins, codon optimized genes were synthesized by GenScript and subcloned into a pet21b vec-

tor between the NdeI and XhoI restriction sites, adding a C-terminal poly-histidine tag with sequence LEHHHHHH. The *Ha* and *Ph* LDH sequences in the NCBI database were missing N-termini, so residues 2-20 of our *Ha* LDH and 1-16 of our *Ph* LDH were taken from BP M1. All enzymes were expressed and purified with the C-terminal tag; this tag was not removed. Primers for site-directed mutagenesis were synthesized by IDT. Site-directed mutagenesis was done with QuikChange Multi kit (Agilent) using the standard protocol in a PTC-200 Peltier thermocycler (MJ Research). Mutants were confirmed by Sanger sequencing from the T7 primer (Genewiz). The BP/PM M1 “7 Å swap” was mutated from M1 R91L in the order G230W, then F89V/T155S/S239T, then I233L; M1 “10 Å swap” from the 7 Å swap in the order A235M, then R156M/N188E, then M95Q; and the M1 “10 Å/loop swap” from the 7 Å swap in the order G87A/G106A, then E102S.

Vectors were transformed into BL1(DE3)pLysS cells (Promega). Additionally, for the protein used in oxaloacetate kinetic assays alone, the BP L1 vector was transformed into BL21(DE3)pLysS ΔMDH cells. These cells were created by previous work⁴¹ using the λ Red recombinase system.⁵¹ After overnight growth in 5 mL LB, cells were transferred into 1-1.5 L LB. Expression was induced with 200 μM IPTG (Fisher) when cells reached an OD600 of 0.4-0.8. Expression continued overnight at 25°C. For the standard cells containing PM/MB L1 and *Tv* MDH R91L vectors as well as the ΔMDH cells containing the BP L1 vector, expression was additionally induced with 1 mM IPTG and continued for only four hours.

Pelleted cells were frozen, then incubated for 30–60 min in a buffer of 50 mM sodium phosphate pH 8.0, 300 mM NaCl, and 10 mM imidazole (Fisher). Cells were sonicated with a PTC-200 Sonic Dismembrator (Fisher) at 35% intensity. Lysate was flowed over a 5 mL HisTrap HP column (GE Healthcare Life Sciences) on an ÄKTAPrime FPLC (GE), then protein was eluted with a gradient to 500 mM imidazole. Protein product was concentrated with Amicon Ultra-15 concentrators with a 10 kDa molecular weight cutoff (Millipore). Product was then exchanged into 50 mM Tris pH 7.6, 0.1 mM EDTA, and 0.02% NaN₃ (Fisher) over a HiPrep Sephacryl column on an ÄKTAPrime Plus or ÄKTAPurifier FPLC (GE), and concentrated for final product. Concentration was determined by measuring absorbance at 280 nm with a Nanodrop 2000c spectrophotometer (Thermo Scientific), with extinction coefficient and molecular weight calculated using exPasy's ProtParam tool (<http://web.expasy.org/prot-param/>). Protein was flash frozen with 50% glycerol in liquid nitrogen then stored at –80°C.

Kinetics

Reagents for each reaction were obtained from Sigma, except for 2-ketoisocaproate, which was purchased from Santa Cruz Biotechnology. Reactions were done in 50 mM triethanolamine (Sigma), pH 7.6, at 25°C. Absorbance at 340 nm for each reaction was monitored by a Cary 100Bio UV-visible spectrophotometer (Varian) in kinetics mode. Three reactions were run simultaneously in quartz self-masking cuvettes with 10 mm path length (Starna), either 1 mL (catalog number 9B-Q-10) or 0.5 mL (18B-Q-10) in volume. Reactions were monitored each second for 600 s total. Maximum rates during the reaction were taken from reaction traces with Cary WinUV Kinetics software (Varian). The concentration of NADH was 200 μ M, well above saturating levels for MDHs and LDHs which generally have a K_M of \ll 50 μ M for NADH²⁴ (including cytosolic MDH).⁵² For L2, L3, and the modern LDHs, we assayed oxaloacetate activity with high concentrations of enzyme (50–200 nM) and low concentrations of oxaloacetate (5–100 μ M, kept on ice until assay), which resulted in a biphasic absorbance curve. We used the second, slow, phase to estimate oxaloacetate kinetic parameters; this was to compensate for pyruvate contamination in oxaloacetate stocks (see previous work⁴¹).

