

Rapid Evolution of β -Glucuronidase Specificity by Saturation Mutagenesis of an Active Site Loop*

Received for publication, February 9, 2004, and in revised form, April 6, 2004
Published, JBC Papers in Press, April 6, 2004, DOI 10.1074/jbc.M401447200

Melissa L. Geddie and Ichiro Matsumura‡

From the Department of Biochemistry, Center for Fundamental and Molecular Evolution, Rollins Research Center, Emory University School of Medicine, Atlanta, Georgia 30322

Protein engineers have widely adopted directed evolution as a design algorithm, but practitioners have not come to a consensus about the best method to evolve protein molecular recognition. We previously used DNA shuffling to direct the evolution of *Escherichia coli* β -glucuronidase (GUS) variants with increased β -galactosidase activity. Epistatic (synergistic) mutations in amino acids 557, 566, and 568, which are part of an active site loop, were identified in that experiment (Matsumura, I., and Ellington, A. D. (2001) *J. Mol. Biol.* 305, 331–339). Here we show that site saturation mutagenesis of these residues, overexpression of the resulting library in *E. coli*, and high throughput screening led to the rapid evolution of clones exhibiting increased activity in reactions with *p*-nitrophenyl- β -D-xylopyranoside (pNP-xy). The xylosidase activities of the 14 fittest clones were 30-fold higher on average than that of the wild-type GUS. The 14 corresponding plasmids were pooled, amplified by long PCR, self-ligated with T4 DNA ligase, and transformed into *E. coli*. Thirteen clones exhibiting an average of 80-fold improvement in xylosidase activity were isolated in a second round of screening. One of the evolved proteins exhibited a ~200-fold improvement over the wild type in reactivity (k_{cat}/K_m) with pNP-xy, with a 290,000-fold inversion of specificity. Sequence analysis of the 13 round 2 isolates suggested that all were products of intermolecular recombination events that occurred during whole plasmid PCR. Further rounds of evolution using DNA shuffling and staggered extension process (StEP) resulted in modest improvement. These results underscore the importance of epistatic interactions and demonstrate that they can be optimized through variations of the facile whole plasmid PCR technique.

Adaptive evolution is arguably the most fundamental biological process, although it remains poorly understood. A better understanding of molecular adaptation would facilitate the engineering of enzymes with novel specificities. Theoreticians have proposed that a relatively small number of catalytically inefficient, broad specificity proto-enzymes diversified through gene duplication and adaptive evolution into the multitude of efficient and specialized catalysts present in the contemporary

biosphere (2–6). This supposition cannot be proven nor refuted through experimentation, but recapitulation of the diversification process in the laboratory would support its feasibility and reveal possible underlying structural mechanisms.

The glycoside hydrolases exemplify enzyme diversification. Natural selection has matched the great structural diversity of carbohydrates with a multitude of enzymes that selectively catalyze their cleavage. These enzymes have been classified into 91 families based on amino acid sequence similarity (7). The GH-A clan (sometimes called the 4/7 superfamily) comprises >3000 GenBank™ sequences from Families 1, 2, 5, 10, 17, 26, 35, 39, 42, 53, 59, 72, 79, and 86 (8). All GH-A enzymes retain the same (β/α)₈-barrel fold and catalytic mechanism (9) and are thought to have diverged from a common ancestor (10). Consider the specific example of two well characterized glycoside hydrolases: human β -glucuronidase (GUS,¹ Family 2) and *Thermoanaerobacterium saccharolyticum* β -xylosidase (XYL, Family 39). These distant homologs catalyze the hydrolysis of similar substrates that differ only in their C5 substituents (carboxylate for β -glucuronides, hydrogen for β -xylosides), but overlap very little in specificity. Their amino acid sequences are too divergent to align, but six conserved active site residues are superimposable (Fig. 1) (11, 12). Our goal is to understand how enzymes like GUS and XYL evolved to become so specific.

X-ray crystallography is essential to understanding the structural basis of enzyme specificity, but the complexity of protein structures necessitates the development of complementary functional approaches. We randomly mutate genes or parts of genes, and express the resulting libraries in populations of microorganisms. High throughput screening enables the functional evaluation of thousands of sequence variants in parallel. Iterative cycles of mutagenesis, point mutant recombination, and high throughput screening lead to the accumulation of beneficial mutations in selected clones (13). Directed evolution has been used to alter properties of enzymes including substrate specificity (reviewed in Refs. 14–18).

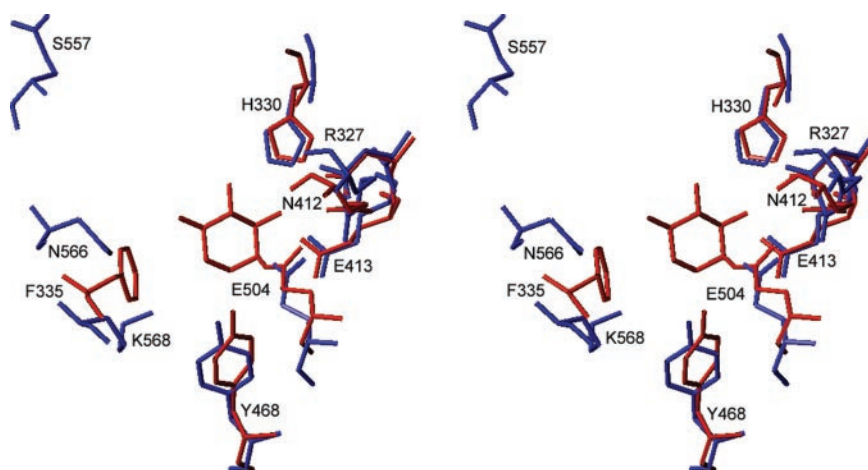
The *Escherichia coli* GUS is a good model system for the study of adaptive evolution. As noted above, the human GUS protein has been crystallized; the amino acid sequences of the *E. coli* and human GUS homologs are 50% identical, with highly conserved active sites (12). The *E. coli* β -glucuronidase gene (*gusA*, formerly *uidA*) is more amenable to directed evolution, because it can be overexpressed at high levels in *E. coli*, enabling the detection of weak catalytic activities in reactions with β -xylosides and other non-native substrates. We have previously directed the evolution of GUS variants with in-

* This work was supported by Grant BIO/MCB 0109668 from the National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Department of Biochemistry, Center for Fundamental and Molecular Evolution, Rollins Research Center, Emory University School of Medicine, 1510 Clifton Rd., Rm. 4119, Atlanta, GA 30322. Tel.: 404-727-5625; Fax: 404-727-3452; E-mail: imatsum@emory.edu.

