An evolutionary route to xylanase process fitness

Nisha Palackal, Yali Brennan, Walter N. Callen, Paul Dupree, Gerhard Frey, Florence Goubet, Geoffrey P. Hazlewood, Shaun Healey, Young E. Kang, Keith A. Kretz, Edd Lee, Xuqiu Tan, Geoffery L. Tomlinson, John Verruto, Vicky W.K. Wong, Eric J. Mathur, Jay M. Short, Dan E. Robertson and Brian A. Steer

Protein Sci. 2004 13: 494-503; originally published online Jan 10, 2004;
Access the most recent version at doi:10.1110/ps.03333504

An evolutionary route to xylanase process fitness

NISHA PALACKAL,[1] YALI BRENNAN,[1] WALTER N. CALLEN,[1] PAUL DUPREE,[2] GERHARD FREY,[1] FLORENCE GOUBET,[2] GEOFFREY P. HAZLEWOOD,[1] SHAUN HEALEY,[1] YOUNG E. KANG,[1] KEITH A. KRETZ,[1] EDD LEE,[1] XUQIU TAN,[1] GEOFFERY L. TOMLINSON,[1] JOHN VERRUTO,[1] VICKY W.K. WONG,[2] ERIC J. MATHUR,[1] JAY M. SHORT,[1] DAN E. ROBERTSON,[1] AND BRIAN A. STEER[1]

[1]Diversa Corp., San Diego, California 92121, USA
[2]Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK

(RECEIVED July 29, 2003; FINAL REVISION September 22, 2003; ACCEPTED October 27, 2003)

Abstract

Directed evolution technologies were used to selectively improve the stability of an enzyme without compromising its catalytic activity. In particular, this article describes the tandem use of two evolution strategies to evolve a xylanase, rendering it tolerant to temperatures in excess of 90°C. A library of all possible 19 amino acid substitutions at each residue position was generated and screened for activity after a temperature challenge. Nine single amino acid residue changes were identified that enhanced thermostability. All 512 possible combinatorial variants of the nine mutations were then generated and screened for improved thermal tolerance under stringent conditions. The screen yielded eleven variants with substantially improved thermal tolerance. Denaturation temperature transition midpoints were increased from 61°C to as high as 96°C. The use of two evolution strategies in combination enabled the rapid discovery of the enzyme variant with the highest degree of fitness (greater thermal tolerance and activity relative to the wild-type parent).

Keywords: Directed evolution; thermal stability; thermophilic; xylanase

Naturally thermostable proteins have been studied extensively to gain insight into the adaptive mechanisms for achieving thermostability and to decipher the general rules that could be applied to stabilizing other mesophilic proteins. Enzyme sequences from mesophilic organisms have been compared to their homologs from thermophilic and hyperthermophilic microorganisms in efforts to identify the interactions responsible for conferring enhanced thermostability (Adams 1993; Ladenstein and Antranikian 1998; Britton et al. 1999; Kinjo and Nishikawa 2001). Numerous mutagenesis studies have also explored this phenomenon; however, the mechanisms that mediate protein thermostabilization are still poorly understood (Adams and Kelly 1998; Farinas et al. 2001). Considerable amino acid sequence differences between mesophilic and thermophilic proteins greatly complicate attempts to identify specific substitutions that affect thermostability. Moreover, large changes in thermostability are usually the result of cumulative small contributions of a number of mutations (van den Burg et al. 1998; Gray et al. 2001). Rules for engineering substantial protein stability by rational design are likely to be protein specific, and any such design effort in combination with structural information is currently beyond state of the art approaches.

A more facile, nonrational approach to improving the thermal properties of an enzyme is directed evolution. We have used Gene Site Saturation Mutagenesis (GSSM) technology (Short 2001b) in combination with GeneReassembly technology (Short 2003) to evolve thermostability characteristics into an enzyme. The GSSM method is a highly efficient mutagenesis technique that can be used to random-

Reprint requests to: Brian A. Steer, Diversa Corp., 4955 Directors Place, San Diego, CA 92121, USA; e-mail: bsteer@diversa.com; fax: (858) 526-5764.

Abbreviations: ANTS, 8-aminonaphthalene-1,3,6-trisulfonic acid; CP, citrate-phosphate; T_m, denaturation temperature transition midpoint; GSSM, Gene Site Saturation Mutagenesis; PACE, polysaccharide analysis using carbohydrate electrophoresis.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.03333504.

ize successive codons in a gene, systematically generating a library of enzyme variants representing all possible 19 amino acid substitutions at each residue position. It is most powerful when used in combination with GeneReassembly technology to randomly combine single beneficial amino acid substitutions, allowing the identification of protein variants with optimal fitness, in this case, increased thermostability, while preserving optimal catalytic activity.

