Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 3104–3109, April 1997
Evolution

Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase

(isopropylmalate dehydrogenase/NAD/NADP/ancient adaptations)

ANTONY M. DEAN* AND G. BRIAN GOLDING†

*Department of Biological Chemistry, The Chicago Medical School, North Chicago, IL 60064-3095; and †Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada

Communicated by Walter M. Fitch, University of California, Irvine, CA, January 21, 1997 (received for review July 18, 1996)

ABSTRACT **Evolutionary analysis indicates that eubacterial NADP-dependent isocitrate dehydrogenases (EC 1.1.1.42) first evolved from an NAD-dependent precursor about 3.5 billion years ago. Selection in favor of utilizing NADP was probably a result of niche expansion during growth on acetate, where isocitrate dehydrogenase provides 90% of the NADPH necessary for biosynthesis. Amino acids responsible for differing coenzyme specificities were identified from x-ray crystallographic structures of *Escherichia coli* isocitrate dehydrogenase and the distantly related *Thermus thermophilus* NAD-dependent isopropylmalate dehydrogenase. Site-directed mutagenesis at sites lining the coenzyme binding pockets has been used to invert the coenzyme specificities of both enzymes. Reconstructed ancestral sequences indicate that these replacements are ancestral. Hence the adaptive history of molecular evolution is amenable to experimental investigation.**

Phylogenetic reconstructions using molecular sequences form the basis of many modern evolutionary studies. With use of molecules as markers, many insights have been gained and hypotheses tested. However, the lack of a close association between the sequences and their phenotypic expression compels, with few exceptions (1), the adaptive events of molecular evolution to remain invisible. Yet any claim to an understanding of molecular evolution demands that its adaptive history be investigated. Here, we return form and function to a molecular phylogeny and, in so doing, uncover an adaptive event that occurred between 2 and 3.5 billion years ago (2, 3).

Bacteria growing on energy-rich compounds generate NADPH in various catabolic pathways leading to the Krebs cycle and assorted fermentation pathways (4). Those capable of growth on acetate use isocitrate dehydrogenase (IDH; EC 1.1.1.42) in the Krebs cycle to generate 90% of the NADPH necessary for biosynthesis (5, 6). Indeed, bacteria with NAD-dependent IDHs (NAD-IDHs) are incapable of growth on acetate and lack either a respiratory chain or a complete Krebs cycle (7). There appears to be no other major source of NADPH; the flux through NADP-dependent malic enzyme provides approximately 10% of the NADPH (5, 6), whereas transhydrogenase, a membrane bound enzyme that utilizes energy in the form of the protonmotive force to reduce NADP, has less than 5% of the activity of IDH (8–10). Hence, the NADP dependence of eubacterial IDHs appears to be an adaptation to growth on acetate. Our goal is to identify when this event occurred and the amino acid replacements responsible for it.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/943104-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

MATERIALS AND METHODS

Sequence Alignments. Initial alignments obtained using CLUSTAL W (11) were modified based on a knowledge of the high resolution x-ray crystallographic structures of *Escherichia coli* NADP-dependent IDH (NADP-IDH) (12) and *Thermus thermophilus* NAD-dependent isopropylmalate dehydrogenase (NAD-IMDH) (13). Indeed, so divergent are these sequences (between 17% and 24% identical among the four major groups in Fig. 1) that residues known to be critical to substrate binding and catalysis are misaligned for a wide range of gap penalties. Instead, residues critical to substrate binding were identified from high resolution x-ray crystallographic structures of *E. coli* NADP-IDH with bound isocitrate (14). Residues critical to coenzyme specificity were identified from high resolution x-ray crystallographic structures of *E. coli* IDH complexed to NADP (15) and *T. thermophilus* IMDH complexed to NAD (16). Eukaryotic NAD-IDHs and NADP-IDHs were then realigned against this set using residues critical to specificity as key landmarks, and with gaps and insertions introduced into surface loops rather than in the hydrophobic cores of the domains. (The complete alignment of 67 full-length sequences is available from the authors.)

Phylogenetic Reconstructions. Maximum likelihood trees (17) of the decarboxylating dehydrogenases were reconstructed using a modified point accepted mutation matrix (18) with random orders of taxa input. Consensus trees were calculated based on 1,000 neighbor joining trees (19) or 1,000 parsimony trees (20). Bootstrap values and the evaluation of maximum likelihood trees were used to assess the stability of the phylogenies. Ancestral sequences were reconstructed using a maximum likelihood algorithm (21) that assumes the tree given in Fig. 1 and equal rates of replacement among amino acids.