¹ The abbreviations used are: GUS, β -glucuronidase; XYL, β -xylosidase; PSQ, β -glucuronidase mutants containing Pro⁵⁵⁷/Ser⁵⁶⁶/Gln⁵⁶⁸; PAF, β -glucuronidase mutants containing Pro⁵⁵⁷/Ala⁵⁶⁶/Phe⁵⁶⁸; pNP-gluc, *p*-nitrophenyl- β -D-glucuronide; pNP-xy, *p*-nitrophenyl- β -D-xylopyranoside; StEP, staggered extension process; LB, Luria broth.

FIG. 1. Conserved active site residues of *T. saccharolyticum* β -xylosidase (11) and human β -glucuronidase (12). The residues are numbered according to the *E. coli* β -glucuronidase. β -Xylosidase residues are in red, and conserved β -glucuronidase residues are in blue. The hexagonal structure in the center is a covalently bound intermediate of the xylosidase reaction.



creased β -galactosidase activity. We expressed a library of randomly mutated *gusA* alleles in a *lacZ* *E. coli* strain. Clones exhibiting increased β -galactosidase activity on LB agar plates containing 5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside (X-gal) were isolated. After two additional rounds of DNA shuffling (point mutation recombination, Ref. 19) and screening, clones exhibiting 500-fold increases in β -galactosidase activity were isolated. All clones were sequenced, and the fittest allele contained 13 mutations. We dissected the effects of these mutations through site-directed mutagenesis and found that four amino acid changes in an active site loop (T509A, S557P, N566S, K568Q) accounted for the changes in substrate specificity (1).

Here we extend a recently developed mutagenesis strategy (20–22) to direct the evolution of GUS variants with increased XYL activity. We randomized specificity-determining residues that were identified in our previous study (Ser⁵⁵⁷, Asn⁵⁶⁶, Lys⁵⁶⁸) by saturation mutagenesis (also called combinatorial cassette mutagenesis, Refs. 22–27). This strategy effected vast improvements in a single round of directed evolution. The selected clones were further evolved through random mutagenesis and point mutation recombination and high throughput screening (Table I).

EXPERIMENTAL PROCEDURES

Materials—Most of the materials, including the constitutive His₆-*gusA* expression vector, were previously described (13). The oligonucleotides (Fig. 2) were synthesized by IDT (Coralville, IA); the pNP- β -D-xylopyranoside (pNP-xyI) was from Sigma-Aldrich. The GeneAmp XL polymerase kit, including the rTth and Vent polymerases, was from Applied Biosystems (Foster City, CA).

Saturation Mutagenesis—The codons encoding *gusA* amino acids 557, 566, and 568 were randomized by whole plasmid PCR (28) using a mixture of Tth and Vent polymerase (29) and back-to-back 5'-phosphorylated degenerate primers: 5'-GCGCAATATGCCTTGKNNG-GTCGAAAATCGG-3' (*gusA* S557X) and 5'-GTTGGCGGTNNKAAGN-NKGGGATCTTCACTCGC-3' (*gusA* N566X,K568X), where K is T or G. The long PCR reactions contained 50 ng of the ancestral *gusA* expression vector, 500 nM primers, 200 nM of each dNTP, 1 \times ABI XL buffer II, and 1.5 mM magnesium acetate. The reactions were overlaid with light mineral oil and heated to 80 °C in a thermal cycler for a hot start. 0.5 μ l of the Tth/Vent mixture was added, and the temperatures were immediately raised to 94 °C for 1 min, followed by 30 cycles of 94 °C \times 15 s, 68 °C \times 5 min (1 min per kb). The reaction was further incubated at 72 °C for 10 min and stored at 4 °C.

The mineral oil was removed, and the PCR reaction was incubated with 0.5% SDS and 50 μ g/ml proteinase K at 65 °C for 15 min to eliminate the thermostable polymerases. The PCR product was purified using the Promega Wizard PCR prep kit (Madison, WI) as directed by the manufacturer, with the final elution in 50 μ l of water. The DNA was digested with restriction enzyme DpnI (to eliminate methylated ancestral template), and the desired PCR product was gel-purified using a QiaQuick spin column as directed by the manufacturer (Qiagen, Valen-

TABLE I
Summary of evolution experiments

The library was constructed in a different way for each round of evolution. The average selection coefficient (improvement in fitness over the ancestor) for each population was determined in quantitative whole cell xylosidase assays. We assayed 16 replicates of each clone, determined the average of those clones that exhibited improvement over its immediate ancestor, and divided by the xylosidase activity of the wild type.

Round of evolution	Diversity generation	Average selection coefficient
1	Saturation mutagenesis	30-fold
2	Random mutagenesis	80-fold
3	DNA shuffling (catalytic domain)	90-fold
4	StEP/shuffling	100-fold

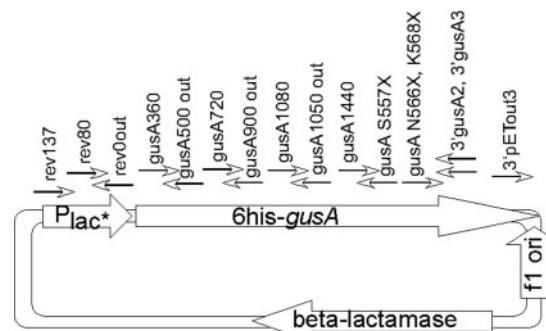


FIG. 2. PCR primers employed in this study. 14 primers were utilized for site-saturation mutagenesis, whole plasmid PCR, DNA shuffling, StEP, and overlap PCR (to subclone the products of DNA shuffling and StEP). Their orientations and approximate positions on a map of the ancestral *gusA* expression vector are represented as arrows.

cia, CA). The concentration of eluted DNA was estimated via agarose gel electrophoresis.

For intramolecular self-ligation of blunt-ended DNA, we recommend low concentrations of DNA and high concentrations of T4 DNA ligase. 4 fmol of purified PCR product were reacted with 2.5 units of T4 DNA polymerase and 5 Weiss units of T4 DNA ligase in 20- μ l reactions containing 200 nM each dNTP, 100 μ M ATP, 50 mM Tris, pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol, and 25 μ M bovine serum albumin, for 1 h at 23 °C. The T4 DNA ligase was heat-killed at 65 °C for 10 min, and DNA was purified by butyl alcohol precipitation (30). Freshly prepared *E. coli* Inv α F' cells were transformed by electroporation as described by Dower *et al.* (31). The transformed bacteria were titered, distributed into 384-well plates, and evaluated in our high throughput microplate assay (32) using pNP-xyI as the substrate. Clones associated with the highest absorbance at 405 nm after 16 h at 37 °C were streaked onto LB-ampicillin agar plates and characterized in a whole cell activity assay.

