A modified consensus approach to mutagenesis inverts the cofactor specificity of Bacillus steatornerophilus lactate dehydrogenase

Humberto Flores¹ and Andrew D. Ellington²,³

¹Instituto de Biotecnología/UNAM, Apartado Postal 510-3, Cuernavaca, Morelos, 62271, México and ²Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, University of Texas at Austin, 1 University Station A4800, Austin, TX 78712, USA
³To whom correspondence should be addressed. E-mail: andy.ellington@mail.utexas.edu

Lactate dehydrogenase from Bacillus steatornerophilus is specific for NAD⁺. There have been several attempts to alter the cofactor specificity of this enzyme, but these have yielded enzymes with relatively low activities that still largely prefer NAD⁺. A modified consensus approach was used to create a library of phylogenetically preferred amino acids situated near the cofactor binding site, and variants were screened for their ability to utilize NMN⁺. A triple mutant (Mut31) was discovered that proved to be more catalytically efficient than wild-type. Mut31 was also better at utilizing NAD⁺ than the wild-type enzyme and was weakly active with NADP⁺ and NMN⁺. An analysis of single amino acid substitutions suggested that all three mutations worked in a concerted fashion to yield robust cofactor utilization. When two previously identified amino acid substitutions were introduced into the Mut31 background, the resultant quintuply substituted enzyme not only utilized NADP⁺ far better than the wild-type enzyme, it actually inverted its preference for NAD⁺ and NADP⁺.

Keywords: cofactor specificity/consensus/directed evolution/lactate dehydrogenase/molecular evolution

Introduction

It is generally true that the natural and directed evolution of proteins is constrained by the large sequence spaces that must be explored by proteins. A variety of mutagenesis, selection, and screening techniques have been developed to more efficiently create and search for functional amino acid substitutions (Neylon, 2004). In particular, random libraries can be constrained to particular positions in a protein that are thought to contribute to function. While the choice of which positions to randomize has frequently been based on structural inspection (Dalby, 2003), the huge amount of phylogenetic data available to researchers may, in many instances, be an even more useful parameter for guiding library construction.

The phylogenetic consensus sequence of a protein has previously been used to guide site-directed mutagenesis and to improve the stabilities of proteins. As an example, Pantoliano et al. (1989) compared the alignment of four subtilisins and chose to alter one residue in subtilisin BPN’ that differed from the consensus; this change greatly increased the stability of the protein. Similarly, multiple p53 homologues were aligned, and 20 positions were changed in the human p53 protein in order to generate a more stable variant (Nikolova et al., 1998). Combinations of single amino acid substitutions yielded p53 variants that were substantially more stable than the wild-type. However, not all 20 changes increased stability; in fact, some were destabilizing. The consensus approach was also used to improve the stability of GroEL minichaperones (short versions of the GroEL protein) (Wang et al., 1999). Based on an alignment of 130 members of the chaperonin 60 (Cpn60) family, 37 single amino acid variants of Escherichia coli GroEL were generated, and six stabilizing single substitutions were combined to yield a hextuple substitution that was much more stable than the wild-type. The stabilizing mutations mapped both on the surface and in the core of the protein (Wang et al., 2000). Again, though, not all single mutants were more stable than the parental protein. Lehmann et al. (2000a,c) constructed a consensus phytase based on a phylogenetic alignment of sequences from 13 fungal species. The consensus phytase showed increased stability over all known fungal phytases. When the effects of individual changes were subsequently examined, some were stabilizing while others were destabilizing. While the consensus method has most often been used to improve protein stability, it has also been proven useful in generating a thermostable phytase and in subsequently modifying its catalytic properties (specific activity, substrate specificity, and pH-activity profile) (Lehmann et al., 2000b).

A ‘consensus’ library for directed evolution experiments would allow more substitutions to be explored than the site-directed mutation experiments described above, but would not scatter mutations throughout a protein sequence, as PCR mutagenesis or DNA shuffling do. A consensus library would delimit the sets of amino acids that could be found at partially conserved positions in a protein. While we have already seen that individual, site-directed consensus substitutions do not necessarily lead to improved function, selection from or screening of a consensus library should yield only those changes that were functionally advantageous. Moreover, screening from or screening of a consensus library should reveal correlative or co-varying substitutions (Pazos et al., 1997) that change in parallel in order to alter protein function.

As a model system for evaluating the efficacy of a consensus library method, we chose to engineer the cofactor specificity of Bacillus steatornerophilus lactate dehydrogenase (BsLDH). The structure of this enzyme has been characterized at atomic resolution (Wigley et al., 1992), and its kinetic parameters are extremely well known (Burgner and Ray, 1978; Deng et al., 1994). More importantly, the functional properties of BsLDH have previously been altered by directed evolution, and a well-characterized assay system for the identification of mutant proteins was available (el Hawrani et al., 1996; Allen and Holbrook, 2000).