GenBank/EMBL accession numbers for IDHs are as follows: *Anabaena* sp. (PCC 7120), A55591; *Bacillus subtilis*, U05257; *Bos taurus*, X69432, U07980; *Caenorhabditis elegans*, C50F4, F59B8, Z46242; *Escherichia coli*, J02799; *Glycine max*, Q06197, S33612; *Homo sapiens*, S55282, U52144, X69433, Z68907; *Macaca fascicularis*, X74124, X82632, X87172; *Medicago sativa*, M93672, S28423; *Mus musculus*, U51167; *Nicotiana tabacum*, S42892; *Rattus norvegicus*, L35317, X74125; *Saccharomyces cerevisiae*, L26312, M57229, M74131, M95203; *Solanum tuberosum*, X67310, X75638; *Sphingomonas yanoikuyae*, U37523; *Sus scrofa*, M86719; *Thermus aquaticus*, A43934; *Thermus aquaticus thermophilus*, P33197; *Vibrio* sp. ABE-1, D14047.

GenBank/EMBL accession numbers for IMDHs are as follows: *Acremonium chrysogenum*, D50665; *Agrobacterium tumefaciens*, M38670; *Bacillus caldotenax*, X04762; *Bacillus coagulans*, M33099; *Bacillus megaterium*, X65184; *Bacillus*

Abbreviations: IDH, isocitrate dehydrogenase; IMDH, isopropylmalate dehydrogenase; NAD-IMDH, NAD-dependent IMDH; NADP-IDH, NADP-dependent IDH.

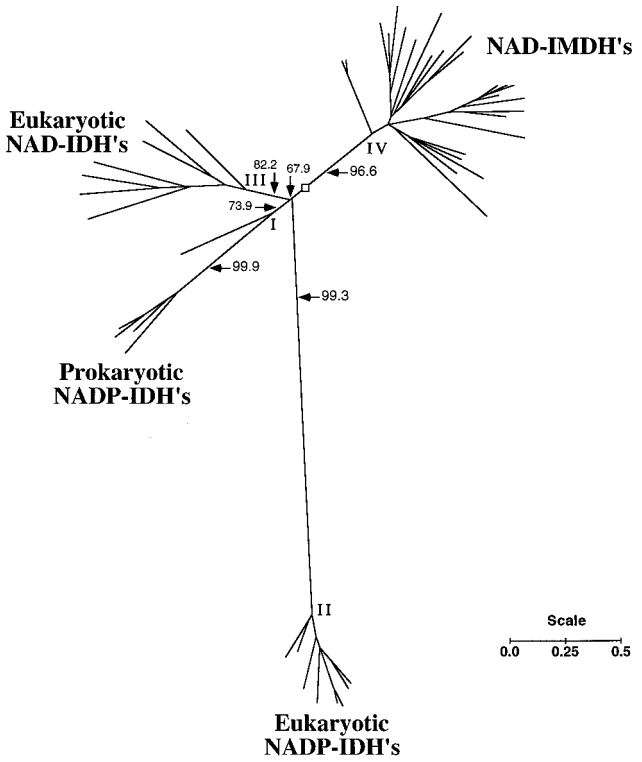

FIG. 1. Unrooted maximum likelihood tree of the decarboxylating dehydrogenases. The numbers adjacent to the arrows are the percentage bootstrap values from 1,000 nearest neighbor joining trees. Roman numerals refer to ancestral nodes, the inferred sequences of which appear in Fig. 3. The square indicates a plausible root for the tree (discussed in the text).

subtilis, P05645; *Bacteroides fragilis*, D45169; *Brassica napus*, X59970; *Buchnera aphidicola*, P48572; *Candida boidinii*, M98832; *Candida maltosa*, P07139; *Clostridium pasteurianum*, P31958; *Debaryomyces occidentalis*, X79823; *Escherichia coli*, P30125; *Haemophilus influenzae*, L45625; *Kluyveromyces lactis*, X55358; *Kluyveromyces marxianus*, X61490; *Lactococcus lactis lactis*, M90761; *Leptospira interrogans*, M59431; *Neisseria lactamica*, S43888; *Neurospora crassa*, U01061; *Pichia angusta*, U00889; *Pichia jadinii*, M16014; *Pichia ohmeri*, Z35101; *Pseudomonas aeruginosa*, U29655; *Saccharomyces cerevisiae*, M12909; *Salmonella typhimurium*, U20795; *Schizosaccharomyces pombe*, M36910; *Spirulina platensis*, M75903; *Thermus aquaticus*, S41223; *Thermus aquaticus thermophilus*, K01444; *Thiobacillus ferrooxidans*, JX0286; *Yarrowia lipolytica*, M37309.

Structural Biology. X-ray crystallographic structures were visualized using a Silicon Graphics (Mountain View, CA) 4D 120/GTX running QUANTA/CHARMM and GRASP software programs. As an aid to aligning the amino acid sequence, the C^{α} backbone trace of each domain of *T. thermophilus* IMDH was independently superimposed by least squares on the corresponding domain of *E. coli* IDH to an rms deviation no greater than 0.42 Å. The nucleotide binding pockets with their respective bound coenzymes were superimposed by least squares to an rms deviation of 0.3 Å using all main chain atoms in loop 2 (Fig. 2, residues Thr-338 through Ala-354) and those of Tyr-391 and Asp-392 common to both IDH and IMDH despite the local difference in secondary structure. Residues in loop 1 (Fig. 2) were not aligned because of obvious and extensive differences in the main chain conformation.