Whole Cell Activity Assays (Secondary Screens)—Microcultures associated with high xylosidase activity in primary screens were streaked

onto LB-ampicillin plates. Individual colonies were picked and grown to saturation in a 96-well microplate. 20 μ l of the saturated culture was placed in 200 μ l of 0.4 mM pNP- β -D-xylopyranoside in a 96-well plate, and the absorbance at 405 nm was monitored for 5 h. Each clone was evaluated in three independent trials, and the average values (and associated S.E.) were calculated. The activity of each clone was normalized to that of the ancestral strain.

Whole Plasmid PCR—The 14 plasmids selected in the first round of evolution were purified, pooled, and PCR-amplified as described above using 5'-GGGAAGCTTCTCATTGTTTGCCTCCCTGCTGCG-3' (3'-*gusA3*) and 5'-GGGAAGCTTGCAGCCGCGACTCGAGCAC-3' (3'-pE-Tout3). The amplification product was purified, digested with restriction enzymes DpnI and Hind III (to digest the sites underlined in the primer sequence), gel-purified, and self-ligated. The ligation reaction was similar to that described above, except that we used a lower concentration of T4 DNA ligase (1 unit instead of 5), no T4 DNA polymerase (nor dNTPs and bovine serum albumin), and a 16 °C incubation (instead of 23 °C). *E. coli* strain Inv α F' was transformed with the plasmid library, and the transformants were screened as described above.

DNA Shuffling—The *gusA* alleles isolated in the second round of evolution were pooled and amplified in a standard PCR using 500 nM primers 5'-GGACTTTGCAAGTGGTGAATCCGCAC-3' (*gusA* 720) and 3'-*gusA3*, 50 ng of template DNA, 60 mM Tris-HCl, pH 8.5, 15 mM (NH₄)₂SO₄, 2.0 mM MgCl₂, 0.2 mM each dNTP, and 5 units of *Taq* polymerase. The mixture was overlaid with mineral oil and cycled 25 times between 94 °C for 30 s and 72 °C for 2 min. The resulting PCR product was purified according to the Wizard PCR procedure and randomly recombined by DNA shuffling (19). We reacted the PCR product with 0.1 units of DNase I in 50 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, for 2 min at 23 °C. The DNA was immediately extracted in phenol/chloroform (50:50) and ethanol precipitated. The fragments were reassembled in a 45 cycle PCR-like reaction with no additional primers. The full-length recombinant products were amplified in the standard PCR reaction as described above and subcloned using overlap PCR as follows (33). The ancestral expression vector and part of the *gusA* gene were amplified in a long PCR (29) with 5'-CGGTTTGTGGTTAATCAG-GAAGTTCG-3' (*gusA* 900out) and 3'-pE-Tout3. Finally, the shuffled *gusA* mutants and the vector were combined in a long overlap PCR using primers 3'-*gusA3* and 3'-pE-Tout3. The resulting full-length PCR product was purified, restriction-digested, and self-ligated. *E. coli* strain Inv α F' was transformed with the plasmid library, and the transformants were screened for clones exhibiting increased XYL activity.

Staggered Extension Process (StEP)—Plasmids derived from clones isolated in the third round of evolution were purified and pooled. Three overlapping regions of the pooled *gusA* alleles were amplified separately in standard PCR reactions, except that 100 cycles with short (15 s) extension times were executed. This produces incomplete fragments that can recombine by switching templates (34). The 5'-end of the *gusA* gene pool was amplified with 5'-TAATCACCATTCCCGCGGATAG-TTC-3' (*gusA* 500out) and 5'-TCTAGATCTGGCAGCAGGTTTCCCG-ACTG-3' (rev137), the central region was amplified with 5'-GCCATT-GAAGCCGATGTCACGCCG-3' (*gusA* 360) and 5'-CCTGTAAGTG-CGCTTGCTGAGTTTCC-3' (*gusA* 1150out), and the 3'-end was amplified with 5'-GGGAAGCTTATGTTTGCCTCCCTGCGGT-TTTTCA-3' (3'-*gusA2*) and 5'-CTGCTGCTGCTGGCTTTAACCTCTC-T-3' (*gusA* 1080). The three amplification products combined and amplified in an overlap PCR using primers rev137 and 3'-*gusA2*. The vector was amplified with 5'-AGCTGTTTCCCTGTGTGAAATTGTATC-C-3' (rev0out) and 3'-pE-Tout3. The vector and insert were then assembled in a long overlap PCR using primers 3'-pE-Tout3 and 3'-*gusA2*. The resulting library was purified, restriction-digested, and self-ligated. *E. coli* strain Inv α F' were transformed with the resulting library, and the transformants were screened for clones with improved XYL activity.

DNA Sequencing—The evolved *gusA* alleles were sequenced using the Applied Biosystems Big Dye protocol at the Center for Fundamental and Applied Evolution (Emory University).

The 3'-end of every evolved *gusA* allele was sequenced using the primer 5'-GCTCAGCGTGGCAGCAGCCAACTC-3' (GC 3'pET). Particularly interesting alleles (namely 1.15, 2.10, 4.7, and 4.8) were sequenced in their entirety, using the following additional primers: 5'-ATGCTCCGGCTCGTATGTTGTGTGG-3' (rev 80), *gusA* 360, 5'-GGACTTTGCAAGTGGTGAATCCGCAC-3' (*gusA* 720), and *gusA* 1080.

Protein Purification and Characterization—Each GUS protein was fused to an N-terminal His₆ tag, and the enzymes 1.15, 2.10, 4.7, and 4.8 were purified to homogeneity (as judged by SDS-PAGE analysis) using nickel chelate affinity chromatography. The total protein concentration was quantified using the Bradford protein assay (Bio-Rad). 5 nM to 1 μ M

of each purified protein were separately reacted with 1 ml of pNP- β -D-xylopyranoside (concentrations ranging from 10 μ M to 1 mM) in 50 mM Tris-HCl buffer (pH 7.6), and the formation of the pNP product was followed in a spectrophotometer. The steady-state kinetic parameters were determined by fitting the initial velocity values to the Michaelis-Menten equation (35). Each of the values (k_{cat} , K_m , k_{cat}/K_m) reported in Table III (and the associated S.E.) is an average of three independent trials.

Molecular Modeling—We visualized crystal structures, made site-directed mutants *in silico* and performed energy minimization calculations using SYBYL (Tripos, St. Louis, MO) at the Biomolecular Computing Resource (BimCore) at Emory University.