There have been several previous attempts to modify cofactor specificity in this enzyme, and these have achieved varying degrees of success, allowing us to benchmark the consensus library method. Feeney et al. (1990) introduced a single
site-directed change, D38S, into BsLDH, but succeeded only in increasing the $k_{cat}$ for NADH while concomitantly decreasing the ability of the enzyme to discriminate against NADPH. Holmberg et al. (1999) introduced a second substitution, I37K, into BsLDH, and achieved an overall improvement of 4-fold in specific activity ($k_{cat}/K_m$) for NADPH. However, the activities of the mutated BsLDH enzymes with NADPH were still extremely low. A single change in BsLDH has also been identified that allowed 3-acetylpyridine adenine dinucleotide to be used as a cofactor (Hewitt et al., 1997).

In order to identify BsLDH variants with altered cofactor binding sites, we screened a targeted consensus library. A number of residues in the Rossmann fold that were thought to interact with NADPH based on the high-resolution crystal structure (Wigley et al., 1992) were targeted, and the library was limited to the most common residues amongst a series of LDH homologues. After screening for activity, a triple mutant was identified that showed a 25-fold increase in catalytic efficiency with NAD$^+$ and also had better activity than the wild-type enzyme when NADP$^+$ was used as a cofactor. When these substitutions were combined with previously identified BsLDH mutations that altered cofactor specificity (Feeney et al., 1990; Holmberg et al., 1999), it proved possible to actually invert the preference for cofactors. The final, quintuple substitution variant utilized NADP$^+$ as well ($k_{cat}$ 23 $s^{-1}$) as the wild-type enzyme utilized NAD$^+$. Interestingly, the cofactor specificity of the variant was also influenced by an allosteric activator of the enzyme, $\alpha$-fructose 1,6-diphosphate (FBP).

### Materials and methods

#### Materials

The compounds $\beta$-NMN, $\beta$-NADP, $\beta$-NAD, $\alpha$-(+)-lactate, phenazine methosulfate (PMS), nitro blue tetrazolium (NBT), FBP, MES buffer, antibiotics, imidazole and Tris base were purchased from Sigma (St Louis, MO). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). DNasel was purchased from Epicentre technologies (Madison, WI). IPTG was purchased from Gold Biotechnology (St Louis, MO). Filter paper was purchased from Whatman (Maidstone, UK). Butterfly nitrocellulose membranes were purchased from Schleicher and Schuell (Keene, NH). B-PER was purchased from Pierce (Rockford, IL). Plasmid pET28a(+) was purchased from Novagen (Madison, WI). Plasmid pKKBsLDH was a generous gift of Professor Holbrook of the University of Bristol, UK. All oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA).

#### Structural and phylogenetic analyses

The three-dimensional structure of BsLDH-NADH-oxamate-FBP (PDB # 1LDN) (Wigley et al., 1992) was examined using the InsightII 98 program (Molecular Simulations, San Diego, CA) on a Silicon Graphics IRIS 6.3 work station (SGI, Mountain View, CA). The sequence of BsLDH (GenBank Accession No. M14788; Barstow et al., 1986) was submitted to the PredictProtein server (http://www.embl-heidelberg.de/predictprotein/predictprotein.html), and a multiple sequence alignment routine, MaxHom (Sander and Schneider, 1991), was used to determine which amino acids were phylogenetically conserved or variable between BsLDH and 67 other sequences: 6 malate dehydrogenases (mdh_halura, mdh_bacsu, mdh_bacs, mdh_chlte, mdh_chlau, and mdh_rhivl) and 62 $\alpha$-lactate dehydrogenases (ldh_bacst, ldh_bacca, ldhp_bacps, ldhx_bacps, ldhx_bacsu, ldhx_bacme, ldhx_lismo, ldhx_strmu, ldhx_lacca, ldhx_strpn, ldhx_strtr, ldhx_lacpe, ldhx_lacpl, ldhx_strbo, ldhx_lismo, ldhx_petma, ldhx_chick, ldhx_anapl, ldhx_xena, ldhx_xena, ldhx_mouse, ldhx_rat, ldhx_pig, ldhx_human, ldhx_chick, ldhh_sen, ldhh_funpa, ldhx_human, ldh_caen, ldh_vulv, ldh_bovin, ldh_myke, ldh_drome, ldh_funhe, ldh_funhe, ldh_human, ldh_mouse, ldh_rabid, mdh_halma, mdh_chick, mdh_squac, mdh_petma, mdh_chick, mdh_anapl, mdh_xena, mdh_xena, mdh_mouse, mdh_rat, mdh_pig, mdh_human, mdh_xena, mdh_metja, ldh_funhe, ldh_funa, mdh_bacis, ldhx_mouse, ldhx_rat, mdh_chlvi, dhl2_lacco, mdh_chlau, mdh_rhivl, ldh_toxgo, and ldh1_plafd). All of the dehydrogenases were specific for NAD. Sequence identities with BsLDH ranged from 30 to 90%, and sequence similarities ranged from 38 to 93%.