Coordinates for IDH and IMDH structures can be obtained from the Brookhaven National Laboratory Protein Data Bank. Accession numbers for *E. coli* IDH are 3icd (apo-IDH), 5icd (IDH with bound Mg^{2+}-isocitrate), 9icd (IDH with NADP bound), and 1ikb (an inactive pseudo-Michaelis complex of IDH with Ca^{2+}-isocitrate and NADP bound). Accession numbers for *T. aquaticus thermophilus* IMDH are 1ipd (apo-IMDH) and 1hex (IMDH with NAD bound).

RESULTS

Phylogeny of the Decarboxylating Dehydrogenases. IDHs belong to an ancient and divergent family of decarboxylating dehydrogenases (Fig. 1) that includes NAD-IMDH (22), which catalyzes the penultimate step in leucine biosynthesis (23). An evaluation of maximum likelihood trees and the bootstrap values of nearest neighbor joining and maximum parsimony trees indicate that each of the major groups (eukaryotic NAD-IDHs, eukaryotic NADP-IDHs, eubacterial NADP-IDHs, and the NAD-IMDHs) is monophyletic. Some species within each major group are rearranged by the different algorithms, but the trees remain very similar in overall topology.

Both evaluation of the maximum likelihood trees and the bootstrap value of 67.9% from 1,000 nearest neighbor joining trees indicate that the relative branching order of the major groups is not delimited. The highly divergent eukaryotic NADP-IDHs (<17% identical with eubacterial NADP-IDHs)

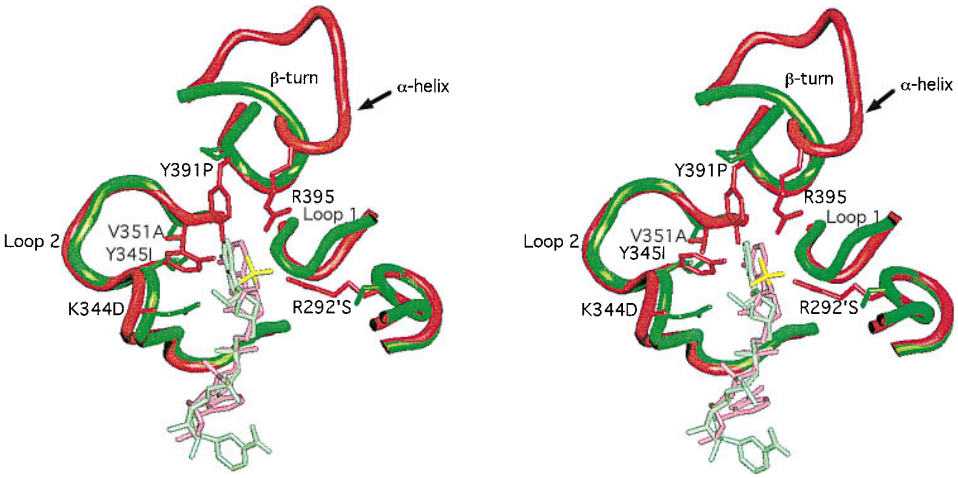

FIG. 2. Stereo view of the *E. coli* NADP-IDH coenzyme binding pocket (15) (red, with NADP in pink and the 2′-phosphate in yellow) superimposed on that of *T. thermophilus* NAD-IMDH (16) (green with NAD in light green). Side chains (*E. coli* IDH numbering) critical to specificity are shown emanating from the backbone worms. The side chain of Lys-344 is disordered in this IDH structure, but it can interact with the 2′-phosphate of NADP.

either are monophyletic with eubacterial NADP-IDHs or branch slightly earlier along the NAD-IMDH limb (as shown). Indeed, the NADP binding pockets of eubacterial NADP-IDHs and eukaryotic NADP-IDHs are sufficiently different that specificity toward NADP may have evolved twice (Fig. 3).

There is no phylogenetic means to identify the root of the tree in the absence of a defined outgroup or a known ancient ancestral gene duplication. However, the common ancestor of all these taxa was undoubtedly an autotroph, capable of synthesizing both leucine and glutamate, and hence having both IMDH and IDH activities. With this in mind, a root (indicated by the square in Fig. 1) has been tentatively placed along the limb joining the IDHs to the IMDHs. This placement, together with the number and complexity of sequence changes, is consistent with the notion that specificity toward isocitrate may have evolved before specificity toward NADP, and that the latter evolved around the time the eukaryotes first appeared.

