Protein Science (2000), *9*:2344–2353. Cambridge University Press. Printed in the USA.

Functional prediction: Identification of protein orthologs and paralogs

RIDONG CHEN AND SOON-SEOG JEONG

Department of Biochemistry, College of Medicine, University of Saskatchewan, 107 Wiggins Road, Saskatoon SK S7N 5E5, Canada

(RECEIVED June 6, 2000; FINAL REVISION September 10, 2000; ACCEPTED October 3, 2000)

Abstract

Orthologs typically retain the same function in the course of evolution. Using β-decarboxylating dehydrogenase family as a model, we demonstrate that orthologs can be confidently identified. The strategy is based on our recent findings that substitutions of only a few amino acid residues in these enzymes are sufficient to exchange substrate and coenzyme specificities. Hence, the few major specificity determinants can serve as reliable markers for determining orthologous or paralogous relationships. The power of this approach has been demonstrated by correcting similarity-based functional misassignment and discovering new genes and related pathways, and should be broadly applicable to other enzyme families.

Keywords: homoisocitrate dehydrogenase; isocitrate dehydrogenase; isopropylmalate dehydrogenase; tartrate dehydrogenase

The genomes of *Escherichia coli*, yeast, worm, and 300 other microbial species have been completely sequenced. The sequencing of the *Arabidopsis* and human genomes is scheduled to be completed by the end of the year 2000 and 2003, respectively, and the sequence of many other genomes may become available in the coming decade. The emergence of genomics has changed the way by which genes are discovered and isolated. In the past, a peptide sequence or phenotype has lead to the isolation of a new gene. Now an investigator begins with the sequence of a key gene and searches for homologous genes in an organism of interest (Henikoff et al., 1997). Genomics, as a powerful approach, will certainly have tremendous impact on both fundamental and applied research. However, the extent to which using genomic approaches for gene discovery actually succeeds will depend on how accurately the function of a new protein can be defined without any a priori functional knowledge.

Over a period of more than three billion years, a large variety of protein molecules has evolved to run the complex machinery of the present-day cells and organisms. The relationships between contemporary genes from different genomes are represented as a system of homologous families that include both orthologs and paralogs. Orthologs are genes in different species that evolved from a common ancestral gene by speciation and retained the same function in the course of evolution. By contrast, paralogs are genes related by duplication within an organism and have evolved a related but different function (Tatusov et al., 1997). Protein families are further grouped into superfamilies that are usually homologous, and folds, and classes that share common structural features but are not necessarily homologous (Fig. 1). Hence, a powerful approach to predicting the exact function of a new protein is to use bioinformatics (e.g., BLAST, PSI-BLAST, and domain database searches) to find its characterized orthologs (Tatusov et al., 1997; Hofmann, 1998). Nevertheless, with incompletely sequenced genomes or large phylogenetic distances, there is always the chance that the best match in a database hit is just a well-conserved paralog (Tatusov et al., 1997; Hofmann, 1998). The analysis is further complicated when the divergence between orthologous genes approaches and even exceeds the level of divergence among paralogs within a species. Moreover, a significant fraction of functional annotations in databases is wrong or dubious (Henikoff et al., 1997; Tatusov et al., 1997; Hofmann, 1998). In this study, using the β-decarboxylating dehydrogenase family as a model, we demonstrate that with insight into how distinct functions of orthologs and paralogs are conferred and evolved, it is feasible to identify orthologs with certainty.

Reprint requests to: Ridong Chen, Monsanto Company, O3C, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167; e-mail: ridong.chen@monsanto.com.

Abbreviations: HDH, homoisocitrate dehydrogenase; IDH, isocitrate dehydrogenase; IMDH, isopropylmalate dehydrogenase; TDH, tartrate dehydrogenase.

Results and discussion

Evolution of β-decarboxylating dehydrogenase gene family

β-Decarboxylating dehydrogenases are a family of bifunctional enzymes that catalyze the Mg^{2+}- and $NAD(P)^{+}$-dependent de-

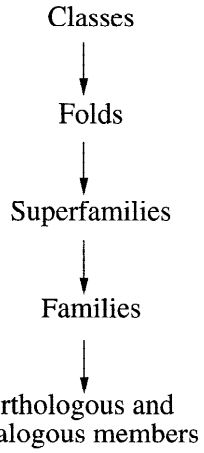

Fig. 1. Hierarchy in protein taxonomy. Enzyme family groups the sequences that are clearly related and catalyze chemically equivalent reaction(s). Superfamily groups enzyme families that share a common overall 3D structure and some catalytic elements.

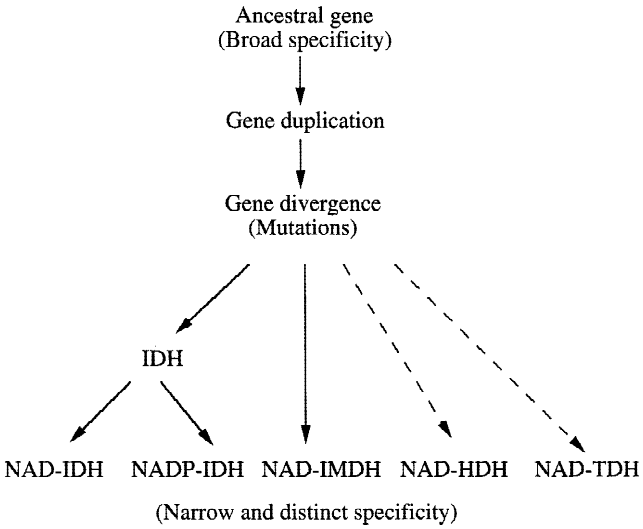

Fig. 2. Evolution of the β-decarboxylating dehydrogenase gene family. Dash line indicates that the progenitor enzymes for NAD-HDH and NAD-TDH are uncertain.

hydrogenation at C2 followed by their Mg^{2+}-dependent decarboxylation at C3 of (2R, 3S) 2-hydroxy acids:

$$\text{2-Hydroxy acid} + \text{NAD(P)}^+ \overset{Mg^{2+}}{\rightleftharpoons} \text{2-Oxalo acid} + \text{NAD(P)H} + H^+$$

$$\text{2-Oxalo acid} + H^+ \overset{Mg^{2+}}{\rightleftharpoons} \text{2-Keto acid} + CO_2.$$

Three orthologs have been identified so far: NAD-dependent isocitrate dehydrogenase (EC 1.1.1.41, NAD-IDH), NADP-dependent isocitrate dehydrogenase (EC 1.1.1.42, NADP-IDH), and NAD-dependent isopropylmalate dehydrogenase (NAD-IMDH, EC 1.1.1. 85). NAD-IDH is limited to eukaryotic organisms and participates in the supply of NADH used for respiratory ATP production in mitochondria, while NADP-IDH is present ubiquitously in both prokaryotes and eukaryotes and involved in the generation of both NADPH and α-ketoglutarate for biosynthetic pathways (Chen & Gadal, 1990). NAD-IMDH is found in bacteria, fungi, and plants and catalyzes the third step of the pathway for leucine biosynthesis (Imada et al., 1991). NADP-IDH and NAD-IMDH are homodimers, while all the NAD-IDHs purified so far have a hetero-oligomeric structure (Chen & Gadal, 1990). The crystal structures of NADP-IDH from *E. coli* and NAD-IMDH from *Thermus thermophilus* have been solved (Hurley et al., 1989; Imada et al., 1991). Both enzymes share a common protein fold that lacks the $\beta\alpha\beta\alpha\beta$ motif characteristic of the nucleotide binding Rossmann fold (Rossmann et al., 1974).