RESULTS

Saturation Mutagenesis—Our objective was to direct the evolution of GUS variants with XYL activity. The ancestral plasmid for these experiments was a constitutive *gusA* expression vector constructed for a previous study (13). We first employed gene site saturation mutagenesis to randomize the sequences of three residues: 557, 566, and 568, which were identified in a different directed evolution experiment (1). We encoded all twenty amino acids in a library containing 32³ (or 32,768) different clones, which is about three times the throughput of our screen. The library was generated by whole plasmid PCR (28, 37, 38)² using primers containing the degenerate NNK (where K is T or G) sequence. The PCR product was purified, self-ligated, and transformed into *E. coli* strain Inv α F'. *E. coli* K-12 and its derivatives, including Inv α F', do not contain endogenous XYL activity (39).

We evaluated ~10,000 GUS mutants in a semi-automated, high throughput microplate screen (32). We used a microplate dispenser to distribute transformed cells into 77 \times 384 well microtiter plates such that each of the 29,568 wells received an average of one cell in 5 μ l of LB-ampicillin medium. As a control, 3 \times 384 microcultures were seeded with cells transformed with the ancestral *gusA* expression vector. The 80 microplates were manually sealed with autoclaved silicone seals and inverted end-over-end in an environmental rotator for 16 h at 37 °C. The microcultures grew to saturation under these conditions, and the GUS proteins were constitutively expressed at high levels. The seals were removed, and 75 μ l of 1.0 mM pNP-xyI in 50 mM Tris, pH 7.6, were added to each well. The microplates were incubated at a 45° angle (so that the cells settled into the corner of the wells) at 37 °C for 16 h. The XYL activity associated with each of the 29,568 microcultures was measured in a microplate spectrophotometer. Reconstruction experiments showed that all XYL activity was associated with the cells, rather than the supernatant, suggesting that pNP-xyI was somehow entering the cytoplasm. A Microsoft Excel macro facilitated the identification of the microcultures containing the most XYL activity.

The vast majority of the GUS mutants exhibited less XYL activity than the ancestral controls (Fig. 3). We were not surprised because nearly all of the clones were supposed to contain random mutations in three active site residues. The 24 fittest clones were characterized in a more quantitative assay to demonstrate that the increases in xylosidase activity were reproducible (secondary screen). 16 replicates of each clone were propagated in a 384-well plate, in parallel with 48 replicates of the ancestral clone. Each of the cultures was reacted with 1.0 mM pNP-xyI for 16 h at 37 °C. The activity of each clone was calculated by averaging the ΔA_{405} of 16 replicates. The improvement initially exhibited by 10 of 24 selected clones was not reproducible, perhaps because the corresponding plasmids contained mutations that caused GUS expression to become genetically unstable. These ten clones

² L. A. Rowe and I. Matsumura, submitted manuscript.

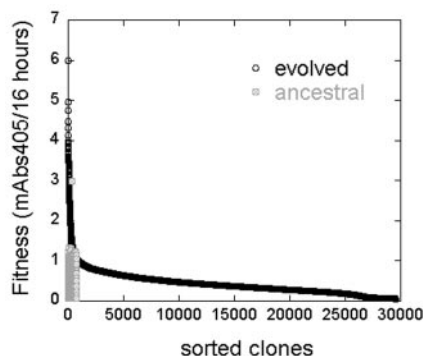


FIG. 3. β -Xylosidase activity of round 1 GUS mutants. Mutants were generated using saturation mutagenesis of residues 557, 566, and 568. The resulting library was subcloned, transformed, and expressed in *E. coli*. The transformed cells were distributed into 77×384 -well plates and propagated to saturation. 1 mM pNP-xylopyranoside (or 0.4 mM in rounds 2–4) were added to of the $\sim 30,000$ microcultures. The reactions were incubated at 37 °C for 16 h. Most of the mutant microcultures exhibited less xylosidase activity than the ancestral cultures (gray). The best 24 were isolated and characterized for further study.

were discarded. The average XYL activity of the remaining 14 clones was 30-fold higher than that of the ancestor. The best clone evinced 70-fold improvement in fitness (clone 1.11, Table II and Fig. 4). In our previous study, we found that random mutagenesis of the whole gene generated mutants with only 2–4-fold increases in β -galactosidase activity (40). The higher values in the current study support the validity of our saturation mutagenesis strategy.

The 3'-quarter of each of the 14 selected *gusA* genes were sequenced. This region (nucleotides 1320–1812) contains the three randomized codons and about half of the catalytic domain (12). The 14 selected clones collectively contained 10 of 20 amino acids at position 557, 6 amino acids at position 566, and 9 amino acids at position 568 (Table II). These data suggest that a variety of amino acids, including ones that differ in size and charge, can be accommodated in the active site loop. The relatively conservative nature of the mutations in the fittest clone (1.11), S557C/N566S/K568, suggests that certain mutation combinations are epistatic (non-additive in effect). We found four mutations in positions other than 557, 566, and 568 (Table II), excluding clone 1.15. This corresponds to a mutation frequency (1:1600) similar to the published value for long PCR reactions using the Tth and Vent polymerases (1:700, Ref. 41). Clone 1.15 contains a total of 14 mutations, and is identical to the *gusA* variant with increased β -galactosidase activity that was evolved in our laboratory (1). Its inclusion in the library was very likely a result of a cross-contamination event. We did not realize this until much later and did not exclude this clone from the evolving population. Clone 1.15 was not the fittest, however, and only mutations in positions 557, 566, and 568 survived subsequent rounds of evolution (see below).

Random Mutagenesis and Recombination—We sought beneficial mutations in residues other than 557, 566, and 568, so we randomly mutated the selected *gusA* genes by mutagenic PCR. The entire vector was amplified in a slightly mutagenic PCR² using a mixture of Tth and Vent polymerases (29) and back-to-back phosphorylated primers. The PCR product was purified, polished with T4 DNA polymerase, and self-ligated with T4 DNA ligase. *E. coli* Inv α F' was transformed with the plasmid library and distributed into 384-well microplates. The microcultures were propagated overnight and reacted with 0.4 mM pNP-xyl for 3 h. In a secondary screen, the 24 clones exhibiting the most improvement in XYL activity were assayed in parallel with the ancestor, in this case the fittest clone from the first round of screening, and 13 proved reproducibly fitter (Fig. 4).