Rate versus substrate concentration was fit either to the Michaelis-Menten equation, Haldane equation for substrate inhibition,⁵³ or Haldane equation with K_i set equal to K_M . Curve fitting was done with Kaleidagraph (Synergy Software). Catalytic efficiency (or activity) for a given substrate is reported as k_{cat}/K_M or $\log(k_{cat}/K_M)$. Relative specificity [$\log(RS)$] for each substrate is reported relative to oxaloacetate using [$\log(k_{cat}/K_M(\text{substrate})) - \log(k_{cat}/K_M(\text{oxaloacetate}))$].

X-ray crystallography

Tv MDH and M1 were exchanged into 5 mM Tris, 0.1 mM EDTA and 0.02% NaN₃ (Fisher) with a PD-10 desalting column (GE Healthcare). All proteins were crystallized by hanging-drop (2 μ M protein and 2 μ M well solution) in VDX greased plates with siliconized cover slides (Hampton). Initial crystallization conditions were found with Hampton crystal screens HR2-110 and HR2-112. *Tv* LDH WT crystallized at 44 mg/mL in 100 mM HEPES pH 7.6, 8.3% PEG-6000 (Fluka), and 5% MPD and was cryoprotected in the same solution with 16.7% MPD. LDH L91R crystallized at 21 mg/mL in 100 mM HEPES pH 7.0, 10% PEG-6000, and 5% MPD and was cryoprotected with 15% MPD. LDH WT also crystallized at 20 mg/mL with 10X NAD in 100 mM HEPES pH 6.8, 8% PEG-6000 and 5% MPD and was cryoprotected with 22% MPD; additionally, it crystallized at 20 mg/mL with 4X NADH and 10X oxamate in HR2-

112 reagent 14 (0.2M potassium sodium tartrate tetrahydrate, 0.1M sodium citrate tribasic dihydrate pH 5.6, 2.0M ammonium sulfate). *Tv* MDH crystallized at 8.3 mg/mL in 4.0M sodium formate. M1 crystallized at 5 mg/mL with 4X NAD and sodium citrate in 100 mM Tris, pH 8.0, 25% PEG 4000, and 0.2M Li₂SO₄.

Diffraction data of all crystals were taken at 100 K. Data for *Tv* LDH and LDH L91R crystals were collected at the X29 beamline at the NSLS and for *Tv* MDH and M1 at the SIBYLS beamline (12.3.1) at the ALS. High and low intensity exposures were taken for LDH L91R and M1. Data for the LDH/NADH and LDH/NADH/oxamate crystal were collected on an Oxford Xcalibur PX system (Varian). Data were processed with XDS,⁵⁴ using all reflections with a $CC_{1/2}$ above 10%. Molecular replacement and refinement, using all processed reflections, was done with the Phenix software suite,⁵⁵ versions 1.8.1, 1.8.2, and 1.9. Modeling was done with Coot 0.6 and 0.7.⁵⁶ Superpositions were done with THESEUS in alignment mode,^{57–59} all other analysis was done with Pymol.⁶⁰ Residues in the binding interface were defined as those within 3.5 Å of any atom on the other subunit, and residues in the NADH-binding pocket were defined as those within 3.5 Å of any atom on the NADH. Figures were made in Pymol, with electron density maps generated by Phenix.