NAD-IDH, NADP-IDH, and NAD-IMDH catalyze different reactions and fulfill distinct biological roles. However, phylogenic analyses indicate that these enzymes divergently evolved from a common ancestral gene present in a progenitor of all extant organisms (Chen et al., 1997b; Fig. 2). The gene might encode the primitive enzyme that possessed a very broad specificity, permitting it to react with a wide range of related substrates that shared a common 2-R malate moiety. This would maximize the catalytic versatility of an ancestral cell that functioned with limited genetic information and enzyme resources (Jensen, 1976). The ancestral gene was then duplicated and acquired additional genetic information. Copies of the gene diverged via mutational modification, giving rise to contemporary enzymes with strict substrate and coenzyme specificity. It has been suggested that specificity toward isocitrate may have evolved before specificity toward NADP, and that the latter evolved around the time the eukaryotes first appeared (Chen et al., 1997b). Specialization of gene functions presumably allowed the improvement of metabolic efficiency and the evolution of new biochemical pathways (Jensen, 1976).

Determinants of substrate and coenzyme specificity

Over evolutionary time, a large number of sequence differences have accumulated among the members of the β-decarboxylating dehydrogenase family. The present-day phylogenies are highly divergent, and the divergence between orthologous genes approaches and even exceeds the level of divergence among paralogs within a species. Thus, it is not reliable to make assignment of new protein sequences to a particular ortholog based on the best BLAST hit or similarity searches. In fact, residues critical to substrate and coenzyme binding and catalysis can rarely be aligned properly using BLAST or CLUSTAL W. This is especially the case when the sequence of *E. coli* NADP-IDH is compared to eukaryotic NADP-IDHs, which share no significant identities (<17%). The substrates for isocitrate and isopropylmalate dehydrogenases, isocitrate, and isopropylmalate are structurally similar, i.e., $^-OOC(HO)CHCH(X)COO^-$, where X represents the γ-moiety: the $-CH_2COO^-$ of isocitrate and the $-CH(CH_3)_2$ of isopropylmalate (Fig. 3). Coenzyme NADP differs from NAD only by a phosphate group esterified at the 2′C of the adenosine ribose. Presumably, distinct functions of NAD-IDH, NADP-IDH, and NAD-IMDH are conferred by their ability to recognize between alternative substrates and coenzymes. Using protein engineering, we have demonstrated that out of hundreds of amino acid substitutions accumulated in *E. coli* NADP-IDH and *T. thermophilus* NAD-IMDH only a few are involved in specificity determination (Chen et al., 1995, 1996a, 1996b, 1997a; Dean & Dvorak, 1995; Dean et al., 1996; Hurley et al., 1996; Yaoi et al., 1997; Chen, 1999).

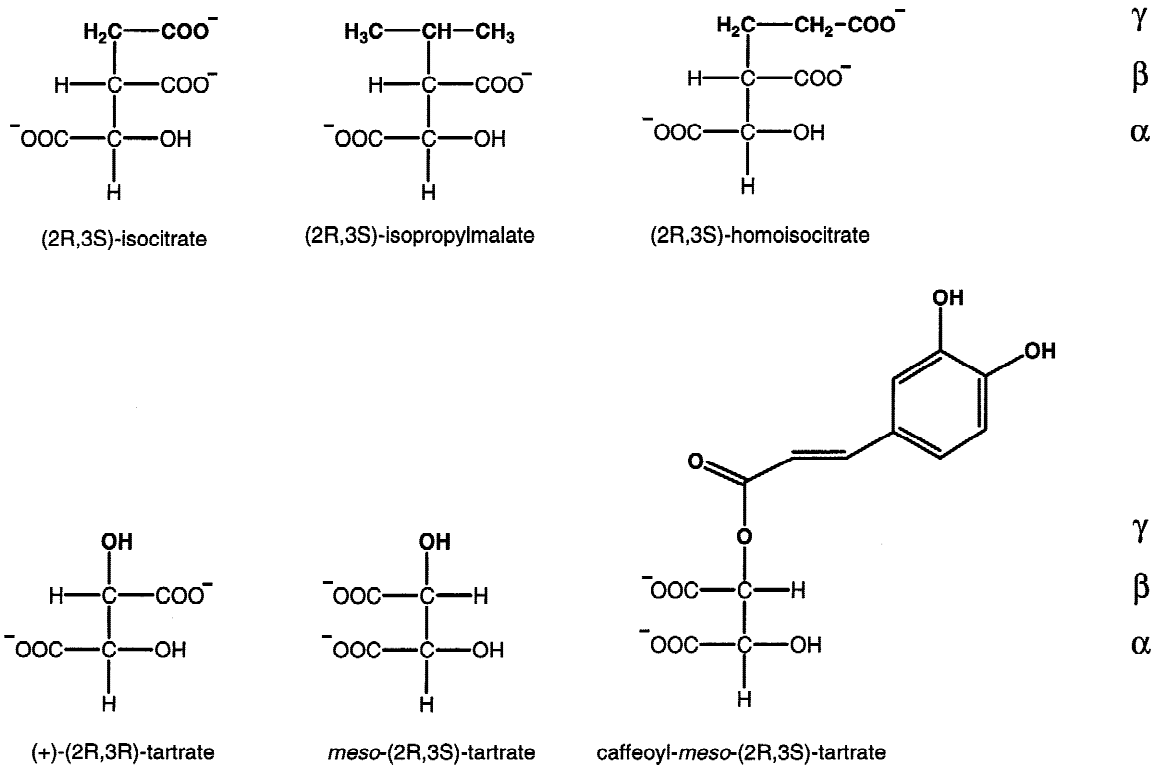

Fig. 3. Structures of substrates for β-decarboxylating dehydrogenases. α, β refers to the α- and β-carboxyl group, respectively. The unique γ-moieties are presented in bold.

In *E. coli* NADP-IDH, all the residues involved in binding and catalysis have been identified (Fig. 4A) (Stoddard et al., 1993; Bolduc et al., 1995; Chen et al., 1995, 1996a, 1996b, 1997a; Dean et al., 1996; Dean & Dvorak, 1995; Hurley et al., 1989, 1991, 1996; Yaoi et al., 1997; Chen, 1999). The true substrate for the NADP-IDH is a Mg^{2+}-isocitrate complex that binds in a pocket, formed from residues donated from both monomers. The amino acid residues Ser113 and Asn115 on the helix d, and the γ-carboxylate of isocitrate are the major determinants of substrate specificity (Fig. 4A) (Chen et al., 1996b; Dean et al., 1996). Ser113 forms a hydrogen bond with the γ-carboxylate of bound isocitrate that, in turn, forms a salt bridge to the nicotinamide ring of the coenzyme (Fig. 4B). Asn115 also interacts with the γ-carboxylate of bound isocitrate. The interaction has a nonideal hydrogen bond angle and is probably electrostatic in nature. This residue also is coordinated to the amide of NADP in the Michaelis complex. The polar environment formed with Ser113 and Asn115 is not compatible with the hydrophobic γ-isopropyl group of bound isopropylmalate. More importantly, this γ-isopropyl group is unable to form the binding site for the nicotinamide ring that is critical for hydride transfer. Consequently, in spite of the structural similarity of isocitrate and isopropylmalate, IDH does not catalyze a reaction with isopropylmalate (with reduced catalytic efficiency by a factor of at least greater than 10^8; R. Chen & S.-S. Jeong, unpubl. results). Coenzyme specificity in the NADP-IDH is conferred by hydrogen bonds between the side chains of Arg395, Tyr345, Tyr391, and Arg292′ and the 2′-phosphate of bound NADP (Fig. 5A). The side chains of Lys344 may ion pair with the 2′-phosphate. These interactions cannot be formed with the 2′-hydroxyl of NAD. Hence, this enzyme is highly specific for NADP. Calculated as the ratio of k_{cat}/K_m, the enzyme displays a 7,000-fold preference for NADP over NAD (Chen et al., 1995).