Determinants of Coenzyme Specificity. Residues critical to coenzyme specificity were identified from high resolution x-ray structures of *E. coli* IDH complexed with NADP (15) and *T. thermophilus* IMDH complexed with NAD (16). These enzymes share a unique nucleotide binding domain. Instead of the $\beta\alpha\beta\alpha\beta$ motif characteristic of the Rossmann fold common to most dehydrogenases (24), the binding pocket is constructed from two loops and an α-helix in *E. coli* IDH, the latter being replaced by a β-turn in *T. thermophilus* IMDH (Fig. 2).

Specificity in *E. coli* IDH is conferred by interactions between Arg-395, Lys-344, Tyr-345, Tyr-391, and Arg-292′ (the prime indicates the second subunit of the homodimer) and the 2′-phosphate of bound NADP (15). These residues are conserved in eubacterial NADP-IDHs (Fig. 3). In *T. thermophilus* IMDH (16), there is no site equivalent to Arg-395 because the α-helix and loop of *E. coli* NADP-IDH is replaced by a β-turn (Fig. 2). Replacements by Ser-292′, Ile-345, and Pro-391 eliminate all favorable interactions with the 2′-phosphate of NADP. Specificity in IMDH is conferred by the rigidly conserved Asp-344, which forms a double hydrogen bond with the 2′- and 3′-hydroxyls of the adenosine ribose of NAD, shifting its position and changing the ribose pucker from C3′-endo to C2′-endo. The shift in the position of the coenzyme, which is also accommodated by replacements Ile-345 and Ala-351, allows an additional H-bond to form between the N2 of the adenine ring and the main chain amide of residue 352 (data not shown). Not only are these movements incompatible with the strong 2′-phosphate interactions seen in NADP-IDH, but the negative charge on Asp-344 will repel NADP from the pocket in much the same way as phosphorylation of Ser-113, or replacement by Asp or Glu, inactivates IDH through repulsion of the γ-carboxylate of isocitrate (14, 16, 25–30).

Inferred Ancestral Sequences. Ancestral character states were reconstructed using maximum likelihood (Fig. 3) rather than parsimony (31), which gives less information of the support for any one amino acid over another. The ancestral sequences were inferred with a maximum likelihood algorithm similar to that of Yang *et al.* (32). Errors in inferred ancestral character states often appear as homoplastic replacements (33). In our case, each inference presented is based on many species, whereas the number of possible character states is large (there are 20 amino acids). This, combined with the slow replacement rate in these proteins, reduces the number of erroneous inferred homoplastic replacements (33).

Ancestral eubacterial NADP-IDHs possess Arg-292, Lys-344, Tyr-345, Lys/Tyr-391, and Arg-395, all of which contact

```
                                    Substrate         Coenzyme
                                  Binding Pocket      Binding Pocket
Sites (E. coli IDH numbering)      11111112233     2 3          3      33           3   33    3
                                   11112563801     9 1          3      44           5   99    9
                                   23599300371     2 7          9      45           1   01    5
NADP-IDH
                                                     loop 1     loop 2                  α-helix
E. coli                            RSNRRRYKDDD     R VGG-IGIAP  HGTAPKYA-----G-QDKVNP  TYDFERLMDGA--------KLLKC
Anabaena sp.                       RSNRRRYKDDD     R VGG-LGMGP  HGTAPKHA-----G-LDRINP  TYDLARLLEPPV-------EPLKC
B. subtilis                        RSNRRRYKDDD     R VGG-IGIAP  HGTAPKYA-----G-LDKVNP  TYDFARLMDGA--------TEVKC
Eubacterial Ancestor    (I)        KSNRRRYKDDD     R VGG-LGLAP  HGTAPKYA-----G-KTSINP  TKDLALCIGG---------AYLKT
                                                     F S   I                           Y   RLLR            --V

H. sapiens (mitochonrial)          KSNRRRYKDDD     S FGS-LGLMT  HGTVTRHYREHQKGRPTSTNP  TKDLAGCIHGLSNVKL-NEHFLNT
S. cerevisiae (cytosolic)          KSNRRRYKDDD     S FGS-LGLMT  HGTVTRHLTDYDKGRETSTNS  TKDLALILGK-----SERSAYVTT
Eukaryotic Ancestor    (II)        KSNRRRYKDDD     S FGS-LGLMT  HGTVTRHYRQHQKGKETSTNP  TKDLALCIGG-------R-AYLTT
                                                                                            IL              V
NAD-IDH
                                                     loop 1     loop 2                  β-turn
H. sapiens (α-subunit)             PSNRRRYKDDD     D IGG-LGVTP  HGTAPDIA-----G-KDMANP  TKDLG--------------GNAKC
S. cerevisiae (α-subunit)          RSNRRRYKDDD     N SAGSLGLTP  HGSAPDIA-----G-QDKANP  TGDLA--------------GTATT
H. sapiens (β-subunit)             KSNRIRYKDNN     N VGG-PGLVA  RNTGKSIA-----N-KNIANP  TPDIG--------------GQGTT
S. cerevisiae (β-subunit)          GSNRARFKDTN     K IGG-PGLVA  RHVGLDIK-----G-QNVANP  TRDIG--------------GSSST
Eukaryotic Ancestor  (III)         RSNRRRYKDDD     N VGG-LGLTP  HGSAPDIA-----G-KNIANP  TPDLG--------------GTATT
                                                           V                                              Y
NAD-IMDH
                                                     loop 1     loop 2                  β-turn
T. thermophilus                    ETLRRRYKDDD     S PGS-LGLLP  HGSAPDIA-----G-KGIANP  PPDLG--------------GSAGT
A. chrysogenum                     ESLRRRYKDDD     D TGT-LGLMP  HGSAPDIS-----G-KGLANP  TGDLG--------------GRATC
S. cerevisiae                      EQLRRRYKDDD     N PGS-LGLLP  HGSAPDLP-------KNKVDP  TGDLG--------------GSNST
B. subtilis                        EKLRRRYKDDD     A TGS-LGMLP  HGSAPDIA-----G-KGMANP  TRDLARSE-----------EFSST
E. coli                            ERLRRRYKDDD     D TGS-MGMLP  GGSAPDIA-----G-KNIANP  TGDLARGA-----------AAVST
Ancestral IMDH         (IV)        ESLRRRYKDDD     N TGS-LGLLP  HGSAPDIA-----G-KGIANP  TPDLG--------------GVAST
                                                     V                                                    T
Engineered Enzymes
                                                     loop 1     loop 2                  α-helix
NAD-IDH (E. coli )                 RSNRRRYKDDD     R VGG-IGIAP  HGTAPDIA-----G-QDKANP  TKDFESLMDGA--------KLLKC
NADP-IMDH (T. thermophilus )       ETLRRRYKDDD     R PGS-LGLLP  HGSAPKYA-----G-KGIVNP  TYDLERLADGA--------GSAGT
```