On average, these 13 clones isolated in the second round of evolution evinced an 80-fold improvement in xylosidase activity over its wild-type ancestor. We sequenced the 3'-quarter of each of the selected alleles and observed, in contrast to round 1, that only two sequence combinations emerged from the second round of screening, Pro⁵⁵⁷/Ala⁵⁶⁶/Phe⁵⁶⁸ (hereafter called PAF) and Pro⁵⁵⁷/Ser⁵⁶⁶/Gln⁵⁶⁸ (PSQ). 10 of 13 round 2 isolates contained PAF, and the remaining three contained PSQ. All of the other 557/566/568 sequence combinations selected in the first round of evolution were driven to extinction by the PSQ and PAF combinations (Fig. 4).

The PAF combination did not exist after the first round of screening, but was most frequent among the 13 clones selected in the second round. The mostly likely first round parent contained Lys⁵⁵⁷, Ala⁵⁶⁶, and Phe⁵⁶⁸ (KAF, clone 1.1). Lysine cannot be converted into proline with a single nucleotide change, so the 10 PAF clones probably arose through recombination (between codons 557 and 566) during whole plasmid PCR. Recombination occurs at low frequency during regular (42) and long (43) PCR through template switching of incomplete strands. The PSQ combination was represented by a single clone among those isolated in the first round of screening (clone 1.15), but this clone contained eleven additional mutations in residues other than 557, 566, and 568, including five in the latter quarter of the gene (three 5' of codon 557 and two 3' of 568). Its putative second round descendants did not contain any of the latter mutations. We believe that the additional mutations were collectively slightly deleterious and that the round 2 PSQ clones were products of two recombination events during long PCR. In other words, all 13 of the clones selected in the second round were probably produced through recombination rather than random mutagenesis.

The sequence analysis of the clones selected in the second round underscores the utility of random recombination in laboratory evolution relative to that of recursive random mutagenesis (13, 19). The recombination rate associated with PCR, however, is relatively modest (42). We sought higher recombination frequencies, and therefore employed DNA shuffling (19) to recombine the existing mutations (including those at 557, 566, 568, and other loci) in the catalytic domain. We used overlap PCR to subclone the recombinant library back into the expression vector (33). The library was transformed into *E. coli* Inv α F' and propagated in 384-well microtiter plates as described above, and $\sim 10,000$ clones were screened for activity in reactions with pNP-xyl. The secondary screen showed that eleven clones reproducibly exhibited improved XYL activity, but on average the improvement over the second round was relatively modest (90-fold improvement over the ancestor, compared with an 80-fold average improvement among round 2 clones, Table I and Fig. 4).

The conventional DNA shuffling of point mutations is associated with a 0.7% mutation rate (19), which corresponds to an average of ~ 13 mutations per 1.8 kilobase allele. To decrease the mutation frequency, we randomly recombined the mutations using the alternative, possibly less mutagenic StEP (34). The *gusA* alleles isolated in the third round of evolution were pooled and purified. Three overlapping ~ 700 -bp regions were separately PCR-amplified using *gusA*-specific primers with a 15-s extension time (Fig. 2). The three StEP amplification products were combined, and a full-length product was amplified in an overlap PCR using primers external to the gene. The recombinant library was subcloned back into the expression vector by overlap PCR reaction as described above. The recombinant library was transformed into *E. coli*. The expression library was propagated and screened for clones with improved XYL activity. The secondary screen revealed that one clone

TABLE II
Analysis of mutants isolated in first round of evolution

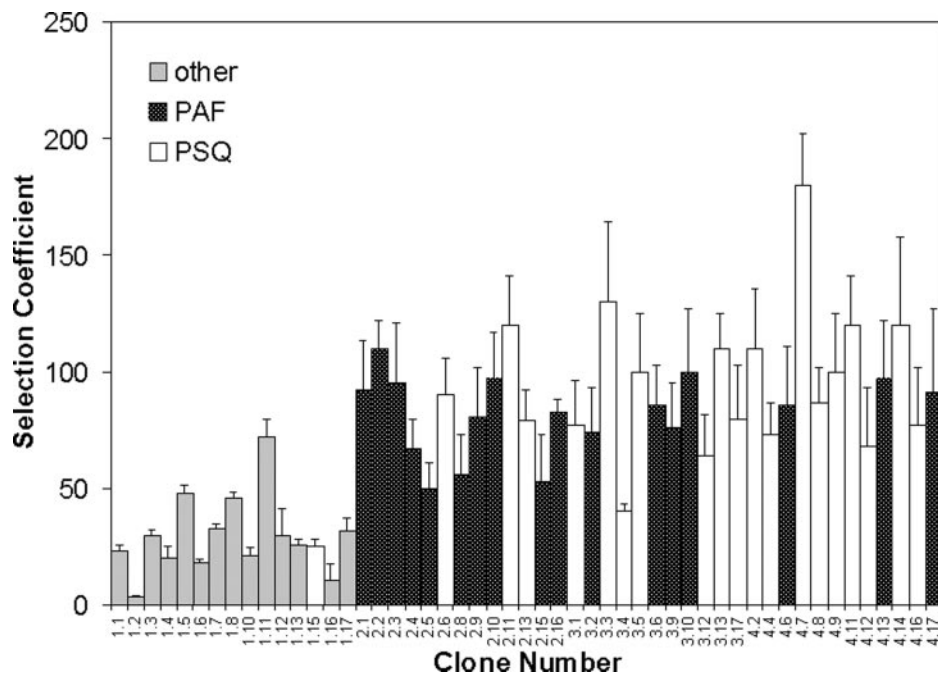
Variant	557	566	568	Other mutations ^a	Relative fitness ^b
WT ^c	Ser	Asn	Lys		1
1.1	Lys	Ala	Phe		23 ± 3
1.2	Gln	Lys	Ser	Q598stop	3.3 ± 1
1.3	Pro	Ser	Val		30 ± 2
1.4	Tyr	Glu	Gly		20 ± 5
1.5	Asp	Asn	Phe		48 ± 4
1.7	Leu	Cys	Met		33 ± 2
1.8	Asp	Asn	Leu		46 ± 2
1.10	Ser	Ser	Lys		21 ± 3
1.11	Cys	Ser	Lys		72 ± 8
1.12	Cys	Ser	Lys		30 ± 10
1.13	Val	Cys	Met	D531E	26 ± 2
1.15 ^a	Pro	Ser	Gln	S22N, G81S, K257E, T509A, Q598R, stop604W	25 ± 4
1.16	Ile	Ala	Arg	A580V	11 ± 7
1.17	Pro	Cys	Phe		32 ± 5

^a The 3'-end of each *gusA* allele (encoding codons 440–604) isolated in the first round of screening was sequenced. Clone 1.15 was sequenced in its entirety. It is identical to a variant that was evolved to utilize β -galactosides (1), and is almost certainly a result of cross-contamination.