Despite sharing only 25% sequence identity, *T. thermophilus* NAD-IMDH and *E. coli* NADP-IDH share a common protein fold and their tertiary structures can be superimposed (Imada et al., 1991). The amino acid residues involved in binding the 2R-malate core, common to 2R,3S-isocitrate and 2R,3S-isopropylmalate, are identical (Fig. 4C). However, the two enzymes differ in the amino acid residues involved in binding the γ-moieties of substrates. In NAD-IMDH, Asn115 is replaced by Leu90, which forms hydrophobic interactions with the γ-isopropyl group of bound isopropylmalate (Fig. 4C; Table 1). In addition, the helix d begins three amino acids earlier with the consequence that the side chain of Glu87 occupies a position in close proximity to both Ser113 and the γ-carboxylate of isocitrate. The available structural data and the kinetic results suggest that the carboxyl group of Glu87 interacts with the nicotinamide ring and helps stabilize the Michaelis complex, mimicking the role played by the γ-carboxylate of the bound isocitrate in NADP-IDH (Fig. 4D) (Dean & Dvorak, 1995). Meanwhile, this negatively-charged residue lies close to an analogous position of the γ-moiety of bound isocitrate and thus would electrostatically repel this molecule as a substrate. Indeed, no activity is detectable with isocitrate (Dean & Dvorak, 1995). Consequently, Glu87 and Leu90 are major determinants of substrate specificity in NAD-IMDH (Dean & Dvorak, 1995; Hurley et al., 1996). The coenzyme specificity toward NAD is mainly conferred by Asp278, which forms a double hydrogen bond with the 2′- and 3′-hydroxyls of the adenosine ribose of NAD (Fig. 5B) (Chen

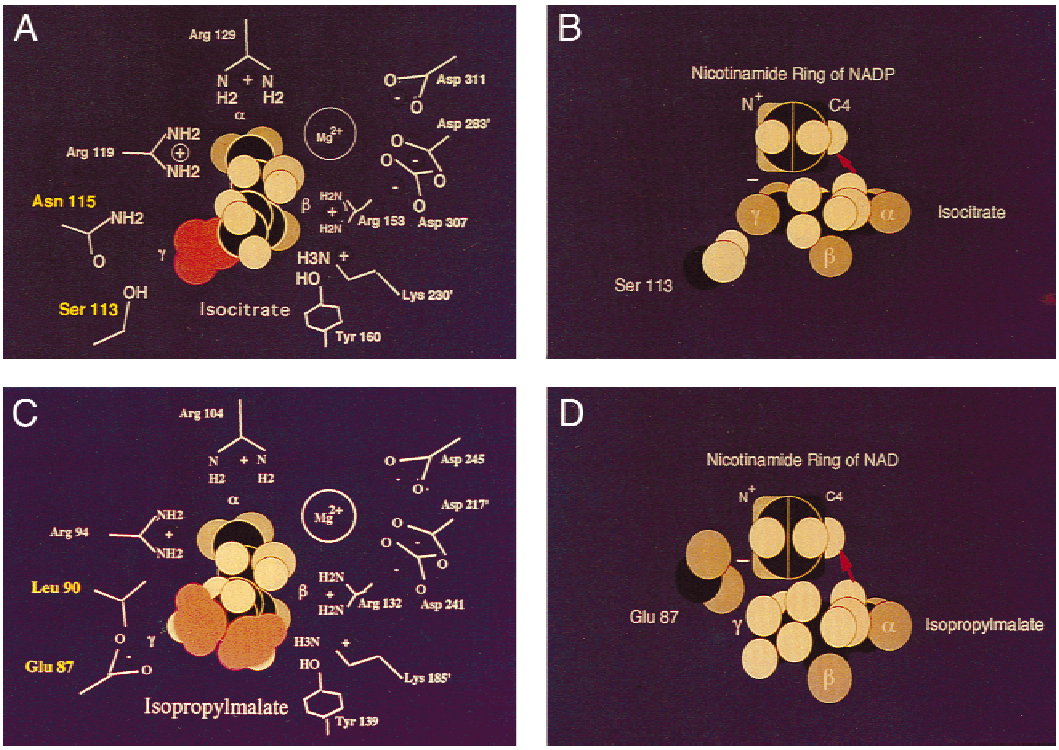

Fig. 4. Schematic diagram of (**A**) the active site of the *E. coli* NADP-IDH with bound 2R, 3S-isocitrate and (**B**) the active site of *T. thermophilus* NAD-IMDH with bound 2R, 3S-isopropylmalate. A prime indicates the residues donated from the second subunit. α and β refer to the α- and β-carboxyl groups, respectively. The major determinants of substrate specificity are residues Ser113 and Asn115 in the NADP-IDH (yellow), and Glu87 and Leu90 in the NAD-IMDH (yellow), which interact with the unique γ-moieties of the substrates (pink). Model for the catalytic complex of the *E. coli* NADP-IDH with isocitrate and NADP bound (**C**) and the *T. thermophilus* NAD-IMDH with isopropylmalate and NAD bound (**D**). The arrow shows how C_4 of nicotinamide ring is positioned to receive the α-carbon hydride of substrates. Both schematic drawings were modified from Dean and Dvorak (1995).

et al., 1997a). Meanwhile, this negatively-charged residue repels the 2′-phosphate of NADP through electrostatic repulsion. As such, this enzyme is 100-fold more active with NAD than with NADP (Chen et al., 1997a).

The roles of the major specificity determinants have been confirmed by site-directed mutagenesis and protein engineering. The S113E and N115L mutants of the NADP-IDH and E87G and L90N mutants of the NAD-IMDH dramatically shifted the specificity toward alternative substrate binding and catalysis (Dean & Dvorak, 1995; Chen et al., 1996b; Dean et al., 1996). More strikingly, guided by X-ray crystal structures and molecular modeling, we introduced seven amino acid substitutions, K344D, Y345I, V351A, Y391K, R395S, C332Y, and C201M, in the *E. coli* NADP-IDH that converted the coenzyme specificity from a 7,000-fold preference for NADP to 200-fold preference for NAD. The kinetic parameters and coenzyme preference of the engineered enzyme compared favorably with, and even exceeded those of natural NAD-IDH from various sources (Chen et al., 1995, 1997a; Hurley et al., 1996; Chen, 1999). Conversely, a NADP-IMDH that is not found naturally was created from *T. thermophilus* NAD-IMDH. This contains five amino acid substitutions, D278K, I279Y, A285V, P335Y, S226R, and an α-helix and loop that replaces the original β-turn. The engineered enzyme is twice as active as the wild-type with the coenzyme specificity inverted from a 100-fold preference for NAD to a 1,000-fold preference for NADP (Chen et al., 1996a; Chen, 1999). The successful protein engineering clearly demonstrate that the substrate and coenzyme specificities in the β-decarboxylating dehydrogenases are principally determined by a few amino acid residues.