FIG. 3. Amino acids critical to substrate binding and catalysis and sequences surrounding the coenzyme binding pockets in decarboxylating dehydrogenases from several extant species, from two engineered enzymes, and from inferred ancestral sequences (numbers refer to the nodes in Fig. 1) reconstructed using a maximum likelihood algorithm (21) that assumes the tree given in Fig. 1 and equal rates of replacement among amino acids. Residues in boldface type are critical to specificity. Dashes indicate gaps. Below the most likely inferred ancestral sequences are alternative amino acids with likelihoods >10% of the preferred residue.

the 2′-phosphate of NADP, and Ile/Val-351, which contacts the adenine ring. All inferred ancestral sequences of NAD-IMDH and NAD-IDH possess Asp-344, which H-bonds to the ribose hydroxyls; Ile-345 and Ala-351, which facilitate H-bond formation by allowing the adenosine to shift in the pocket; and Pro-391, which eliminates an H-bond to the 2′-phosphate of NADP. Arg-395, which forms an H-bond to the 2′-phosphate of NADP bound in *E. coli* IDH, appears to be present in several NAD-IMDHs (e.g., *E. coli* and *B. subtilis* in Fig. 3). However, this site is entirely absent in other NAD-dependent enzymes in which the α-helix is replaced by a β-turn. While reconstructions favor Asn replacing Arg-292, this highly solvated surface residue is frequently replaced by other small polar residues that are too short to contact bound coenzymes.

Ancestral eukaryotic NADP-IDHs possess Arg-344, His-345, and Lys-391, all of which might interact with the 2′-phosphate of NADP. However, replacements Ser-292, Leu-395, and Tyr-351 are unlikely to interact with NADP, and in general, the sequences surrounding the eukaryotic nucleotide binding pocket are so divergent that alternative means to stabilize NADP probably operate. Hence, any conclusions regarding the importance and role of these replacements should be treated with great caution.

DISCUSSION

The Evolution of NADP-IDHs. Phylogenetic reconstructions reveal that eubacterial NADP-IDHs are monophyletic, include both Gram-positive and Gram-negative species, and diverge near the branch points of the eukaryotic NAD- and NADP-IDHs. This suggests that NADP dependence evolved early in the history of life, on or about the time that the eukaryotes first appeared, between 2 and 3.5 billion years ago (2, 3).

Limitations of Phylogenetic Analyses. Identifying amino acids responsible for functional differences between similar sequences is trivial. All one needs do is search among those few replacements to identify likely candidates. More sophisticated approaches use maximum parsimony (31) or maximum likelihood (21) to reconstruct ancestral sequences. Nested analyses of variance (34) can also be applied.