#### Construction of the 16/81/85/122/123 library

The BsLDH gene on the plasmid pKK223-3 (el Hawrani et al., 1996) served as a starting point for directed evolution. A library of variants was constructed by PCR using four mutagenic oligonucleotides (Bsldh16: 5'-AECTGGCGCGACGCGTG (ACG)GCCCGGCGCGATG; Bsldh81/85: 5'-TTTTCG(GTG)GCGCGGCGGCGG(CTG)ATATGCAACG; Bsldh122/ 123com: 5'-CGACCGGATTG(CG)TC(AG)CGACGAGAACACAG; and Bsldh122/123sen: 5'-TTTTCCTCGTGC(TG)CAATCACGGTCG; bases in parentheses represent an equimolar mixture) and two oligonucleotides (FWDBsldh: 5'-CGCCAGGCAAATTCTGTTTTATCAGACCGC and REVBsldh: 5'-TCGATAAGTGGATGATTGGAGCGGAT) that corresponded to plasmid sequences. Initially, REVBsldh and Bsldh16 were used to generate a fragment that was in turn used as a megaprimer (Sarkar and Sommer, 1990) for an amplification reaction in which Bsldh181/85 was the second primer. This product was in turn used as megaprimer for an amplification reaction in which Bsldh122/123com was the second primer. Ultimately, this set of reactions generated a gene fragment that encompassed about half of the BsLDH gene and that contained all of the mutagenized positions. We then used Bsldh181/85sen and FWDBsldh to create the other half of the gene. The two halves overlapped by 24 nucleotides, and these halves were used to assemble the full-length gene in an amplification reaction with FWDBsldh and REVBsldh. The final product was cleaved with EcoRI and PstI restriction endonucleases, purified, and ligated into pKK223-3, which had been cut with the same enzymes. The ligation mixture was transformed into Inv25 E.coli cells (Invitrogen, Carlsbad, CA). We had previously observed that basal expression of the BsLDH protein was sometimes harmful to cells, and so plasmid pREP4 (Qiagen, Valencia, CA), which contains the lacI$^\beta$ repressor and tightly represses expression from promoters that are under the control of the lac operator, was also present in these cells.

#### Screening for variants with cofactor specificity changes

The protocol described by el Hawrani et al. (1996) was used with some modifications to screen the consensus library for variants that had activity with NADP$^+$ or NMN$^+$. In this protocol, colonies that have LDH activity develop a blue to
purple halo. The differences from the published protocol were that we induced expression with 0.5 mM IPTG for 12 h, the cells were lysed with chloroform gas, and the substrate (L-lactate) concentration was 20 mM. During screening, the concentration of each cofactor was 5 mM. Screens were sometimes performed in the presence of 5 mM FBP, the natural allosteric activator of BsLDH.

Those colonies that appeared to have LDH activity greater than wild-type or a parental control were re-plated. Plasmids were purified, re-transformed into the same bacterial strain and re-screened for dehydrogenase activity with a given cofactor. Clones were grown in 3 ml of LB media for 3 h, and protein expression was induced for 3 h with 0.5 mM IPTG. Cells were lysed with 100 μl of B-PER (Pierce, Rockford, IL) containing 0.1 U Dnase I, 5 mM MgCl₂ and 0.5 M NaCl₂. Extracts (10 μl) were assayed on microplates in 100 μl of screening buffer (el Hawrani et al., 1996). Instead of looking for a blue halo on solid media, the formation of a blue precipitate corresponding to reduced NBT was followed (Scott et al., 1990; Allen and Holbrook, 2000). Genes corresponding to protein variants that were reproducibly more active than wild-type or parental controls were sequenced.

**Construction of I37K:D38S mutants**

The site-directed substitutions were constructed using the megaprimer method (Sarkar and Sommer, 1990). The primers Bsldh37-38 (5’-GGCGGTATTCGCTGAG CAG-CATGATCTCATC (nucleotides that were altered are in bold) and Bslldh66s (5’-CGCCTGAGCGGATATGAAGAA AACAAGGGAGCCCGA, which introduces a NdeI restriction site, underlined, into the start codon of the gene) were used to amplify a portion of the wild-type and Mut31 BsLDH genes. The resultant PCR product was then used as a megaprimer along with forward and reverse primers for amplification reactions with two different templates, the wild-type and Mut31 genes. The full-length, mutated genes were cut with NdeI and HindIII and cloned into pET28a(+). Ligation reactions were transformed into DH5αlac(DE3) pLysS (Matsumura and Ellington, 2001). The sequences of the recombinant genes were confirmed. Ultimately, proteins produced from these constructs had His tags fused to their N-termini.