Markers for orthologs and paralogs

The sequences of the NADP-IDHs and NAD-IMDHs from other species were compared with that of *E. coli* NADP-IDH and *T. thermophylus* NAD-IMDH (Table 1). It is noteworthy that the NADP-IDH/NAD-IMDH phylogenies are so divergent that correct alignment can be made only based on detailed knowledge of the X-ray crystallographic structures of *E. coli* NADP-IDH and *T. thermophilus* NAD-IMDH (Dean & Dvorak, 1995). Both substrate specificity determinants, Ser113 and Asn115, identified in *E. coli* NADP-IDH are absolutely conserved in other NADP-IDHs. The coenzyme specificity determinants, Lys344 and Tyr345, are present in all the bacterial NADP-specific enzymes. However, these residues are replaced with the positively-charged residue Arg and His respectively in highly divergent eucaryotic NADP-IDHs (Table 1). Nevertheless, site-directed mutagenesis analysis of the *Arabidopsis* NADP-IDH confirms that both the Arg and His residues are the equivalent coenzyme specificity determinants in the eucaryotic enzymes, which electrostatically interact with the bound NADP (R. Chen & L. Wan, unpubl. results). As expected, Ser113 and Asn115 are replaced by Glu and Leu, respectively, while Lys344 and Tyr345 are substituted by Asp and Ile or Leu in all known

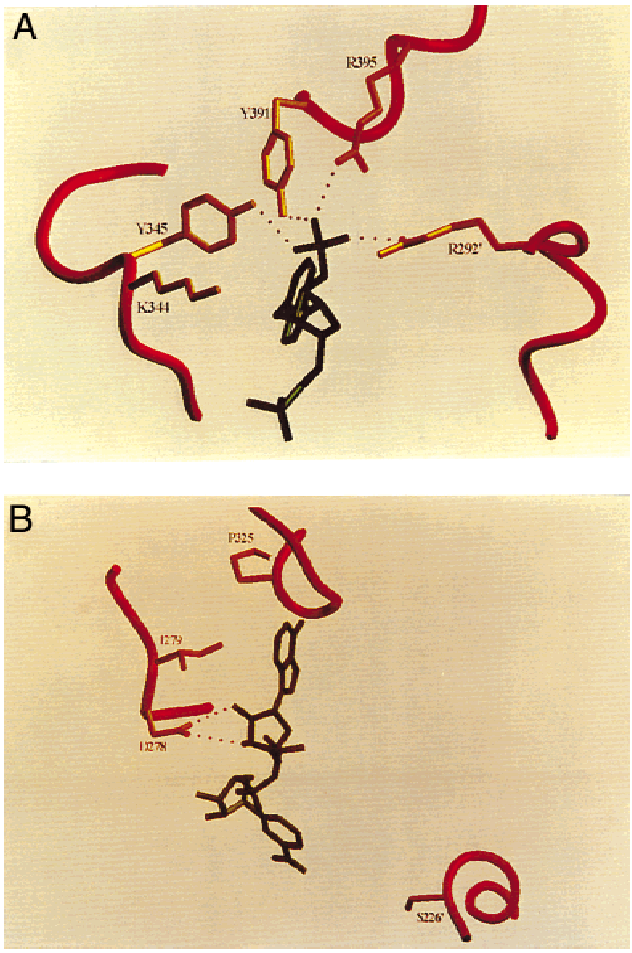

Fig. 5. A: Interaction of the nucleotide-binding pocket of the *E. coli* NADP-IDH (white) with NADP (green). **B:** *T. thermophilus* NAD-IMDH (white) with NAD (green). NADP-IDH numbering is used in both figures. The active site residues of the enzymes are in yellow and the 2′-phosphate of NADP and 2′ and 3′-hydroxyl groups of NAD in red. It is noted that due to the difference of the local secondary structures, the NAD-IMDH does not have the counterpart of the Arg395 found in the NADP-IDH.

NAD-IMDHs (Table 1). Hence, these major specificity determinants can be used as reliable landmarks for predicting the substrate and coenzyme specificities of new sequences belonging to the isocitrate and isopropylmalate dehydrogenases.

No crystal structure is available for NAD-IDHs. Different from NADP-IDHs and NAD-IMDHs, which are homodimers, all the NAD-IDHs purified so far have a hetero-oligomeric structure (Chen & Gadal, 1990). Yeast NAD-IDH exists as a hetero-octamer consisting of four copies of the regulatory subunit (NAD-IDH1) and four copies of the catalytic subunit (NAD-IDH2) (Keys & McAlister-Henn, 1990), while mammalian NAD-IDHs consist of three different subunits (α, β, and γ) having the tetrameric form of $\alpha_2\beta\gamma$ (Ramachandran & Colman, 1983; Nichols et al., 1995). Catalytic residues are located in the α-subunit (Cupp & McAlister-Henn, 1993). Site-directed mutagenesis analysis suggests that the yeast NAD-IDH has an active center formed from two of the multiple subunits, which is essentially identical to that of *E. coli* NADP-IDH (Cupp & McAlister-Henn, 1993). Structure-based alignment shows that the catalytic subunits of NAD-IDHs retain all the active residues involved in Mg^{2+}-isocitrate binding and catalysis (Table 1). Similar to NAD-IMDH, the residues Lys344 and Tyr345, which interact with NADP in *E. coli* NADP-IDH, are replaced by Asp and Ile in these NAD-IDH proteins (Table 1). This is consistent with their coenzyme preference toward NAD. Most of the active residues, including the residues equivalent to Ser113 and Asn115 of *E. coli* NADP-IDH, are found in the regulatory subunits of both yeast and mammalian NAD-IDHs. However, both Asp307 and Asp311, which coordinate the Mg^{2+} of Mg^{2+}-isocitrate in the *E. coli* NADP-IDH, are substituted by Asn, Ser, or Thr in these sequences (Fig. 4A; Table 1). This replacement should abolish the binding with Mg^{2+}-isocitrate, the true substrate for IDHs. Indeed, only two Mg^{2+}-isocitrate sites per $\alpha_2\beta\gamma$ tetramer of yeast NAD-IDH have been observed (Rutter & Denton, 1989). Meanwhile, the Arg129 residue, which binds via ionic hydrogen bonds to the α-carboxylate of bound isocitrate in the *E. coli* NADP-IDH, is replaced by a hydrophobic branched-chain residue such as Leu, Val, or Ile (Fig. 4A; Table 1). This should significantly weaken the binding affinity with isocitrate. The substitutions of these important residues agree with the regulatory roles of the subunits and allow us to differentiate them from catalytic subunits. The residue Asp278 identified in the *T. thermophilus* NAD-IMDH is conserved in yeast IDH1 but is replaced by polar residues Gln or Ser in the mammalian β and γ subunits (Table 1). Thus, identifying the critical specificity determinants is sufficient to define the function of NAD-IDH proteins as well.

Correcting functional misassignment

Insight into the structural and functional determinants for enzyme specificity allowed us to access functional assignment of β-decarboxylating dehydrogenase sequences. It shows that a significant fraction of the cDNA or genomic sequences, which have been published or deposited in databases, are incorrectly annotated (Table 2). Case studies are presented here to demonstrate that correct function can be assigned with certainty.

The sequence (gene number: Aq1512) from the *Aquifex aeolicus* genome sequencing project (GenBank accession number: AE000744), was predicted to encode a NADP-IDH protein of 426 amino acid residues (Table 1) (Deckert et al., 1998). All of the substrate binding and catalytic residues identified in *E. coli* NADP-IDH are conserved in this protein, including Ser113 and Asn115, which are the major determinants of specificity toward isocitrate. In contrast, the residues Lys344 and Tyr345 interacting with NADP in the *E. coli* NADP-IDH are replaced by Asp and Ile, as seen in NAD-IMDH. This observation allows us to assign the function of the *A. aeolicus* protein as NAD-IDH. The coding region of the genomic DNA was subcloned and the enzyme was expressed in an *idh*$^-$ strain of *E. coli*. As expected, the enzyme is NAD-dependent (Table 3). Calculated as the ratio of k_{cat}/K_m, this thermophilic enzyme favors NAD over NADP by a factor of 86 at 65 °C. These results demonstrate that, similar to NADP-IDH, NAD-IDH is present in both procaryotic and eucaryotic organisms.