However, the problem of identifying the key amino acids becomes intractable by these means when, as in the case of the IDHs, the sequences are highly divergent (35). Under these circumstances, sequence alignments may identify "sites" in one enzyme that have no equivalent in another where local secondary structures differ (e.g., the α-helix and β-turn in Fig. 2). Confusion inevitably arises when sequences sharing little identity, and even encoding different secondary structures, confer the same function (e.g., nucleotide binding pockets of the eubacterial and eukaryotic NADP-IDHs). Even given common secondary structures, critical amino acids need not be conserved because there are frequently various solutions to local steric packing problems (e.g., Ala-351 in the NAD-dependent enzymes is replaced by Val in certain eukaryotic NAD-IMDHs; Fig. 3). Critical amino acids also need not be conserved when particular side chain properties (e.g., a positive charge) are more important than what is actually encoded. Substituting amino acids that are highly conserved in one part of the phylogeny may prove disruptive when they have little or nothing to do with specificity *per se*, and everything to do with an alternative means to stabilize similar secondary structures (e.g., replacements in loop 1 of IDH in Fig. 2 by residues found in NAD-dependent enzymes disrupt the alignment of the domains that comprise the catalytic site, so reducing catalytic efficiency). Finally, homologous sequences in some regions may be so divergent that alignments based solely on maximizing sequence identities become highly dubious (e.g., alignment of the eukaryotic NADP-IDHs with the other enzymes requires a detailed knowledge of crystallographic structures so that functionally critical residues can be identified and aligned).

Determinants of Coenzyme Specificity. Three-dimensional structures provide invaluable information in identifying residues critical to specificity (Fig. 2). The high resolution x-ray structure of the binary complex of *E. coli* IDH with bound NADP (15) reveals that specificity toward NADP is conferred by H-bonds between Arg-395, Lys-344, Tyr-345, Tyr-391, Arg-292′, and the 2′-phosphate of bound coenzyme. These residues are conserved in eubacterial NADP-IDHs (Fig. 3). The high-resolution x-ray structure of the binary complex of *T. thermophilus* IMDH with bound NAD (16) reveals that all potentially favorable interactions with the 2′-phosphate of NADP are eliminated and that the negatively charged Asp-344, which forms a double hydrogen bond with the 2′- and 3′-hydroxyls of the adenosine ribose of NAD, is in a position to repel the 2′-phosphate of NADP. Although Asp-344 is rigidly conserved in all the NAD-dependent decarboxylating dehydrogenases, Ile-345 is sometimes replaced by Leu, whereas Ala-351, which is critical to H-bonding between Asp-344 by allowing the coenzyme to shift position, is replaced by Val in eukaryotic NAD-IMDHs.

Protein Engineering Confirms Structural Analyses. Descriptions of the determinants of specificity based on x-ray structures demand experimental verification. Site-directed mutagenesis has been used to invert the coenzyme specificity of *E. coli* NADP-IDH from a 7,000-fold preference for NADP to a 200-fold preference for NAD (26). Five amino acid replacements (Lys-344 → Asp, Tyr-345 → Ile, Val-351 → Ala, Tyr-391 → Lys, and Arg-395 → Ser) introduced into the nucleotide binding pocket of wild-type *E. coli* IDH cause a shift in preference from NADP to NAD by a factor exceeding 1 million. Two additional replacements at sites remote to the nucleotide binding pocket improve overall performance to a level comparable with the mitochondrial NAD-IDH from yeast (Table 1).

Site-directed mutagenesis has also been used to invert the coenzyme specificity of *T. thermophilus* NAD-IMDH (27). Replacement of the seven residues of the β-turn in IMDH with a 13-residue sequence modeled on the α-helix and loop in *E. coli* NADP-IDH, together with four additional replacements (Ser-292′ → Arg, Asp-344 → Lys, Ile-345 → Tyr, Ala-351 → Val), causes a shift in preference from NAD to NADP by a factor of 100,000 and a modest improvement in performance (Table 1).

Table 1. Kinetic parameters of wild-type and engineered enzymes toward NADP and NAD

		NADP			NAD			Specificity	
Enzyme	Ref.	K_m, μM	k_{cat}, sec^{-1}	k_{cat}/K_m (A), $\mu M^{-1}\cdot sec^{-1}$	K_m, μM	k_{cat}, sec^{-1}	k_{cat}/K_m (B), $\mu M^{-1}\cdot sec^{-1}$	(A)/(B) NADP/NAD	(B)/(A) NAD/NADP
E. coli NADP-IDH	26	17	80.5	4.7	4700	3.22	0.00069	6900	0.00015
S. cerevisiae NAD-IDH	36				210	40	0.190		
Engineered NAD-IDH	37	5800	4.70	0.00081	99	16.2	0.164	0.005	200
T. thermophilus NAD-IMDH	27	1750	0.26	0.00015	12	0.15	0.0125	0.012	80
Engineered NADP-IMDH	27	20	0.39	0.020	25560	0.52	0.00002	1000	0.001

A = $k_{catNADP}/K_{mNADP}$; B = k_{catNAD}/K_{mNAD}.