^b The selection coefficient (relative fitness) was determined by reacting clonal cultures with *p*-nitrophenyl- β -D-xylopyranoside, measuring the rate of product formation by monitoring A_{405} in a microplate spectrophotometer and dividing by the ancestral rate.

^c WT, wild type.

FIG. 4. Selection coefficients of evolved xylosidases. The fittest clones from each of four rounds of directed evolution (50 in all) were isolated from single colonies and propagated side-by-side in a 96-well plate (data represent averages of independent experiments conducted on three separate days). 20 μ l of each of the saturated cultures were reacted with 0.4 mM *p*-nitrophenyl- β -D-xylopyranoside in 50 mM Tris, pH 7.6. Product formation was followed at 405 nm for 5 h at 23 °C. Each of the rates were divided by the ancestral value, so that the wild-type value on this graph would be 1.



(4.7) exhibited ~200-fold improvement over the wild type, whereas the remainder were comparable in fitness to those isolated in the third round (Fig. 4).

Kinetics of Evolved Mutants—We selected the most interesting clones for further purification and characterization, including the fittest PAF (clone 2.10) and three PSQ variants: the fittest one (clone 4.7), a sibling PSQ with activity closer to the average of all PSQs (clone 4.8), and the likely ancestor of the latter two (clone 1.15). These clones and the ancestral His₆-GUS were separately expressed in *E. coli* and purified to homogeneity by immobilized metal affinity chromatography. Each of the purified enzymes was reacted with varying concentrations of pNP- β -D-glucuronide (pNP-gluc, native substrate) or pNP-xyI (novel substrate). The second order rate constant (k_{cat}/K_m), which is a measure of the rate of capture of substrate by an enzyme into a productive complex (44) of each enzyme-substrate reaction was determined (Table III). Clone 2.10, the fittest PAF variant, exhibited a ~200-fold improvement in k_{cat}/K_m in reactions with pNP-xyI. It displayed a slight (~2.5-fold = 590/240) preference for pNP-xyI over pNP-gluc, in con-

trast to the 117,000-fold (340,000/2.9) preference of the wild-type enzyme for β -glucuronides. Overall the PAF mutations in clone 2.10, which is otherwise wild type, impart a 290,000 ($2.5 \times 117,000$) shift in specificity with respect to the wild-type enzyme.

The PSQ variants were also purified, and their steady-state kinetic parameters in reactions with pNP-xyI and pNP-gluc were determined. Enzyme 4.7, like 2.10, exhibits a slight (2.4-fold) substrate preference for pNP-xyI over pNP-gluc, its k_{cat}/K_m in reactions with pNP-xyI is 80-fold higher than the comparable wild-type value. Clone 4.7 was the fittest in the context of the whole cell time course assay, but the corresponding purified enzyme exhibits less xylosidase activity (k_{cat}/K_m) than does enzyme 2.10. These results demonstrate that k_{cat}/K_m is not the sole determinant of fitness in the whole cell assay. Enzymes 4.7, 4.8, and 1.15 all contain the PSQ mutations, but 4.8 and 1.15 also have additional mutations. Enzyme 4.7 has the highest k_{cat} of all enzymes characterized, but also the highest K_m . Enzyme 4.8 has two missense (V473A and G601S) and 5 silent mutations. Enzyme 4.7 contains three additional

TABLE III
 Kinetic parameters of the evolved enzymes

Substrate enzyme	pNP- β -D-xylopyranoside			pNP- β -D-glucuronide		
	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m
	s^{-1}	μM	$s^{-1} M^{-1}$	s^{-1}	μM	$s^{-1} M^{-1}$
WT ^a GUS	0.0010 \pm 0.0001	360 \pm 60	2.9 \pm 0.3	68 \pm 6	200 \pm 10	340,000 \pm 20,000
1.15	0.054 \pm 0.003	660 \pm 50	81 \pm 3	0.039 \pm 0.001	210 \pm 10	190 \pm 10
2.10	0.063 \pm 0.005	110 \pm 10	590 \pm 30	0.042 \pm 0.002	180 \pm 20	240 \pm 20
4.7	ND ^b	ND	240 \pm 10	0.012 \pm 0.001	220 \pm 10	54 \pm 2
4.8	ND	ND	190 \pm 10	0.021 \pm 0.004	280 \pm 90	80 \pm 10

^a WT, wild type.

^b Not determined.

silent mutations whereas 1.15 contains six missense (S22N, G81S, K257E, T509A, Q598R, stop604W) and five silent mutations. These eleven mutations in clone 1.15 are collectively deleterious, both with respect to fitness in the whole cell assay and the *in vitro* xylosidase activity, so it is clear why recombination and high throughput screening led to their extinction.

The steady-state kinetic parameters of enzymes 2.10, 4.7, and 4.8 may explain why the PAF and PSQ combinations persisted over four rounds of intense selection. The k_{cat} of 4.8 in reactions with pNP-xyl is higher than that of 2.10, but its K_m is also higher. The two enzymes are similar in specific activity when the pNP-xyl concentration is 0.4 mM, which was the substrate concentration in the high throughput screen. If we had used higher pNP-xyl concentrations, it seems likely that the PSQ mutations would have become fixed (ubiquitous). Conversely, if we had used lower substrate concentrations the PAF form would have driven its competitor into extinction.

Are the PAF and PSQ Combinations Special?—A wide variety of missense mutations at positions 557, 566, and 568 can impart increases in XYL activity (Table II, round 1 clones). All combinations except PAF and PSQ, however, were driven into extinction in the second round of evolution. We wished to learn whether the emergence of these forms occurred through neutral drift or through adaptive evolution. In the drift scenario, all of the 557/566/568 mutation combinations were similar in evolutionary potential, and the PAF and PSQ combinations were amplified by chance. In the adaptation scenario, the PAF and PSQ mutation combinations were selected because they were beneficial and synergistic in effect (epistasis). We investigated these scenarios by using site-directed mutagenesis to impose the recombinant PSF (instead of PSQ) and PAQ (instead of PAF) mutation combinations upon otherwise wild-type scaffolds. *E. coli* cells expressing these mutant proteins were viable, indicating that the newly formed protein was not toxic, but neither enzyme exhibited detectable XYL activity. These results in combination suggest that the PAF and PSQ combinations each represent synergistic mutation sets.