The potato "NAD-IMDH" clone (assession number P29696; Table 1) was isolated during a differential screen of a potato leaf library designed to recover genes induced upon tuberisation (Jackson et al., 1993). A database search revealed that the deduced amino acid sequences of the enzyme showed 33, 27, and 32% identity to the NAD-IMDH from rape, yeast, and *Bacillus subtilis*, respectively. The transit peptide of the rape NAD-IMDH has been shown to direct chloroplastic uptake (Ellerstrom et al., 1992). There

Table 1. *Structure-based alignment of β-decarboxylating dehydrogenase sequences*[a]

Enzymes	Sequences		
NADP-IDH	113 115 119 129	307 311	344 345
E. coli	IR**S**L**N**VAL**R**QELDLYICL**R**P	NLNG**D**YIS**D**A	HGTA-P**KY**AG
T. thermophilus	EK**S**A**N**VTL**R**KLFETYANV**R**P	NMNG**D**ILS**D**L	HGSA-P**KY**AG
A. fulgidus	YR**S**L**N**VTI**R**QVLDLYANV**R**P	NLNG**D**YLS**D**A	HGSA-P**KY**AG
S. cerevisiae	WK**S**P**N**GTI**R**NILGGTV-F**R**E	NLYG**D**ILS**D**I	HGTV-T**RH**FR
H. sapiens	WK**S**P**N**GTI**R**NILGGTV-F**R**E	NYDG**D**VQS**D**I	HGTV-T**RH**YR
A. thaliana	WR**S**P**N**GTI**R**NILNGTV-F**R**E	NYDG**D**VQS**D**F	HGTV-T**RH**FR
NAD-IMDH	87 90 94 104	241 245	278 279
T. thermophilus	**E**TG-**L**LSL**R**KSQDLFANL**R**P	NIFG**D**ILS**D**L	HGSA-P**DI**AG
E. coli	**E**RGA**L**LPI**R**KHFKLFSNL**R**P	NLFG**D**ILS**D**E	AGGSAP**DI**AG
S. cerevisiae	**E**QG-**L**LKI**R**KELQLYANL**R**P	NLYG**D**ILS**D**E	HGSA-P**DL**PK
B. napus	**E**TG-**L**LQL**R**AGLKVFANL**R**P	NIFG**D**ILS**D**E	HGSA-P**DI**AG
NAD-IDH			
Catalytic subunit			
S. cerevisiae IDH2	HR**S**L**N**LTL**R**KTFGLFANV**R**P	NLYG**D**ILS**D**L	HGSA-P**DI**AG
K. lactis IDH2	HR**S**L**N**LTL**R**KTFGLFANV**R**P	NLYG**D**ILS**D**L	HGSA-P**DI**AG
M. fascicularis α	PR**S**M**N**LLL**R**KTFDLYANV**R**P	NLYG**D**ILS**D**L	HGTA-P**DI**AG
Regulatory subunit			
S. cerevisiae IDH1	HG**S**L**N**VAL**R**KQLDIYANV**A**L	SMYG**T**ILG**NI**	HVGL–**DI**KG
K. lactis IDH1	HG**S**L**N**VAL**R**KQLDIFANV**A**L	NLYG**S**ILG**NI**	HVGL–**DI**KG
M. fascicularis β	LA**S**Y**D**MRL**R**RKLDLFANV**V**H	NLYG**N**IID**N**L	HPFA–**QA**VG
M. fascicularis γ	HK**S**R**N**NIL**R**TSLDLYANV**I**H	NLYG**N**IVN**N**V	RNT-GK**SI**AN
Misannotated sequence			
A. aeolicus "NADP-IDH"	VR**S**I**N**SAL**R**RAFDYYSAV**R**P	NLNG**D**YVS**D**L	HGTA-W**DI**AG
Potato "NAD-IMDH"	HR**S**L**N**LTL**R**KELNLYANV**R**P	NLYG**D**IIS**D**L	HGSA-P**DI**AG
New gene: NAD-HDH			
S. cerevisiae "IMDH"	**Y**SSP**I**VAL**R**REFDQYVNI**R**P	NLYG**D**ILS**D**L	HGSA-P**DI**FG
"NAD-TDH"			
P. putida	**L**WGS**L**LKF**R**REFDQYVNI**R**P	NLFG**D**ILS**D**L	HGSA-P**DI**FG
E. coli	**L**WGS**L**LKF**R**REFDQYVNL**R**P	NLFG**D**ILS**D**L	HGSA-P**DI**YG

[a]Numbers are for the residues in bold for the first enzyme in each group.

were some homologies between the N-terminal sequences of the potato and rape proteins. Consequently, the gene product has been reported as chloroplastic NAD-IMDH (Jackson et al., 1993). Structure-based alignment shows that the equivalent Ser113 and Asn115 residues are both found in the sequence (Table 1), indicating that the enzyme belongs to IDH but not IMDH. Like NAD-IMDHs and yeast and mammalian NAD-IDHs, the Lys344 and Tyr345 residues found in *E. coli* NADP-IDH are replaced by Asp and Ile (Table 1). Hence, this IDH is NAD-dependent. Since all the active residues involved in Mg^{2+}-isocitrate binding and catalysis are found in the sequence, this must be a catalytic subunit of a NAD-IDH. It has been shown that plant NAD-IDHs are exclusively localized in mitochondria (Chen & Gadal, 1990). Hence, it is concluded that the gene encodes a catalytic subunit of mitochondrial NAD-IDH. Recently, three cDNA clones encoding different NAD-IDH subunits have been isolated from tobacco plants (Lancien et al., 1998). This study confirms that the current functional prediction for potato cDNA is correct.

Discovering new orthologs: Homoisocitrate dehydrogenase

It has been suggested that new enzyme functions are established most easily and most commonly by recruitment of proteins already catalyzing analogous reactions (Jensen, 1976). NAD-IDHs, NADP-IDHs, and NAD-IMDHs are enzymes of very ancient origin. Protein engineering demonstrates that a few amino acid residue substitutions are sufficient to alter their specificities. Such a strategy may have been used to modify these enzymes into new orthologs that catalyze identical chemical reactions but have distinct substrate specificities. One example of this strategy is discussed below.

In fungi, lysine is synthesized via the α-aminoadipate pathway (Bhattacharjee, 1992). The conversion of homoisocitric acid to α-ketoadipic acid occurs due to the enzyme homoisocitrate dehydrogenase (EC 1. 1. 1. 87, NAD-HDH):

$$\text{Homoisocitric acid} + NAD^+ \overset{Mg^{2+}}{\rightleftharpoons} \alpha\text{-Ketoadipic acid} + CO_2 + NADH + H^+.$$

This reaction is clearly analogous to that of IDHs and IMDHs. Early studies showed that the enzyme of *Saccharomyces cerevisiae* was separated from the NAD-IDH and has a different pH optimum (Bhattacharjee, 1992). The molecular mass of the enzyme is 48 kD. These results suggest that NAD-HDH is a novel member of β-decarboxylating dehydrogenase family (Fig. 2). In spite of a