The experimental demonstration that changes at the sites and secondary structures identified by x-ray crystallography are sufficient to generate changes in specificity is crucial. Parallel work on the substrate specificities of these enzymes has yet to yield significant changes while retaining catalytic efficiency (28–30).

That the coenzyme specificities of the decarboxylating dehydrogenases are determined by residues lining the nucleotide binding pocket demonstrates that the many differences outside the nucleotide binding pockets (these enzymes share only 25% sequence identity) contribute little to discrimination between NAD and NADP. Assuming that a similar number of replacements is involved with determining substrate specificity leaves some 240 replacements that play little or no role in determining specificity, toward either substrate or coenzyme. Many may be neutral, but many others are likely to have been subject to selection. For instance, in engineering the NAD-IDH, two replacements remote from the active site improved activity (but not specificity) 16-fold. Such treadmill replacements, selected at one time or another as evolving populations track environmental changes, do not alter the overall function of enzymes in any significant way.

Existing Sequence Differences Reflect Ancestral Adaptive Events. Ancestral character states were reconstructed using maximum likelihood (Fig. 3). All inferred ancestral sequences of NAD-IMDH and NAD-IDH possess Asp-344, Ile-345, and Ala-351, which were introduced into the engineered *E. coli* NADP-IDH. Pro-391 and Pro-395 were not introduced because eliminating interactions with the 2′-phosphate of NADP does not require engineering of the α-helix to a β-turn. Arg-292 was retained to prevent a loss of catalytic efficiency that accompanied a further improvement in specificity. Hence, the engineered NAD-IDH contains three inferred ancestral replacements.

Ancestral eubacterial NADP-IDHs possess Arg-292, Lys-344, Tyr-345, Lys-391, and Arg-395, all of which were introduced into the engineered IMDH, and Ile-351, which was replaced by Val in the engineered IMDH and which maximum likelihood suggests Val is also highly plausible. Hence, the engineered NADP-IMDH possesses all the replacements associated with ancient adaptive events which collectively inverted the coenzyme specificity of an ancestral eubacterial IDH from NAD to NADP.

Neither of the engineered enzymes can be considered ancestral. Indeed, inspection of the likelihoods across the 250 or so replacements among the 340 sites (excluding additions and deletions) that characterize the differences between any pair of eubacterial NADP-IDHs and NAD-IMDHs suggests that the probability of accurately reconstructing an ancient ancestral enzyme is vanishingly small.

Activity of the Engineered Enzymes. The ratio k_{cat}/K_m is both a measure of enzyme performance and a measure of the degree to which an enzyme stabilizes the transition state (38). The activity of the engineered NAD-IDH is comparable with that of the mitochondrial NAD-IDH of yeast (Table 1) when fully activated by AMP, yet both are rather less active than the wild-type NADP-dependent enzyme. This is expected. The strong ionic H-bonds formed between Arg-292, Lys-344, and Arg-395 and the 2′-phosphate of NADP of wild-type *E. coli* IDH must be replaced by weaker H-bonds between Asp-344 and the two ribose hydroxyls. With less energy available to bind NAD, less energy is available to stabilize the transition state and the k_{cat}/K_m must drop.

Order of Adaptive Substitutions. Phylogenetic analyses provide no clue as to the order in which the replacements accumulated during the shift from NAD to NADP utilization. However, during the mutagenesis experiments with *T. thermophilus* IMDH, the activity toward NADP progressively increased as replacements Arg-292, Asp-344, Ile-345, and Tyr-391 and the Arg-395 of the α-helix were sequentially introduced. This suggests one possible route for the acquisition of specificity toward NADP.

Constraints in Molecular Evolution. In contrast to the IDHs, all naturally occurring IMDHs are NAD-dependent. Hence, NAD utilization is either favored or constrained. The possibility that fitness is maximized by NAD-IMDH contributing to maintenance of optimal an NAD/NADH ratio seems implausible given that the Leu operon is repressed in the presence of excess Leu (37). A more plausible scenario is that NAD utilization is constrained because mutants displaying intermediate specificities toward NAD and NADP have lower overall ($k_{cat}/K_{mNAD} + k_{cat}/K_{mNADP}$) activities, as is the case when engineering these enzymes (26, 27). This hypothesis is directly testable; sites crucial to coenzyme specificity have been identified, site-directed mutagenesis can be used to explore possible intermediate states, and chemostat competition experiments can be used to determine Darwinian fitnesses in a model organism such as *E. coli* (36, 40). Hence, an understanding of the relations between structure and function at the molecular level provides a means to assess the importance of constraints in guiding adaptive evolution.

Conclusion. The engineered NADP-IMDH possesses all the replacements associated with adaptive events that occurred billions of years ago in an ancestral eubacterial IDH and which permitted growth on acetate as the sole source of carbon and energy. We suggest that much of the adaptive history of molecular evolution is similarly amenable to reconstruction through experimental investigation.