**Protein purification**

The Mut31 gene in pKK223-2 was amplified using Bslldh66s and FWDBsldh. The product was cleaved with NdeI and HindIII restriction endonucleases and cloned into pET28a(+). Cloned genes were sequenced. Variants WT I37K:D38S and Mut31 I37K:D38S were initially constructed in pET28a(+), as described above, and no subsequent manipulations were necessary.

The expression constructs were transformed into *E. coli* strain DH5αlac (DE3) cells carrying the plasmid pLysS. Transformants were grown at 37°C in 350 ml of LB containing kanamycin and chloramphenicol to mid-log phase (A600 ~ 0.6). IPTG was added to induce protein expression, and the cells were allowed to grow for 3 h. The induced cells were collected by centrifugation and lysed by adding 5 ml of B-PER (Pierce, Rockford, IL) containing 1 U Dnase I, 5 mM MgCl₂ and 0.5 M NaCl₂. The lysates were centrifuged at 10 000 r.p.m. for 30 min and passed over a nickel chelate affinity column. The columns were washed with 10 column volumes of wash buffer (50 mM phosphate, pH 7; 0.5 M NaCl₂; 60 mM imidazole) and with 2 ml of 125 mM imidazole (100 mM MES, pH 6.5). Bound proteins were eluted with 2 ml of 0.2 M imidazole (100 mM MES, pH 6.5). The concentration of eluted proteins was determined using a Bradford protein assay (Bio-Rad, Hercules, CA). The homogeneity of protein samples, as determined by SDS–PAGE, was at least 95%.

**Results and discussion**

**Construction of a consensus library**

The structure of BsLDH was examined, and residues within 4 Å of NAD⁺ were chosen as potential targets for mutagenesis (Figure 1). Seventeen positions were identified as being adjacent to the cofactor. Complete randomization of all of these positions would have required that we screen on the order of 20¹⁷ (1 × 10²²) variants, not including potential skewing due to codon distribution. Since the plate assay that we and others have employed for screening BsLDH for functional variants is limited to roughly 10000 colonies per round, it was necessary to devise a method that would further delimit the number of variants that would be screened. An examination...
of the phylogenetic alignment of the LDH family indicated that only five of the 17 structurally identified residues actually underwent any significant change during evolution (Table I).

The five identified positions (Phe16, Cys81, Asn85, Ala122 and Thr123) were partially randomized such that a given position included both the wild-type amino acid and other, phylogenetically prevalent amino acids. For example, position 16 in BsLDH is phenylalanine, which is present 19% of the time in the LDH family. However, alanine (30%) and glutamine (22%) were even more prevalent. Therefore, in the consensus library the sequence of codon 16 was chosen to be (TGC)/(TCA)/(CG) (first, second and third codon positions, respectively) in order to include these two more predominant amino acids (as well as phenylalanine and seven other amino acids). Similarly, position 81 in BsLDH is cysteine, yet this amino acid is present in only 9% of family members, while the vast majority (84%) contain threonine (Table I). Again, the sequence of codon 81 was (TA)/(GC)/C, which allowed for the incorporation of either cysteine or threonine, as well as serine. To our knowledge, this is one of the first times a consensus library has been generated, as opposed to merely constructing individual consensus variants through site-directed mutagenesis. The number of different variants in the library was only 1728 (18×4×6×2×2), a number which could be readily screened by colony-lift assays, as described below.

<table>
<thead>
<tr>
<th>F 16/30</th>
<th>T 81/95</th>
<th>A 122/136</th>
<th>T 123/137</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (30)b</td>
<td>M (7.5)</td>
<td>R (3)</td>
<td>S (16)</td>
</tr>
<tr>
<td>Q (22)</td>
<td>T (1.5)</td>
<td>Q (1.5)</td>
<td>N (1.5)</td>
</tr>
</tbody>
</table>

Note: The first column shows the wild-type amino acid. The first number is the amino acid position in the gene as in Barstow et al. (1986), while the second number is the assignment in the three-dimensional structure as in Wigley et al. (1992). Values in parentheses are the percent of the total.

Screening for variants with altered cofactor specificities

In order to identify variants that were especially good at utilizing the nicotinamide moiety, the resultant libraries were screened using a colony-lift assay for BsLDH activity with either NADP+ (a related cofactor) or NMN+ (which could potentially be either a cofactor or a competitive inhibitor of endogenous cofactors). An initial screen of 10,000 colonies with NADP+ did not yield variants that had reproducibly better activity than the wild-type enzyme. However, a second screen of 20,000 colonies with NMN+ as the cofactor yielded three active variants, one of which had reproducibly better activity than the wild-type enzyme upon re-screening.