Table 2. *Functional reassignment of β-decarboxylating dehydrogenase genes*

Accession numbers	Sources	Database annotation	Corrected assignment
X92486	*S. tuberosum*	NAD-IDH	NADP-IDH
Z68343	*C. elegans*	IDH-like	NADP-IDH
AAC07444	*A. aeolicus*	NADP-IDH	NAD-IDH
Q59940	*Streptococcus* sp	NADP-IDH	NAD-IDH
Q59985	*S. salivarius*	NADP-IDH	NAD-IDH
CAA14727	*R. prowazekii*	IDH	NAD-IDH
P29696	*S. tuberosum*	NAD-IMDH	NAD-IDH *Cat*
Z79755	*C. elegans*	IDH-like	NAD-IDH *Cat*
Z46242	*C. elegans*	IDH-like	NAD-IDH *Reg*
Z81046	*C. elegans*	NAD-IDH-like	NAD-IDH *Reg*
Y16126	*L. esculentum*	NADP-IDH	NAD-IDH *Reg*
U81993	*A. thaliana*	NAD-IDH 1	NAD-IDH *Reg*
U81994	*A. thaliana*	NAD-IDH 2	NAD-IDH *Reg*
CAB16208	*S. pombe*	IDH	NAD-IDH *Reg*
CAA86325	*C. elegans*	IDH	NAD-IDH *Reg*
U92974	*L. lactis*	unknown	NAD-IMDH
H64389	*M. jannaschii*	NADP-IDH	NAD-IMDH
AAD03852	*S. aromaticivorans*	NAD-IMDH	NAD-TDH
CAB12208	*B. subtilis*	NAD-IMDH	NAD-TDH
Q48806	*L. pneumophila*	unknown	NADP-HDH
P40495	*S. cerevisiae*	NAD-IMDH	NAD-HDH
T38621	*S. pombe*	NAD-IMDH	NAD-HDH
Q58991	*M. jannaschii*	NADP-IDH	NAD-HDH
C64499	*M. jannaschii*	IDH	NAD-HDH
AAB84690	*M. thermoautotrophicum*	NADP-IDH	NAD-HDH
BAA30836	*P. horikoshii*	NAD-IMDH	NAD-HDH

number of biochemical and genetic studies, the gene encoding the enzyme has not been identified.

The substrate, homoisocitrate, is structurally similar to isocitrate and isopropylmalate where the γ-moiety is $-CH_2CH_2COO^-$ (Fig. 3). With the principles governing the relationship between structure and function in IDHs and IMDHs, it is predicted that all of the substrate binding and catalytic residues shared by *E. coli* NADP-IDH and *T. thermophilus* NAD-IMDH should be conserved in NAD-HDH. What makes this enzyme distinct would be the amino acid residues involved in binding the γ-moieties of substrate. Compared to isocitrate, the γ-carboxylate of homoisocitrate is one carbon further away from C2. This difference prevents the formation of a salt bridge between the bound homoisocitrate and the nicotinamide ring of the coenzyme in Michaelis complex that would otherwise pull the nicotinamide C4 out of the catalytic trajectory during hydride transfer. In this enzyme, the nicotinamide ring is probably aligned with the bound homoisocitrate by the protein itself, as seen in IMDH where the carboxyl group of Glu87 interacts with the nicotinamide ring (Fig. 4D). However, the presence of a negatively-charged residue such as Glu at this position would electrostatically repel the γ-moiety of bound substrate. Hence, the residue equivalent to Ser113 of the NADP-IDH and Glu87 of the NAD-IMDH should be Tyr or Gln, which is able to interact with the nicotinamide ribose ring via hydrogen bonding. Meanwhile, the residue equivalent to Asn115 of NADP-IDH and Leu90 of NAD-IMDH is likely a nonpolar residue that hydrophobically interacts with the extra-carbon portion of the γ-moiety of the bound homoisocitrate (Fig. 4A; Table 1).

Table 3. *Kinetic parameters of NAD-dependent isocitrate dehydrogenase (IDH) and NAD-dependent homoisocitrate dehydrogenase (HDH) from A. aeolicus and S. pombe, respectively*

Kinetic parameter	IDH at 65°C	HDH at 21°C
k_{cat} (s^{-1})	1.98	3.7
K_{mNAD} (μM)	27.1	182.0
$K_{misocitrate}$ (μM)	40.1	
$K_{mhomoisocitrate}$ (μM)		87.0

A search of the entire genomic sequence of *S. cerevisiae* leads to the identification of a candidate sequence with functional annotation as IMDH (accession number: P40495) (Table 1). This protein of 385 amino acid residues shares 32, 34, and 35% identity with the yeast NAD-IMDH, and the regulatory and catalytic subunits of NAD-IDH, respectively. The molecular mass is estimated as 41 kD. As expected, all the substrate binding and catalytic residues identified in both *E. coli* NADP-IDH and *T. thermophilus* NAD-IMDH are found in this protein. Significantly, the major substrate specificity determinants in the NADP-IDH, Ser113 and Asn115, are replaced by Tyr and Ile, which makes this protein different from IDH and IMDH. Meanwhile, the residues Lys344 and Tyr345, which interact with NADP in the *E. coli* enzyme, are replaced by Asp and Ile. This observation allows us to confidently assign the function of the protein as NAD-HDH. The NAD-HDH gene was also identified in *Schizosaccharomyces pombe* (CAB 11688). Due to the ease of subcloning, the coding region of the *S. pombe* genomic DNA was inserted in the expression vector pEMBL18$^-$ and the protein expressed in *E. coli*. Enzyme assay shows that the gene product is indeed NAD-HDH (Table 3). The enzyme is NAD-specific and shows no activity with isocitrate, isopropylmalate, or D-malate. Both IDH and IMDH exhibit detectable activity with homoisocitrate. It would be interesting to elucidate whether IDH or IMDH has been recruited for creation of NAD-HDH.

Deducing true function of "tartrate dehydrogenase"

NAD-dependent tartrate dehydrogenase (EC 1.1.1.93, NAD-TDH) is one of the numerous examples in which substrate-ambiguous enzymes are recruited for new function. This enzyme was identified from a strain of *Pseudomonas putida* capable of growth on (+)-(2R, 3R)-tartrate (Fig. 3) (Serfozo & Tipton, 1995). It catalyzes the initial step of the pathway for conversion of tartrate to glyceric acid, the NAD-dependent dehydrogenation of (+)-tartrate to 3R-oxaloglycolate. This enzyme does not catalyze the decarboxylation of 3R-oxaloglycolate; however, *meso*-(2R, 3S)-tartrate is also a substrate for NAD-TDH and undergoes enzyme-catalyzed decarboxylation (Fig. 3) (Serfozo & Tipton, 1995). Further studies show that the enzyme catalyzes the oxidative decarboxylation of (2R,3S)-isopropylmalate with significantly higher catalytic activity and the protein shares 30–40% sequence identity with bacterial NAD-IMDHs and 20–30% with NADP-IDHs (Tipton & Beecher, 1994). These results indicate that the enzyme is a new member of β-decarboxylating dehydrogenase family (Fig. 2). The catalytic activity, measured by k_{cat}/K_m, is 7×10^6 $s^{-1}M^{-1}$ for the *E. coli* NADP-IDH with (2R, 3S)-isocitrate, 2×10^6 $s^{-1}M^{-1}$ for the

S. typhimurium NAD-IMDH with (2R, 3S)-isopropylmalate and 4×10^2 $s^{-1}M^{-1}$ for the *P. putida* NAD-TDH with (+)-(2R, 3R)-tartrate, respectively (Tipton & Beecher, 1994; Chen et al., 1995, 1996b). The sluggish activity of NAD-TDH with (+)-(2R, 3R)-tartrate, for which the configuration is different from the natural substrates for other family members (Fig. 3), strongly suggests that this enzyme did not arise early in evolution as a (+)-(2R, 3R)-tartrate dehydrogenase, but rather, has been recruited recently in response to the appearance of (+)-(2R, 3R)-tartrate in the environment. This is supported by the discovery that NAD-TDH gene is also found in the *E. coli* genome, which is unable to grow on (+)-(2R, 3R)-tartrate (Table 1).