For this enzyme evolution study, we have chosen to optimize the thermostability of a xylanase (EC 3.2.1.8) because of the many potential commercial applications of a stable enzyme for hydrolysis of plant-derived xylans. Xylanases catalyze the endo-cleavage of β1,4 bonds in xylan polymers, and have utility in improving the effectiveness and economics of processing lignocellulosic materials for the generation of liquid fuels and chemicals. Xylanases can act as substitutes for chlorine chemicals in pulp bleaching, and, when stable and active at high temperature, are useful in food production, biomass conversion, dough making, brewing, and as stable animal feed supplements (Saul et al. 1995; Chen et al. 1997; Kulkarni et al. 1999; Clarke et al. 2000; Beg et al. 2001; Irwin et al. 2003).

Results

Evolution by the GSSM method

The xylanase used for this work (*XYL7746*, internal nomenclature) was discovered from a complex environmental DNA library (Robertson et al. 1996; Short 1997, 1999, 2001a) derived from a sample of fresh bovine manure. *XYL7746* is a 20-kD protein belonging to glycosyl hydrolase family 11 (Coutinho and Henrissat 1999a,b). The mature xylanase shares 83% sequence identity with the family 11 xylanase from *Bacillus circulans* (GenBank accession AAM08360). A number of commercial xylanases belong to family 11 (Kulkarni et al. 1999; Beg et al. 2001), and as such, we were interested in exploring the possibility of improving the thermal properties of a family 11 xylanase.

The GSSM method (Short 2001b) was used to create a comprehensive library of point mutations in gene *XYL7746*. A screen for thermotolerance was developed that measures residual xylanase activity following a temperature challenge above the melting point of the wild-type enzyme. This xylanase thermotolerance screen identified nine single-site amino acid mutants (D8F, Q11H, N12L, G17I, G60H, P64V, S65V, G68A, and S79P; Fig. 1A) that had improved thermal tolerance relative to the wild-type enzyme (as measured following a heat challenge at 80°C for 20 min; see Materials and Methods). Wild-type enzyme and all nine single-site amino acid mutants were produced in *Escherichia coli* and purified utilizing an N-terminal hexahistidine tag. There was no noticeable difference in activity due to the tag (data not shown).

To determine the effect of the single amino acid mutations on enzymatic activity, all nine mutants were purified and their xylanase activity (initial rates at the wild-type temperature optimum, 70°C) was compared to that of the wild-type enzyme. Enzyme activities were comparable to wild type (initial rate normalized to 1.0) for D8F, N12L, G17I, G60H, P64V, S65V, G68A, and S79P mutants (relative initial rates 0.65, 0.68, 0.76, 1.1, 1.0, 1.2, 0.98, and 0.84, respectively) confirming that these mutations do not significantly alter the enzymatic activity. (Initial rates were measured three or more times, and variance was typically less than 10%.) In contrast to these eight mutants, a notable reduction in enzymatic activity was observed for the best thermal tolerant, single-site mutant, Q11H (relative initial rate 0.35).

Melting temperature (T_m) of wild-type and single-site amino acid mutant enzymes

The purified wild-type xylanase and the nine thermal tolerant single-site amino acid mutants were analyzed using differential scanning calorimetry (DSC). Aggregation was apparent for the wild-type enzyme as evidenced by a shoulder in the DSC trace for its thermal denaturation (Fig. 1B). The evolved mutant enzymes showed no indication of aggregation. For all enzymes, thermally induced denaturation was irreversible and no discernible transition was observed in a second scan of the sample. Due to the irreversibility of denaturation, only the apparent T_m (melting temperature) could be calculated (Sanchez-Ruiz 1992; Beldarrain et al. 2000). The T_m of the wild-type enzyme was 61°C, while the T_ms of all point mutants were increased and ranged from 64°C to 70°C (Fig. 1A). The Q11H mutation introduced the largest increase (T_m = 70°C) over wild-type followed by P64V (69°C), G17I (67°C), and D8F (67°C).

9X combined GSSM mutant enzyme

The 9X enzyme was constructed by combining the single-site changes of the nine thermal tolerant up-mutants by site-directed mutagenesis (Fig. 1A). The 9X mutant was expressed in *E. coli* and purified to homogeneity. DSC was performed to determine the melting temperature. The T_m of 9X enzyme was 35 degrees higher than the wild-type enzyme, demonstrating a dramatic shift in its thermal stability (Fig. 1B).