The new variant, Mut31, had three amino acid substitutions: Phe16Stop (amber), Cys81Ser and Asn85Arg (Table II). Since the E.coli strain in which the screens were carried out contains an amber suppressor, the amino acid substitution at position 16 is actually Phe16Gln. These substitutions clearly clustered near the ester bond between phosphate and the ribose of nicotinamide (Figure 1). Of the amino acids that were adopted, two (16Gln and 85Arg) were prevalent amongst LDHs, with 85Arg actually predominating (Table I). However, serine at position 81 is only rarely found, with the prevalent residue being threonine. Even though serine was over-represented in the oligonucleotide used to mutagenize this position, enough variants were screened so that threonine was functionally optimal, it could have been selected. The number of clones that were screened theoretically represents all possible variants in the library. In support of this hypothesis, we also sequenced a number of unselected variants, and while these variants contained numerous amino acid substitutions corresponding to the mutated codon positions, they all had lower activities than the wild-type enzyme. Since Mut31 contains both prevalent and more rare substitutions, it can be asserted that the optimal set of substitutions would not have been found by a purely consensus approach, which would have been unlikely to explore such rare combinations. Our results thus highlight the advantages of using a consensus library.

A number of additional screens of both the original library and a Mut31 library generated by error-prone PCR did not reveal any further improved variants (~100,000 colonies were screened). These screens were carried out with either NADP+, NMN+, or NMN+ plus adenosine [as this combination had previously been shown to increase the activity of dehydrogenases with NMN+ (Sicsic et al., 1986)]. The fact that only a single improved variant was identified was not surprising. Our criterion that the enzyme should have better function than the wild-type limited the number of variants that were likely to be identified and concomitantly excluded the selection of the wild-type itself. In addition, previous attempts to engineer BsLDH variants with altered cofactor utilization yielded
enzymes with low activities and at best slight decreases in the $K_m$ for NADPH.

**Kinetic parameters of Mut31**

The kinetic parameters of Mut31 and of the wild-type enzyme were determined with NAD$^+$ as a cofactor (Table III). Kinetic analyses were carried out in the ‘backwards’ direction with lactate and NAD$^+$ in order to discern how the kinetic properties of the enzyme evolved under the conditions actually used in the screen. Moreover, since BsLDH is an allosteric enzyme and is significantly activated by the effector FBP (Cameron et al., 1994; Allen and Holbrook, 2000), kinetic values were derived both in the presence and absence of saturating FBP.

In the absence of FBP, Mut31 has a $k_{\text{cat}}$ value that was over 3-fold higher than that of the wild-type enzyme, and a $K_m$ value for NAD$^+$ that was decreased by 4-fold. Overall, the $k_{\text{cat}}/K_m$ value of Mut31 LDH with NAD$^+$ as a cofactor increased by almost 15-fold.

In the presence of FBP, the $k_{\text{cat}}$ value of Mut31 was 70% that of the wild-type enzyme (23 and 32 s$^{-1}$, respectively). However, the $K_m$ value for NAD$^+$ fell substantially in Mut31, to 8 $\mu$M, a value 13-fold less than the wild-type. The mutant enzyme showed a smaller activation by FBP than the wild-type (15-fold activation of $k_{\text{cat}}$). The mutant amino acid substitutions found in Mut31 may inherently shift the conformation of the enzyme towards the activated structure normally produced by FBP binding. Similarly, previous directed evolution experiments identified a triply substituted variant of BsLDH that was fully active even in the absence of FBP (Allen and Holbrook, 2000), but none of these amino acid substitutions overlapped with the three amino acid substitutions found in the current study. Taken together, the six amino acid substitutions found between these two studies may help to define the allosteric pathway for this enzyme.

Mut31 was also assayed with NADP$^+$ and NMN$^-$. The variant showed very low activity with NADP$^+$ plus FBP (roughly 4% the activity seen with NAD$^+$ plus FBP) and it proved difficult to extract individual kinetic parameters. However, based on initial rate data under standard assay conditions, Mut31 had 3- to 4-fold more activity with NADP$^+$ than did the wild-type enzyme (Table IV and Figure 2). This improvement cannot be attributed to generally enhanced catalytic efficiency in the mutant enzyme, since in the presence of