The hypothesis that TDH may have been optimized through evolution to catalyze a different function triggers us to search for its identity. Structure-based sequence alignment shows that all the substrate binding and catalytic residues identified in both *E. coli* NADP-IDH and *T. thermophilus* NAD-IMDH are found in this protein. On the other hand, both Asp and Ile interacting with NAD are conserved. However, the Glu87 in the NAD-IMDH is replaced with a hydrophobic residue Leu, indicating that NAD-TDH and NAD-IMDH are separate enzymes. This may explain why with isopropylmalate as substrate, NAD-IMDH catalyzes the reaction 100-fold faster than does NAD-TDH (R. Chen, unpubl. results). Another striking feature of the substrate binding pocket is the presence of a Trp residue that is adjacent to the Leu replacing the Glu87. Molecular simulation suggests that this bulky residue may lead to the enlargement of the binding pocket that a bulky and hydrophobic γ-moiety such as aromatic ring can be accommodated. Spectacularly, the NAD-TDH gene in the *E. coli* genome seems to form an operon with the genes that resemble genes for the degradation of the aromatic compounds (positions 1,878,868 to 1,884,868). The putative operon starts with the NAD-TDH gene *yeaU*, followed by transporter gene *yeaV*, the benzoate 1,2-dioxygenase α subunit gene *yeaW*, and the benzoate 1,2-dioxygenase β subunit gene *yeaX*. The NAD-TDH gene is preceded by an open reading frame (*yeaT*) resembling a number of transcriptional regulators. Hence, the operon is likely regulated by the product of *yeaT* gene. The map of this operon is strikingly analogous to that of the *E. coli lac* or *gus* operon, which both give the ability to transport and use exogenous lactose and glucuronides as carbon source, respectively. As such, "NAD-TDH" may catalyze a reaction of a novel pathway for degradation of the aromatic compounds imported via *yeaV* protein. So far, the "NAD-TDH" gene is found in *Agrobacterium tumefaciens*, *B. subtilis*, *E. coli*, *P. putida*, and *Sphingomonas aromaticivorans* (Table 2). These bacteria are either in association with plants or in the environment where they survive on the plant materials, suggesting that the xenobiotics for this newly proposed pathway may be of plant origin. It is known that plants accumulate a high concentration of tartrate ester formed with phenolic compounds, such as caffeic acid (Fig. 3) (Hohlfeld et al., 1996). Further research will be pursued to determine whether these plant metabolites are good substrates for NAD-TDH and can be transported and metabolized by bacterial cells.

Conclusions

NAD-IDH, NADP-IDH, and NAD-IMDH are members of β-decarboxylating dehydrogenase family that arose through gene duplication and divergence. Despite their highly divergent sequences, it is demonstrated that the rules for predicting the precise function of the family members can be established. The strategy is based on our recent findings that substitution of only a few amino acid residues in these enzymes are sufficient to change substrate and coenzyme specificities and thus exchange enzyme functions. Such structural and functional insights led us establish that the few critical specificity determinants can serve as reliable markers for determining orthologous or paralogous relationships. The power of this approach has been demonstrated by correcting similarity-based functional misassignments and discovering new genes and related pathways. The notion that only a few key amino acid residues are principally responsible for distinct binding and catalytic specificity of enzyme family or superfamily members is supported with other studies (Wilks et al., 1988; Babbitt et al., 1995). A striking example is lactate dehydrogenase and malate dehydrogenase, which share 20–40% sequence identity with a common Rossman fold. It has been demonstrated that substitution of one amino acid residue, Gln102 with Arg, in the lactate dehydrogenase from *Bacillus stearothermophilus*, converts the enzyme into a highly specific and efficient malate dehydrogenase (Wilks et al., 1988).

The current strategy integrates three-dimensional (3D) information and functional insights into sequence comparison. Such an approach for precise functional prediction offers substantial advantages over the similarity-based search (BLAST or clusters of orthologous group system), which are based on 1D information (Tatusov et al., 1997; Hofmann, 1998). It is also superior to the existing 3D information-based search (e.g., PROSITE, Pfam, or BLOCKS), because structural homology is insufficient for defining precise protein function. The prerequisite of this approach is a detailed knowledge of structure-function correlations for any particular family, which may require extensive studies. Nevertheless, with the rapid progress of genome-wide efforts to determine representative 3D structures for all protein families and an increasing understanding of enzyme binding and catalytic specificity evolution (Babbitt et al., 1995; Thornton, 1998), the current strategy should become broadly applicable. Already, we have found that the rules established for predicting coenzyme specificity of the β-decarboxylating dehydrogenases can be applied to other dehydrogenase families with NAD(P) as hydrogen acceptor (our results not shown). Hence, extension of similar studies for other protein families would be much needed to take full advantage of the enormous wealth of biological information coming out of the cDNA or genome sequencing projects. It is recommended that, for efficient and accurate functional assignment, one start with similarity-based searches and domain database searches for family or superfamily assignment and then take the current approach to identify characterized orthologs for precise functional prediction. Our strategy is particularly effective for identifying enzymes from poorly characterized genomes and may greatly accelerate the pace for discovering new enzymes and related pathways in these organisms.

Materials and methods

Sequence analysis

Search in the GenBank and SWISS-PROT databases for amino acid sequence similarities was performed using the program BLAST version 2.0 (Henikoff et al., 1997; Tatusov et al., 1997). Alignment of the NAD-IDH, NADP-IDH, and NAD-IMDH amino acid sequences was done with the CLUSTAL W program (Tatusov et al., 1997; Hofmann, 1998). Structure-based alignment was performed as follows: residues involved in binding and catalysis were iden-

tified from the X-ray crystal structures of *E. coli* NADP-IDH and *T. thermophilus* NAD-IMDH. Other sequences were then aligned using active residues as key landmarks. GenBank/EMBL accession numbers for NAD-IDHs are as follows: *Kluyveromyces lactis* AF045154, AF045153; *Macaca fascicularis* X74124, X82632, X87172; *M. fascicularis* X74124, X82632, X87172. Accession numbers for NADP-IDHs are as follows: *Archaeoglobus fulgidus* (AAB90591), AE001060; *E. coli*, J02799; *H*omo *sapiens* (AAC50455), U52144; *S. cerevisiae*, P41939 *T. thermophilus*, P33197. Accession numbers for IMDHs are as follows: *Bacillus subtilis*, P05645; *Brassica napus* X59970; *S. cerevisiae*, M12909; *T. thermophilus*, K01444.

Computer graphics

X-ray crystallographic structures were visualized on Silicon Graphics IRIS machines (Power Indigo2 Extreme Graphics, 4D/30TG) running TURBO-FRODO software. The coordinates of *T. thermophilus* IMDH and *E. coli* IDH and their binary complexes with coenzyme or substrate were retrieved and transferred from the Brookhaven National Laboratory Protein Data Bank. The NADP-IDH and NAD-IMDH structures were superimposed to align the amino acid sequences. Accession numbers for *E. coli* NADP-IDH are 3icd (NADP-IDH apoprotein), 5icd (binary complex with Mg^{2+}-isocitrate), 9icd (binary complex with NADP). Assession numbers for *T. thermophilus* NAD-IMDH are 1ipd (NAD-IMDH apoprotein) and 1hex (binary complex with NAD).

Gene cloning and kinetic analysis

The coding region of the genomic DNA for the *A. aeolicus* NAD-IDH and the *S. pombe* NAD-HDH was amplified, respectively, by polymerase chain reaction and linked to the downstream region of *E. coli* NADP-IDH promotor inserted into vector pRMBL18$^-$. The genes were entirely sequenced and no mutation was found.

The *A. aeolicus* NAD-IDH was expressed in an *idh*$^-$ strain of *E. coli*, SL4, to prevent contamination by the endogenous NADP-IDH (Chen et al., 1995). Sonicated extracts were incubated at 85 °C for 30 min and the denatured proteins removed by centrifugation. Supernatants were subjected to DEAE anion exchange chromatography. As judged by Coomassie blue staining after SDS polyacrylamide gel electrophoresis, enzyme preparation was at least 95% free from contaminating protein. Kinetics assays were conducted at 65 °C in KAC buffer (25 mM MOPS, pH 7.3, 5 mM $MgCl_2$, 100 mM NaCl, 1 mM dithiothreitol (DTT)) as described previously (Chen et al., 1995). Estimates of kinetic parameters were determined using unweighted nonlinear least-squares Newton–Raphson regressions to the Michaelis–Menten model.