Thermal denaturation

Stability was assayed by measuring the circular dichroism (CD) signal of each protein from 250 nm to 195 nm while increasing temperature from 25° to 81°C. Protein was exchanged into 50 mM sodium phosphate, pH 7.4 using a PD-10 column (GE). Spectra were measured in a Jasco J-810 spectropolarimeter with a Peltier temperature controller. Spectra were measured in a quartz cuvette with 1 mm path length and PTFE stopper (Starna Cells, catalogue number 21-Q-1) and were at a concentration of 0.1 mg/mL (0.067 mg/mL for *Tv* LDH). Single spectra were taken every 2°. To test for reversibility, samples were heated to 5° above T_m before returning to 25°C. Reversibility was indicated by recovery of the spectrum at 25° after unfolding. Spectra were analyzed with Kaleidagraph (Synergy). The signal at 222 nm was fit directly to the equation $S(T) = F(T) + [U(T) - F(T)]/[1 + \exp(\Delta H_m(1 - T/T_m)/(RT))]$, where $S(T)$ is the CD signal and $F(T)$ and $U(T)$ are linear functions describing the signal of the folded and unfolded protein, respectively (C_p was expected to have a negligible effect on unfolding and so was set to zero⁶¹).

Acknowledgements

The crystallographic data collection was conducted at the SIBYLS beamline at the Advanced Light

Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research.

References

- Ohno S (1970). Evolution by gene duplication. Springer-Verlag.
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nature Rev Genet* 9:938–950.
- Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 11:97–108.
- Bergthorsson U, Andersson DI, Roth JR (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci USA* 104:17004–17009.
- Jensen RA (1976) Enzyme recruitment in evolution of new function. *Ann Rev Microbiol* 30:409–425.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6:R91–R105.
- Fisher R (1999). The genetical theory of natural selection: a complete variorum edition. Oxford, UK: Oxford University Press.
- Huxley J (2010). Evolution: the modern synthesis. Cambridge, MA: MIT Press.
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98.
- Kimura M (1983). The neutral theory of molecular evolution. Cambridge, UK: Cambridge University Press.
- Orr HA, Coyne JA (1992). The genetics of adaptation: a reassessment. *Am Naturalist* 140:725–742.
- Orr HA (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev Genet* 9:855–867.
- Maynard Smith J (1970). Natural selection and the concept of a protein space. *Nature* 225:563–564.
- Weinreich DM, Watson RA, Chao L (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–1174.
- Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006). Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444: 929–932.
- Goward C, Nicholls D (1994). Malate dehydrogenase: a model for structure, evolution, and catalysis. *Protein Sci* 3:1883–1888.
- Golding GB, Dean AM (1998). The structural basis of molecular adaptation. *Mol Biol E* 15:355–369.
- Madern D (2002). Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J Mol E* 54:825–840.
- Madern D (2003). Evolution of *Cryptosporidium parvum* lactate dehydrogenase from malate dehydrogenase by a very recent event of gene duplication. *Mol Biol E* 21:489–497.
- Wu G, Fiser A, ter Kuile B, Sali A, Müller M (1999). Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase. *Proc Natl Acad Sci USA* 96:6285–6290.
- Adams MJ, Buehner M, Chandrasekhar K, Ford GC, Hackert ML, Liljas A, Rossmann MG, Smiley IE, Allison WS, Everse J, Kaplan NO, Taylor SS (1973). Structure-function relationships in lactate dehydrogenase. *Proc Natl Acad Sci USA* 70:1968–1972.
- Holbrook JJ, Liljas A, Steindel SJ, Rossmann MG (1975). Lactate dehydrogenase. *Enzymes* 11:191–292.
- Wilks HM, Hart KW, Feeney R, Dunn CR, Muirhead H, Chia WN, Barstow DA, Atkinson T, Clarke AR, Holbrook JJ (1988). A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* 242:1541–1544.
- Cendrin F, Chroboczek J, Zaccai G, Eisenberg H, Mevarech M (1993). Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaeobacterium *Haloarcula marismortui*. *Biochemistry* 32:4308–4313.
- Nicholls DJ, Miller J, Scawen MD, Clarke AR, Holbrook JJ, Atkinson T, Goward CR (1992). The importance of arginine 102 for the substrate specificity of *Escherichia coli* malate dehydrogenase. *Biochem Biophys Res Commun* 189:1057–1062.
- Yin Y, Kirsch JF (2007). Identification of functional paralog shift mutations: conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. *Proc Natl Acad Sci USA* 104:17353–17357.
- Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu RY, van der Giezen M, Tielens AGM, Martin WF (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* 76:444–495.
- Kulda J (1999). Trichomonads, hydrogenosomes and drug resistance. *Intl J Parasitol* 29:199–212.
- De Jesus JB, Cuervo P, Junqueira M, Britto C, Costa e Silva-Filho F, Soares MJ, Cupolillo E, Fernandes O, Domont GB (2007). A further proteomic study on the effect of iron in the human pathogen *Trichomonas vaginalis*. *Proteomics* 7:1961–1972.
- Leitsch D, Kolarich D, Binder M, Stadlmann J, Altmann F, Duchêne M (2009). *Trichomonas vaginalis*: metronidazole and other nitroimidazole drugs are reduced by the flavin enzyme thioredoxin reductase and disrupt the cellular redox system. Implications for nitroimidazole toxicity and resistance. *Mol Microbiol* 72:518–536.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *System Biol* 59:307–321.
- Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Redelings BD, Suchard MA (2007). Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. *BMC Evol Biol* 7:40
- Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol E* 24:1586–1591.
- Iwata S, Kamata K, Yoshida S, Minowa T, Ohta T (1994). T and R states in the crystals of bacterial L-lactate dehydrogenase reveal the mechanism for allosteric control. *Nat Struct Biol* 1:176–185.
- Fersht A (1999). Structure and mechanism in protein science. New York: W. H. Freeman.