### Table III. Kinetic parameters of BsLDH variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>$k_{\text{cat}}$ (s$^{-1}$)</th>
<th>$K_m$ (mM)$^a$</th>
<th>$k_{\text{cat}}/K_m$</th>
<th>Cofactor FBP (activator)</th>
<th>Activation$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>32 ± 0.80</td>
<td>105 ± 8.8</td>
<td>0.30</td>
<td>NAD$^+$</td>
<td>+</td>
</tr>
<tr>
<td>Mut31</td>
<td>23 ± 0.26</td>
<td>8.1 ± 0.52</td>
<td>2.8</td>
<td>NAD$^+$</td>
<td>+</td>
</tr>
<tr>
<td>WT</td>
<td>3.6 ± 0.04</td>
<td>167 ± 8</td>
<td>0.02</td>
<td>NAD$^+$</td>
<td>–</td>
</tr>
<tr>
<td>Mut31</td>
<td>13 ± 0.22</td>
<td>42 ± 2.8</td>
<td>0.31</td>
<td>NAD$^+$</td>
<td>–</td>
</tr>
<tr>
<td>Q16</td>
<td>38 ± 0.28</td>
<td>65 ± 2.6</td>
<td>0.59</td>
<td>NAD$^+$</td>
<td>+</td>
</tr>
<tr>
<td>S81</td>
<td>39 ± 0.35</td>
<td>88 ± 3.8</td>
<td>0.43</td>
<td>NAD$^+$</td>
<td>+</td>
</tr>
<tr>
<td>Q16</td>
<td>4.2 ± 0.04</td>
<td>142 ± 6</td>
<td>0.03</td>
<td>NAD$^+$</td>
<td>–</td>
</tr>
<tr>
<td>S81</td>
<td>3.9 ± 0.08</td>
<td>248 ± 20</td>
<td>0.015</td>
<td>NAD$^+$</td>
<td>–</td>
</tr>
<tr>
<td>Mut31 I37K:D38S</td>
<td>23 ± 1.6</td>
<td>4700 ± 630</td>
<td>0.005</td>
<td>NADP$^+$</td>
<td>+</td>
</tr>
</tbody>
</table>

Three independent measurements were performed. The standard error is shown. $^aK_m$ for cofactor. $^bActivation$ parameters were calculated using the following equation:

$$\frac{(k_{\text{cat}}/K_m)_{\text{FBP}}}{(k_{\text{cat}}/K_m)_{\text{FBP}}}.$$
FBP the specific activities of the wild-type and Mut31 enzymes were much the same. Rather, it seems likely that the amino acid substitutions have begun to alter the cofactor specificity of the enzyme.

When NMN⁺ was examined as a cofactor, the measured activities were even lower than those that were observed with NADP⁺, and thus kinetic parameters were again difficult to derive. While no activity greater than background was observed when NMN⁺ and adenosine were incubated with the wild-type enzyme and lactate, even at long time points, a trace of activity was exhibited by Mut31 (Figure 3). This is not the first time that a natural dehydrogenase has been reported to use NMN⁺ as a cofactor (Fisher and McGregor, 1969; Plapp et al., 1986; Sicic et al., 1986), but it is the first time that directed evolution has yielded a dehydrogenase that is active with NMN⁺. This faint activity with NMN⁺, or a combination of activity with endogenous cofactors and NMN⁺, may have been one reason that Mut31 was originally picked in our colony screen; we have previously observed that formazan dye deposition is frequently a much more sensitive assay for activity than the gain or loss of NADH fluorescence (see also Scott et al., 1990).

**Concerted mutational and phenotypic change**

In order to determine which of the three amino acid substitutions contributed to the improvements in kinetic parameters, and whether the substitutions were additive or synergistic, we attempted to express each of the substitutions individually (Table II). Only variants F16Q and C81S were successfully expressed; N85R was produced at such low levels that it was not possible to isolate and characterize the protein. The kinetic parameters of the two expressed, single substitution variants were for the most part very similar to those of the wild-type enzyme (Table III). The biggest kinetic difference was a slight (less than 2-fold) decrease in $K_m$ for NAD⁺ in the presence of FBP, consistent with the much larger decrease in $K_m$ that was seen in the triple mutant. Overall, it appears as though the observed changes in kinetic parameters in the triple mutant were either largely due to N85R or were a synergistic property of the concerted amino acid substitutions.

Other evidence supports the identification of N85R as a functionally important residue. In other dehydrogenases the equivalent position has been shown to make a hydrogen bond with the pyrophosphate of the coenzyme (Smiley et al., 1971). Therefore, changes in this position could influence cofactor affinity and specificity. In fact, it was already known that an adjacent change, Q86R, altered substrate specificity (Wilks et al., 1988; el Hawrani et al., 1996). Nonetheless, the fact that the N85R variant on its own could not be functionally expressed in *E.coli* implies that F16Q and/ or C81S abrogated the instability conferred by N85R. Therefore, all three substitutions were likely necessary to generate a stable protein that also had altered cofactor specificity.

**Mut31 as a basis for further engineering**

All of the enzymes that were initially aligned in order to generate the sequence of the consensus library were specific for NAD⁺. Thus, it is not surprising that only modest changes in cofactor specificity were obtained. However, by generally improving the ability of the enzyme to interact with its cofactor, we may have generated a platform that will be more tolerant of amino acid substitutions that mediate cofactor specificity.