DISCUSSION

The GH-A glycoside hydrolases are believed to share a common (β/α)₃-barrel fold (9). The active site is at the face defined by the C-terminal ends of the β -strands and is composed of residues from seven of the loops that connect the β -strands to the α -helices. Conserved amino acids in the active site can be superimposed (Fig. 1). The β -glucuronide and β -xyloside substrates differ only in their C5 substituents, so it is tempting to suppose that a single amino acid replacement in GUS would be sufficient to convert it into a xylosidase. Straightforward alterations of enzyme substrate specificity have been reported (45), but the functional divergence of the GH-A enzymes was more complicated. The C α traces of the active site loops, including those that contain the superimposable residues (loops 2, 4, 6, and 7), are divergent (not shown).

Loop 8 is the most divergent part of the active site and is

most responsible for the C5 specificity of GH-A enzymes. The residues that we randomized are part of this loop. Molecular modeling suggests that the C6 carboxylate oxygens of the β -glucuronide substrate form hydrogen bonds with Tyr⁵⁴⁹, Asn⁵⁶⁶, Lys⁵⁶⁸ (all on loop 8), Tyr⁴⁶⁸ (loop 6), and Asp¹⁶³ (domain 1) in the wild-type GUS (1). The crystal structure of the *T. saccharolyticum* β -xylosidase (11) shows hydrophobic interactions between C5 of the enzyme-bound xylose intermediate and Phe³³⁵, Trp³¹⁵ (both loop 8), and Tyr²⁸² (loop 7). A crystal structure of the *E. coli* β -galactosidase (GH-A, Family 2) bound to a transition state analog (46) shows that the C6 hydroxyl group interacts directly with Asn⁶⁰⁴, Trp⁵⁶⁸, (both loop 8), His⁵⁴⁰ (loop 7), Asp²⁰¹ (domain 1), and a bound sodium ion.

Molecular modeling of the PAF and PSQ mutations with energy minimization revealed no large-scale structural changes within the active site. The S557P mutation was fixed after the second round of evolution, and we propose that it increases the conformational flexibility of loop 8. We believe that Phe⁵⁶⁸ (clone 2.10) participates in a hydrophobic interaction with C5 of pNP-xyl, and that the N566A mutation allows Phe⁵⁶⁸ to assume a more optimal binding conformation. The PSQ mutation combination is also apparently synergistic. We propose that Ser⁵⁶⁶ and Gln⁵⁶⁸ form a hydrogen bond, thereby stabilizing the conformation of loop 8. These mutant residues do not interact with β -xylosides, so the K_m of the complex with pNP-xyl is higher than the corresponding value for the PAF complex. The conformational stability of the active site, however, results in a higher k_{cat} . The most efficient xylosidases, enzymes 2.10 and 4.7, did not contain mutations other than those at residues 557, 566, and 568. We therefore suppose that the seven such mutations that emerged in our study, Q598stop (clone 1.2), D531E (1.13), A580V (1.16), F582L (3.3), V473A (4.8), G601S (4.8), and F582S (4.14), are neutral or slightly deleterious in effect.

We previously employed whole gene random mutagenesis and DNA shuffling to direct the evolution of *E. coli* GUS variants with increased β -galactosidase activity. In those experiments, 16 of 17 clones isolated in the first round of screening were mutated in one of three positions (509, 557, or 566), and each of the single mutations (T509A, S557P, or N566S) in isolation imparted a 2–4-fold increase in β -galactosidase activity (k_{cat}/K_m). The K568Q mutation is deleterious in isolation but confers a 25-fold increase in k_{cat}/K_m in the context of T509A/S557P/N566S (1). In this study, we randomized positions 557, 566, and 568 and isolated 14 clones that were on average 30-fold fitter in the context of our whole cell time course assay. Whole plasmid PCR amplification of the 17 alleles apparently resulted in random recombination and the generation of even fitter variants (80-fold fitter on average than the ancestor).

Other groups have employed random mutagenesis and screening to identify residues that determine substrate specificity (20, 21) or thermostability (22). These residues were

subsequently randomized, and high throughput screening of the resulting libraries led to the identification of novel sequence variants. These experiments, however, led to relatively modest improvements in fitness. In one study the benzoylformate decarboxylase from *Pseudomonas putida* was randomly mutated and a variant exhibiting a 5-fold increase in carbonylase activity was isolated. A single residue (Leu⁴⁷⁶) was randomized; the resulting library was screened but none of the second generation clones showed improvement over the original L476Q variant (21). In a separate study, the glutaryl acylase of *Pseudomonas* SY-77 was randomly mutated. Sequence variants that enabled growth on a novel adipyl substrate were identified. The fittest variant (Y178H) exhibited a ~3-fold improvement in k_{cat}/K_m when reacted with a novel adipyl-7-ADCA substrate. Three contiguous residues (177–179) were randomized together, and several new clones were identified in the genetic selection. None of these new clones, however, exhibited greater catalytic efficiency than Y178H (20).

In vitro evolution studies of different proteins cannot of course be directly compared. We tentatively propose that the success of any site saturation mutagenesis experiment is predicated upon the generation of synergistic mutations. The residues we chose to randomize in this study were based on information derived from three rounds of mutagenesis and DNA shuffling (1). It is very unlikely that epistatic mutations will emerge in a single round of whole gene random mutagenesis. Had we selected our target residues based on our first round of evolution (509, 557, or 566), we would not have chosen to randomize 568. Sites for saturation mutagenesis could also be selected through the computational analysis of potential mutations, an approach that has recently been actualized by several groups (47–49).

This study also illustrates the remaining challenges facing practitioners of directed evolution. The PAF and PSQ mutations should have been present in the site saturation mutagenesis library, but we did not isolate these until the second and fourth rounds of evolution, respectively. The rate of adaptation decreased after each round of *in vitro* evolution, suggesting that single mutations that improve fitness are uncommon. The most active evolvant (enzyme 2.10) remains ~1000-fold worse (k_{cat}/K_m) than the wild-type *T. saccharolyticum* β -xylosidase (36). Site saturation mutagenesis and screening apparently drove the population to local maxima upon the fitness landscape, so further increases in xylosidase activity will probably require more extensive site saturation mutagenesis. We predict that the randomization of charged and polar residues near Phe⁵⁶⁸ in clone 2.10 will lead to the evolution of a variant completely hydrophobic pocket. This pocket should increase affinity for the xyloside substrate and increase the conformational flexibility of the new active site, thus leading to even greater specificity and catalytic efficiency. We are currently investigating this idea.

Acknowledgments—We thank Lori Rowe and the other members of the Matsumura group for helpful discussion and Dr. Wayne Patrick for critical reading of the manuscript. We are grateful to Dr. Kim Gernert for assistance with the molecular modeling software and to Prof. Dale Edmondson for his guidance with the data analysis.