39. Eventoff W, Rossmann MG, Taylor SS, Torff H-J, Meyer H, Keil W, Kiltz H-H (1977). Structural adaptations of lactate dehydrogenase isozymes. *Proc Natl Acad Sci USA* 2677–2681. [VOL]:
40. Des Marais DL, Rausher MD (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*.
41. Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL (2014). An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases *eLife* 3:[PAGE#S].
42. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW (2012). Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet* 8:e1003072
43. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–4410.
44. Williams PD, Pollock DD, Blackburne BP, Goldstein RA (2006). Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol* 2:e69
45. Bridgham JT, Ortlund EA, Thornton JW (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515–519.
46. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
47. Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
48. Gascuel O (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol E* 14:685–695.
49. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *System Biol* 61:539–542.
50. Darriba D, Taboada GL, Doallo R, Posada D (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.
51. Datsenko KA, Wanner BL (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97:6640–6645.
52. Nishiyama M, Birktoft JJ, Beppu T (1993). Alteration of coenzyme specificity of malate dehydrogenase from *Thermus flavus* by site-directed mutagenesis. *J Biol Chem* 268:4656–4660.
53. Reed MC, Lieb A, Nijhout HF (2010) The biological significance of substrate inhibition: a mechanism with diverse functions. *BioEssays* 32:422–429.
54. Kabsch W (2010). XDS. *Acta Cryst D* 66:125–132. 1-8.
55. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Echols N, Headd JJ, Hung L-W, Jain S, Kapral GJ, Kunstleve RWG, McCoy AJ, Moriarty NW, Oeffner RD, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2011). The Phenix software for automated determination of macromolecular structures. *Methods* 55: 94–106.
56. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010). Features and development of Coot. *Acta Cryst D* 66: 486–501.
57. Theobald DL, Wuttke DS (2006). Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proc Natl Acad Sci USA* 103:18521–18527.
58. Theobald DL, Wuttke DS (2008) Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput Biol* 4:e43.
59. Theobald DL, Steindel PA (2012). Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics* 28:1972–1979.
60. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
61. Swint L, Robertson AD (1993). Thermodynamics of unfolding for turkey ovomucoid third domain: thermal and chemical denaturation. *Protein Sci* 2:2037–2049.