Previous work had identified two amino acid substitutions, I37K and D38S, that could improve the ability of BsLDH to utilize NADPH (Feeney et al., 1990; Holmberg et al., 1999). These positions had originally been chosen for rational mutagenesis based on an analysis of the crystal structure of the enzyme, as they were close to the 2′ phosphate of NADPH. An analysis of mutations at these positions first identified D38S (Feeney et al., 1990) and subsequently I37K (Holmberg et al., 1999). However, these two substitutions improved the $k_{cat}$ of the wild-type enzyme for NADPH by only 4-fold; in contrast, the $K_m$ for NADPH increased and overall $k_{cat}/K_m$ was only marginally improved.

Since it looked as though Mut31 had generally improved kinetic properties, we decided to see how the introduction of these two additional mutants would influence its ability to utilize NADP⁺. Site-directed mutagenesis was used to construct both wild-type and Mut31 genes which carried I37K and D38S substitutions (Table II and Figure 4). The catalytic parameters of the expressed, purified Mut31 protein containing these additional substitutions (Mut31 I37K:D38S) are shown in Table III. The $k_{cat}$ of Mut31 I37K:D38S for NADP⁺ was similar to that of the Mut31 parent for NAD⁺, while the $K_m$ for NADP⁺ was over 4 mM, much higher than that of the other BsLDH variants for NAD⁺.

The low activities of the quintuply substituted variant with NAD and of other BsLDH variants with NADP⁺ made it difficult to extract kinetic parameters for these proteins, but the initial rates of all proteins were compared under standard assay conditions (Table IV and Figure 2). The improvements exhibited by the pentuple substitution relative to its parents were substantial. In the presence of FBP, the Mut31 I37K:D38S variant was 10-fold faster than Mut31 with NADP⁺, and 36-fold faster than the wild-type protein containing the I37K and D38S double substitution. In contrast, in the presence of FBP the rate of the Mut31 I37K:D38S variant with NAD⁺ was 5-fold reduced relative to the Mut31 parent, which had previously shown increased activity with both nicotinamide cofactors. As Holmberg and co-workers (1999) have
Inverting cofactor specificity of B. stearothermophilus LDH

Previously shown, the I37K:D38S double substitution inhibits the utilization of NAD$^+$ and results in a reversal in cofactor specificity.

Interestingly, the activator (FBP) appears to alter the cofactor specificity of the BsLDH pentapeptide substitution variant (Table IV and Figure 2). In its presence, Mut31 I37K:D38S has a higher initial velocity with NADP$^+$; in its absence, the enzyme has a higher initial velocity with NAD$. The ability of FBP to modulate cofactor specificity is not present in variants that lack the I37K:D38S substitutions (Table IV and Figure 2). The role of FBP as a ‘specificity switch’ would represent a new function for this allosteric effector, which is typically thought of solely as a molecule that increases the affinity of the enzyme for its substrate (Cameron et al., 1994; Allen and Holbrook, 2000).

Comparing the consensus library method with consensus mutational methods

Our results have implications for two aspects of protein engineering: the number of mutations required to achieve phenotypic change, and how to best target those mutations. Pazos et al. (1997) have previously suggested that many amino acid substitutions that arise during natural protein evolution are paired, covarying amino acid substitutions. Thus, multiple amino acid substitutions may need to simultaneously occur during directed evolution in order to alter enzyme activities. This hypothesis has in fact been borne out by a number of studies in which high mutation rates were required to yield novel protein phenotypes (Cramer et al., 1998; Zhao et al., 1998; Zaccolo and Gherardi, 1999; Daugherty et al., 2000). For example, Daugherty et al. (2000) compared scFv libraries with different mutation rates, and found that the most highly mutated library had the most active clones. In line with these results, all three mutations in Mut31 had to appear simultaneously in order to begin to alter cofactor specificity. The further addition of the I37K and D38S substitutions might be introduced serially, but again it was only when all five substitutions were present together and could act in concert that the cofactor specificity of the protein was radically altered.

We initially chose sets of sequence substitutions in BsLDH based on the phylogenetic consensus. The consensus approach assumes that Nature has efficiently searched protein sequence space, and this approach has previously been used to improve the stabilities and kinetic parameters of other proteins. For example, Pantoliano et al. (1989) aligned four subtilisins, and altered one position on subtilisin BPN’ to the consensus amino acid; the change increased the stability of the protein. Similarly, Nikolova et al. (1998) aligned 23 proteins that were homologous to p53, selected 20 positions to change, and found combinations of these substitutions that were more stable than the wild-type p53. Wang et al. (1999) aligned 130 members of the chaperonin 60 (Cpn60) family, then combined six consensus substitutions to create a more stable GroEL minichaperone. Lehmann et al. (2000a,c) went beyond mutating a single protein and actually constructed a consensus phytase based on the alignment of 13 diverse, natural sequences. The artificial, consensus phytase was more stable than all of the natural phytases, and it was used as a template to generate additional phytase variants with novel catalytic properties (specific activity, substrate specificity and pH-activity profile). Finally, Steipe et al. (1994) showed that a consensus method could predict stabilizing mutations in immunoglobulin variable domains with a success rate of 60%.