REFERENCES

- Matsumura, I., and Ellington, A. D. (2001) *J. Mol. Biol.* **305**, 331–339
- Kacsar, H., and Beeby, R. (1984) *J. Mol. Evol.* **20**, 38–51
- Lazcano, A., Diaz-Villagomez, E., Mills, T., and Oro, J. (1995) *Adv. Space Res.* **15**, 345–356
- Jensen, R. A. (1976) *Annu. Rev. Microbiol.* **30**, 409–425
- Ycas, M. (1974) *J. Theor. Biol.* **44**, 145–160
- Waley, S. G. (1969) *Comp. Biochem. Physiol.* **30**, 1–11
- Bourne, Y., and Henrissat, B. (2001) *Curr. Opin. Struct. Biol.* **11**, 593–600
- Henrissat, B., and Davies, G. (1997) *Curr. Opin. Struct. Biol.* **7**, 637–644
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7090–7094
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5674
- Yang, J. K., Yoon, H. J., Ahn, H. J., Il Lee, B., Pedelacq, J. D., Liang, E. C., Berendzen, J., Laivenieks, M., Vieille, C., Zeikus, G. J., Vocadlo, D. J., Withers, S. G., and Suh, S. W. (2004) *J. Mol. Biol.* **335**, 155–165
- Jain, S., Drendel, W. B., Chen, Z. W., Mathews, F. S., Sly, W. S., and Grubb, J. H. (1996) *Nat. Struct. Biol.* **3**, 375–381
- Rowe, L. A., Geddie, M. L., Alexander, O. B., and Matsumura, I. (2003) *J. Mol. Biol.* **332**, 851–860
- Farinas, E. T., Bulter, T., and Arnold, F. H. (2001) *Curr. Opin. Biotechnol.* **12**, 545–551
- Zhao, H., Chockalingam, K., and Chen, Z. (2002) *Curr. Opin. Biotechnol.* **13**, 104–110
- Schmidt-Dannert, C. (2001) *Biochemistry* **40**, 13125–13136
- Harris, J. L., and Craik, C. S. (1998) *Curr. Opin. Chem. Biol.* **2**, 127–132
- Powell, K. A., Ramer, S. W., Del Cardayre, S. B., Stemmer, W. P., Tobin, M. B., Longchamp, P. F., and Huisman, G. W. (2001) *Angew. Chem. Int. Ed. Engl.* **40**, 3948–3959
- Stemmer, W. P. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10747–10751
- Sio, C. F., Riemsma, A. M., van der Laan, J. M., Verhaert, R. M., and Quax, W. J. (2002) *Eur. J. Biochem.* **269**, 4495–4504
- Lingen, B., Grotzinger, J., Kolter, D., Kula, M. R., and Pohl, M. (2002) *Protein Eng.* **15**, 585–593
- Miyazaki, K., and Arnold, F. H. (1999) *J. Mol. Evol.* **49**, 716–720
- Coco, W. M., Encell, L. P., Levinson, W. E., Crist, M. J., Loomis, A. K., Licato, L. L., Arensdorf, J. J., Sica, N., Pienkos, P. T., and Monticello, D. J. (2002) *Nat. Biotechnol.* **20**, 1246–1250
- DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z., Tomlinson, G., Huang, H., Tan, X., Bibbs, L., Chen, P., Kretz, K., and Burk, M. J. (2003) *J. Am. Chem. Soc.* **125**, 11476–11477
- el Hawrani, A. S., Sessions, R. B., Moreton, K. M., and Holbrook, J. J. (1996) *J. Mol. Biol.* **264**, 97–110
- Ness, J. E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T. V., Govindarajan, S., Mundorf, E. C., and Minshull, J. (2002) *Nat. Biotechnol.* **20**, 1251–1255
- Graham, L. D., Haggett, K. D., Jennings, P. A., Le Brocq, D. S., Whittaker, R. G., and Schober, P. A. (1993) *Biochemistry* **32**, 6250–6258
- Eisinger, D. P., and Trumpower, B. L. (1997) *BioTechniques* **22**, 250–252, 254
- Cheng, S., Fockler, C., Barnes, W. M., and Higuchi, R. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5695–5699
- Thomas, M. R. (1994) *BioTechniques* **16**, 988–990
- Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16**, 6127–6145
- Geddie, M. L., Rowe, L. A., Alexander, O. B., and I., M. (2004) *Methods Enzymol.* **388**, 134–145
- Temesgen, B., and Eschrich, K. (1996) *BioTechniques* **21**, 828, **830**, 832
- Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) *Nat. Biotechnol.* **16**, 258–261
- Matsumura, I., Wallingford, J. B., Surana, N. K., Vize, P. D., and Ellington, A. D. (1999) *Nat. Biotechnol.* **17**, 696–701
- Vocadlo, D. J., Wicki, J., Rupitz, K., and Withers, S. G. (2002) *Biochemistry* **41**, 9727–9735
- Stemmer, W. P., Morris, S. K., Kautzer, C. R., and Wilson, B. S. (1993) *Gene (Amst.)* **123**, 1–7
- Parikh, A., and Guengerich, F. P. (1998) *BioTechniques* **24**, 428–431
- Whitehead, T. R. (1997) *Curr. Microbiol.* **35**, 282–286
- Matsumura, I., and Ellington, A. D. (2001) in *In Vitro Mutagenesis Protocols* (Braman, J., ed) Vol. 182, pp. 259–267, Humana Press, Totowa, NJ
- Stewart, A. C., Gravitt, P. E., Cheng, S., and Wheeler, C. M. (1995) *Genome Res.* **5**, 79–88
- Judo, M. S., Wedel, A. B., and Wilson, C. (1998) *Nucleic Acids Res.* **26**, 1819–1825
- Shafikhani, S. (2002) *Environ Microbiol.* **4**, 482–486
- Northrup, D. B. (1998) *J. Chem. Education* **75**, 1153–1157
- Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R., and Holbrook, J. J. (1988) *Science* **242**, 1541–1544
- Juurs, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G., and Matthews, B. W. (2001) *Biochemistry* **40**, 14781–14794
- Voigt, C. A., Mayo, S. L., Arnold, F. H., and Wang, Z. G. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3778–3783
- Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) *Nature* **423**, 185–190
- Hayes, R. J., Bentzien, J., Ary, M. L., Hwang, M. Y., Jacinto, J. M., Vielmetter, J., Kundu, A., and Dahiyat, B. I. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15926–15931