We thank Andy Clarke, Dan Hartl, Jack Kirsch, and Phil Bragg for their helpful suggestions. A.M.D. wishes to thank John Gantz for much needed support. This work was financed by Public Health Service Grant GM-48735 from the National Institutes of Health to A.M.D. and by grants from the Natural Sciences and Engineering Research Council (Canada) and Canadian Institute for Advanced Research to G.B.G.

1. Stewart, C. B., Schilling, J. W. & Wilson, A. C. (1987) *Nature (London)* **330,** 401–404.
2. Doolittle, R. F., Feng, D.-F., Tsang, S., Cho, G. & Little, E. (1996) *Science* **271,** 470–477.
3. Knoll, A. H. (1992) *Science* **256,** 622–627.
4. Ingraham, J. L., Maaloe, O. & Neidhardt, F. C. (1983) *Growth of the Bacterial Cell* (Sinauer, Sunderland, MA).
5. Walsh, K. & Koshland, D. E., Jr. (1984) *J. Biol. Chem.* **259,** 9646–9654.
6. Walsh, K. & Koshland, D. E., Jr. (1985) *J. Biol. Chem.* **260,** 8430–8437.
7. Chen, R. & Gadal, P. (1990) *Plant Physiol. Biochem.* **28,** 411–418.
8. Bragg, P. D., Davies, P. L. & Hou, C. (1972) *Biochem. Biophys. Res. Commun.* **47,** 1248–1255.
9. Zahl, K. J., Rose, C. & Hanson, R. L. (1978) *Arch. Biochem. Biophys.* **190,** 598–602.
10. Liang, A. & Houghton, R. L. (1981) *J. Bacteriol.* **146,** 997–1002.
11. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
12. Hurley, J. H., Thorsness, P., Ramalingham, V., Helmers, N., Koshland, D. E., Jr., & Stroud, R. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8635–8639.
13. Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. & Oshima, T. (1991) *J. Mol. Biol.* **222,** 725–738.
14. Hurley, J. H., Dean, A. M., Sohl, J. L., Koshland, D. E., Jr., & Stroud, R. M. *Science* **249,** 1012–1016.
15. Hurley, J. H., Dean, A. M., Koshland, D. E., Jr., & Stroud, R. M. (1991) *Biochemistry* **30,** 8671–8678.
16. Hurley, J. H. & Dean, A. M. (1994) *Structure* **2,** 1007–1016.
17. Adachi, J. & Hasegawa, M. (1992) Molphy: Programs for Molecular Phylogenetics; I. Protml: Maximum Likelihood Inference of Protein Phylogeny (Japanese Institute of Statistical Mathematics, Tokyo), *Comput. Sci. Monogr.* **27.**
18. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) *Comput. Appl. Biosci.* **8,** 275–282.
19. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.

20. Felsenstein, J. (1994) PHYLIP (University of Washington, Seattle, WA), Version 3.5.
21. Felsenstein, J. (1981) *J. Mol. Evol.* **17,** 368–376.
22. Thorsness, P. E. & Koshland, D. E., Jr. (1987) *J. Biol. Chem.* **264,** 10422–10425.
23. Burns, R. O., Umbarger, H. E. & Gross, S. R. (1963) *Biochemistry* **2,** 1053–1057.
24. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) *Nature (London)* **250,** 194–199.
25. Dean, A. M. & Koshland, D. E., Jr. (1990) *Science* **249,** 1044–1046.
26. Chen, R., Greer, A. & Dean, A. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 11666–11670.
27. Chen, R., Greer, A. & Dean, A. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12171–12176.
28. Dean, A. M. & Dvorak, L. (1995) *Protein Sci.* **4,** 2156–2167.
29. Chen, R., Grobler, J. A., Hurley, J. H. & Dean, A. M. (1996) *Protein Sci.* **5,** 287–295.
30. Dean, A. M., Shiau, A. K. & Koshland, D. E., Jr. (1996) *Protein Sci.* **5,** 341–347.
31. Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. (1995) *Nature (London)* **374,** 57–59.
32. Yang, Z., Kumar, S. & Nei, M. (1995) *Genetics* **141,** 1641–1650.
33. Frumhoff, P. C. & Reeve, H. K. (1994) *Evolution* **48,** 172–180.
34. Templeton, A. R., Boerwinkle, E. & Sing, C. F. (1987) *Genetics* **117,** 343–351.
35. McClure, M. A., Vasi, T. K. & Fitch, W. M. (1994) *Mol. Biol. Evol.* **11,** 571–592.
36. Dean, A. M. (1995) *Genetics* **139,** 19–33.
37. Cupp, J. R. & McAlister-Henn, L. (1993) *Biochemistry* **32,** 9323–9328.
38. Hackney, D. D. (1990) *Enzymes* **19,** 1–34.
39. Umbarger, H. E. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. Neidhardt, F. (Am. Soc. Microbiol., Washington, DC), pp. 352–367.
40. Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. (1987) *Genetics* **115,** 25–31.