However, while the consensus method is frequently a useful guide to mutagenesis and protein engineering, it should be noted that it has not been universally successful. For example, not all of the predicted substitutions in consensus-based experiments yielded predicted or functional results (Nikolova et al., 1998; Wang et al., 1999; Lehmann et al., 2000a). While some of the predicted substitutions chosen by Nikolova et al. (1998) increased the stability of p53, others were destabilizing. Likewise, not all of the consensus substitutions chosen by Wang et al. (1999) stabilized GroEL minichaperones. Similarly, some of the amino acid substitutions that made up Lehmann’s consensus phytase (Lehmann et al., 2000a) were destabilizing when introduced in isolation.

In the consensus methods cited above, generally only the phylogenetically most common amino acid at a given position was used. In contrast, in the consensus library method, any of several phylogenetically favored amino acids can be introduced. For example, Mut31 included two amino acid substitutions (F16Q and C81S) that did not conform to the phylogenetically most common residues, but that were nonetheless found in other protein family members. Ultimately, our consensus library selection was able to utilize natural diversity to alter the activity of BsLDH better than previous attempts using other methods.

Comparing the consensus library method with family shuffling methods

Family shuffling has been another popular approach to engineering protein activities (reviewed in Schmidt-Dannert, 2001). Family shuffling techniques have been used to direct the engineering of a wide array of proteins including cephalosporinases with greater moxalactamase activities (Cramer et al., 1998), interferon-2 variants with greater antiviral and antiproliferation activities (Chang et al., 1999), subtilisins with higher enzymatic activities and unique properties (Ness et al., 1999).
and novel carotenoid synthetic pathways (Schmidt-Dannert et al., 2000).

However, an inherent problem with family shuffling is the inability of variants to diversify significantly from parental sequences. In those instances where family members have low homology with one another, it is difficult to obtain shuffled products. This limitation has been addressed in part by newer methods, such as degenerate homoduplex recombination, which makes use of synthetic oligonucleotides as bridge sequences (Coco et al., 2002). In addition, though, deleterious mutations can accumulate in parallel with functional mutations during shuffling, leading to a large mutational load that can potentially limit the number of rounds of selection and the diversities of selected variants. Even in those instances where shuffling or other mutagenic methods can generate multiple mutations in parallel, current selection and screening methods often cannot adequately evaluate all of the random variants.

To obviate these problems, family shuffling techniques have been further modified. Newer computing methods which identify areas of potentially beneficial mutation or minimal disruption may prove useful in reducing variant size, but these are still limited in the number of residues that can be examined (Voigt et al., 2001, 2002). To take advantage of such algorithms and of the functional information that can generally be gleaned from phylogenetic analyses, ‘synthetic shuffling’ methods in which focused libraries are generated based on phylogenetic data have been developed. Govindarajan et al. (2003) looked at 15 synthetically recombinated subtilisin genes, and found that almost all possible pairwise combinations of naturally varying amino acids can exist. Ness et al. (2002) capitalized on the apparent lack of negative interactions and incorporated all of the phylogenetic variation that was found across the 15 subtilisin genes into one library. By screening the library for variants with extreme properties, it proved possible to identify enzymes that were both thermostable and active at pH 10. However, interestingly, a lower proportion of the original library was active (20%) than was the case for conventional family shuffling methods (75%). While such synthetic shuffling experiments are very powerful, in general they may again be limited by the number of variants that can be screened, especially when there is considerable variation between family members. In contrast, the consensus library method we have advanced selects from amongst a phylogenetically validated subset of already likely answers.

Conclusions
We believe that the method we have chosen for mutagenesis may be optimal for engineering radical changes in protein phenotype. By simultaneously mutagenizing multiple amino acids we can readily explore concerted amino acid substitutions, yet the number of possible variants is well within the range of most screening techniques. The inclusion of amino acid substitutions that mimic the phylogenetic consensus guarantees that many of the enzyme variants will be functional. However, because codons are randomized, the choices of amino acids at critically chosen positions are not limited to the phylogenetic consensus, allowing fine-tuning of cumulative correlations (as was the case with the non-consensus Cys81Ser substitution).

Acknowledgements
This research was supported by a grant from the Office of Naval Research and by a MURI award from the Army Research Office. H.F. thanks the Instituto de Biotecnologia/UNAM for assistance with his postdoctoral fellowship. We thank Amos Yan for the critical reading and helpful discussions on the manuscript.

References

Received October 28, 2004; revised May 12, 2005; accepted June 1, 2005

Edited by Andrew Griffiths