The *S. pombe* NAD-HDH was expressed in *E. coli* strain SL4. The enzyme was purified to homogeneity by ammonium sulfate fractionation, hydroxyapatite chromatography, and DEAE anion-exchange chromatography. The enzyme preparation was 95% free from contaminating protein. Kinetics assays were conducted at 21 °C in KKAC buffer (25 mM MOPS, pH 7.3, 5 mM $MgCl_2$, 100 mM KCl, 1 mM DTT) as described previously (Chen et al., 1996a).

Acknowledgments

We thank six anonymous reviewers for valuable comments on the manuscript. This work has been supported by the research grants from the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Health Services and Utilization Research Commission to R.C.

References

Babbitt PC, Mrachko GT, Hasson MS, Huisman GW, Kolter R, Ringe D, Petsko GA, Kenyon GL, Gerlt JA. 1995. A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids. *Science 267*:1159–1161.

Bhattacharjee JK. 1992. Evolution of α-aminoadipate pathway for the synthesis of lysin in fungi. In: Mortlock ed. *The evolution of metabolic function*. London, UK: CRC Press. pp 47–80.

Bolduc JM, Dyer DH, Scott WG, Singer P, Sweet RM, Koshland DE Jr, Stoddard BL. 1995. Mutagenesis and Laue structures of enzyme intermediates: Isocitrate dehydrogenase. *Science 268*:1312–1318.

Chen R. 1999. A general strategy for enzyme engineering. *Trends Biotechnol 17*:344–345.

Chen R, Greer A, Dean AM. 1995. A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. *Proc Natl Acad Sci USA 92*:11666–11670.

Chen R, Greer A, Dean AM. 1996a. Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc Natl Acad Sci USA 93*:12171–12176.

Chen R, Greer AF, Dean AM. 1997a. Structural constraints in protein engineering—The coenzyme specificity of *Escherichia coli* isocitrate dehydrogenase. *Eur J Biochem 250*:578–582.

Chen R, Greer AF, Dean AM, Hurley JH. 1997b. *Engineering secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase*. San Diego, CA: Academic Press.

Chen R, Grobler JA, Hurley JH, Dean AM. 1996b. Second-site suppression of regulatory phosphorylation in *Escherichia coli* isocitrate dehydrogenase. *Protein Sci 5*:287–295.

Chen RD, Gadal P. 1990. Do the mitochondria provide the 2-oxoglutarate needed for glutamate synthesis in higher plant chloroplasts? *Plant Physiol Biochem 28*:411–427.

Cupp JR, McAlister-Henn L. 1993. Kinetic analysis of NAD(+)-isocitrate dehydrogenase with altered isocitrate binding sites: Contribution of IDH1 and IDH2 subunits to regulation and catalysis. *Biochemistry 32*:9323–9328.

Dean AM, Dvorak L. 1995. The role of glutamate 87 in the kinetic mechanism of *Thermus thermophilus* isopropylmalate dehydrogenase. *Protein Sci 4*: 2156–2167.

Dean AM, Shiau AK, Koshland DE Jr. 1996. Determinants of performance in the isocitrate dehydrogenase of *Escherichia coli*. *Protein Sci 5*:341–347.

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, et al. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature 392*:353–358.

Ellerstrom M, Josefsson LG, Rask L, Ronne H. 1992. Cloning of a cDNA for rape chloroplast 3-isopropylmalate dehydrogenase by genetic complementation in yeast. *Plant Mol Biol Int J Mol Biol Biochem Genet Eng 18*:557–566.

Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, Hood L. 1997. Gene families: The taxonomy of protein paralogs and chimeras. *Science 278*:609–614.

Hofmann K. 1998. Protein classification & functional assignment. In: Patterson M, Handel M, eds. *Trends guide to bioinformatics*. Amsterdam, The Netherlands: Elsevier Science. pp 18–21.

Hohlfeld M, Veit M, Strack D. 1996. Hydroxycinnamoyltransferases involved in the accumulation of caffeic acid esters in gametophytes and sporophytes of *Equisetum arvense*. *Plant Physiol 111*:1153–1159.

Hurley JH, Chen R, Dean AM. 1996. Determinants of cofactor specificity in isocitrate dehydrogenase: Structure of an engineered $NADP^+ \rightarrow NAD^+$ specificity-reversal mutant. *Biochemistry 35*:5670–5678.

Hurley JH, Dean AM, Koshland DE Jr, Stroud RM. 1991. Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: Implications from the structures of magnesium-isocitrate and $NADP^+$ complexes. *Biochemistry 30*:8671–8678.

Hurley JH, Thorsness PE, Ramalingam V, Helmers NH, Koshland DE Jr, Stroud RM. 1989. Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. *Proc Natl Acad Sci USA 86*:8635–8639.

Imada K, Sato M, Tanaka N, Katsube Y, Matsuura Y, Oshima T. 1991. Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 A resolution. *J Mol Biol 222*:725–738.

Jackson SD, Sonnewald U, Willmitzer L. 1993. Cloning and expression analysis

of beta-isopropylmalate dehydrogenase from potato. *Mol Gen Genet 238*: 309–314.

Jensen RA. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol 30*:409–425.

Keys DA, McAlister-Henn L. 1990. Subunit structure, expression, and function of NAD(H)-specific isocitrate dehydrogenase in *Saccharomyces cerevisiae*. *J Bacteriol 172*:4280–4287.

Lancien M, Gadal P, Hodges M. 1998. Molecular characterization of higher plant NAD-dependent isocitrate dehydrogenase: Evidence for a heteromeric structure by the complementation of yeast mutants. *Plant J 16*:325–333.

Nichols BJ, Perry AC, Hall L, Denton RM. 1995. Molecular cloning and deduced amino acid sequences of the alpha- and beta-subunits of mammalian NAD(+)-isocitrate dehydrogenase. *Biochem J 310*:917–922.

Ramachandran N, Colman RF. 1983. Chemical characterization of distinct subunits of pig heart DPN-specific isocitrate dehydrogenase. *J Biol Chem 255*:8859–8864.

Rossmann MG, Moras D, Olsen KW. 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature 250*:194–199.

Rutter GA, Denton RM. 1989. The binding of Ca^{2+} ions to pig heart NAD^+-isocitrate dehydrogenase and the 2-oxoglutarate dehydrogenase complex. *Biochem J 263*:453–462.

Serfozo P, Tipton PA. 1995. Substrate determinants of the course of tartrate dehydrogenase-catalyzed reactions. *Biochemistry 34*:7517–7524.

Stoddard BL, Dean A, Koshland DE Jr. 1993. Structure of isocitrate dehydrogenase with isocitrate, nicotinamide adenine dinucleotide phosphate, and calcium at 2.5-A resolution: A pseudo-Michaelis ternary complex. *Biochemistry 32*:9310–9316.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science 278*:631–617.

Thornton JM. 1998. The future of bioinformatics. In: Patterson M, Handel M, eds. *Trends guide to bioinformatics*. Amsterdam, The Netherlands: Elsevier Science. pp 30–31.

Tipton PA, Beecher BS. 1994. Tartrate dehydrogenase, a new member of the family of metal-dependent decarboxylating R-hydroxyacid dehydrogenases. *Arch Biochem Biophys 313*:15–21.

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.

Yaoi T, Miyazaki K, Oshima T. 1997. Substrate recognition of isocitrate dehydrogenase and 3-isopropylmalate dehydrogenase from *Thermus thermophilus* HB8. *J Biochem (Tokyo) 121*:77–81.