To evaluate the effect of the combined mutations and elevated melting temperature on the enzyme's biochemical properties, pH and temperature profiles were constructed and compared to the wild-type enzyme. The wild-type and 9X mutant enzyme had comparable pH/activity profiles with the highest activity between pH 5 and 6 (Fig. 2A). Both enzymes had similar initial rate/temperature optima at 70°C; however, the wild-type enzyme had higher activity at lower temperatures (25°C–50°C), whereas the 9X mutant retained

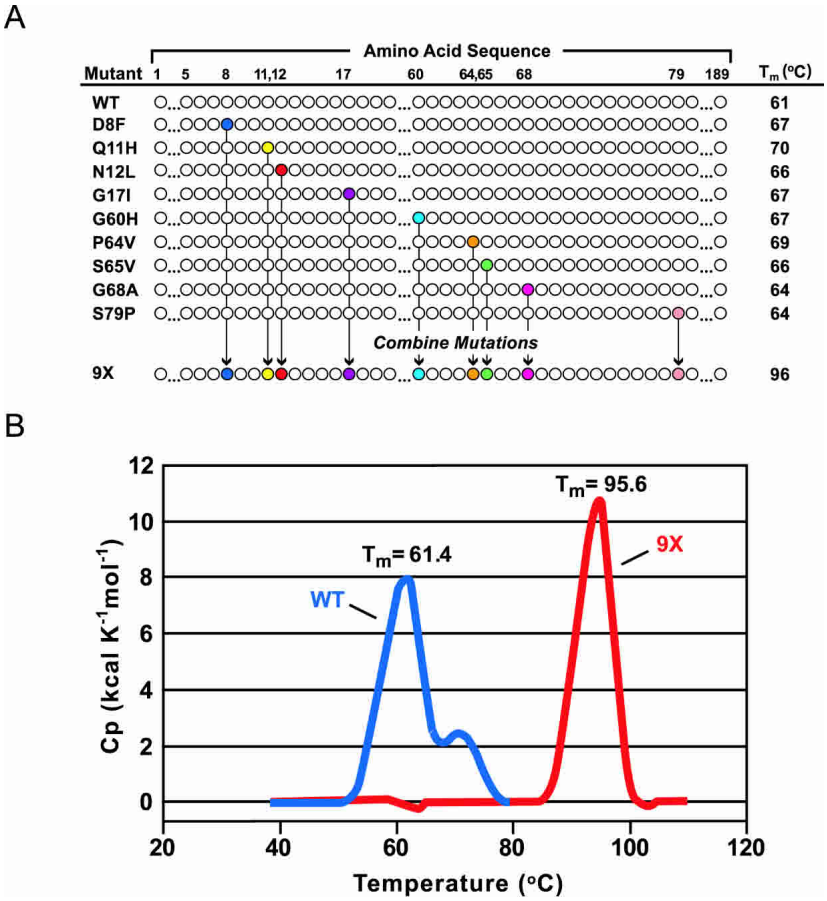

Figure 1. Gene Site Saturation Mutagenesis (GSSM) of wild-type enzyme *XYL7746*. (*A*) The schematic diagram illustrates position, numbering, and the amino acid change for the thermal tolerant point mutants of the wild-type gene. A library of all 64 codons was generated for every amino acid position in the gene (~13,000 mutants) and screened for mutations that increased thermal tolerance. A 9X variant was generated by combining all nine single-site mutants into one enzyme. The corresponding melting temperature transition midpoint (T_m) determined by DSC for each mutant enzyme and the 9X variant is shown on the *right*. (*B*) Unfolding of the wild-type and 9X mutant enzymes was monitored by DSC at a scan rate of 1°C/min. Baseline subtracted DSC data were normalized for protein concentration.

more than 60% of its activity up to 100°C (determined by initial rate) in the presence of substrate (Fig. 2B). The activity of the wild-type enzyme was not detectable at temperatures above 70°C.

To determine the effect of the nine combined mutations on enzyme thermal tolerance, residual activity was measured and compared to the wild-type enzyme. Residual activity was determined by a heat challenge for 5 min at each temperature (37°C, 50°C, 60°C, 70°C, 80°C, and 90°C) followed by activity measurements at 37°C. The wild-type enzyme was completely inactivated above 70°C, while the evolved 9X mutant displayed significant activity after heating at 70°C, 80°C, and even 90°C (Fig. 2C). Furthermore, although the activity of the wild-type enzyme decreased with increasing temperature, the 9X variant was somewhat activated by heating at temperatures up to 60°C.

Generation of combinatorial GSSM variants using GeneReassembly technology

To identify combinatorial variants of the nine single site amino acid mutants with highest thermal tolerance and activity compared to the additively constructed 9X variant, a GeneReassembly library (Short 2003) of all possible mutant combinations (2^9) was constructed and screened. Using thermal tolerance as the screening criterion, 33 unique combinations of the nine mutations were identified including the original 9X variant. A secondary screen was performed to select for variants with higher activity/expression than the evolved 9X. This screen yielded 10 variants with sequences possessing between 6 and 8 of the original single mutations in various combinations (Fig. 3A).

The melting temperature (T_m) of each of the combinatorial variants was at least 28°C higher than the wild type

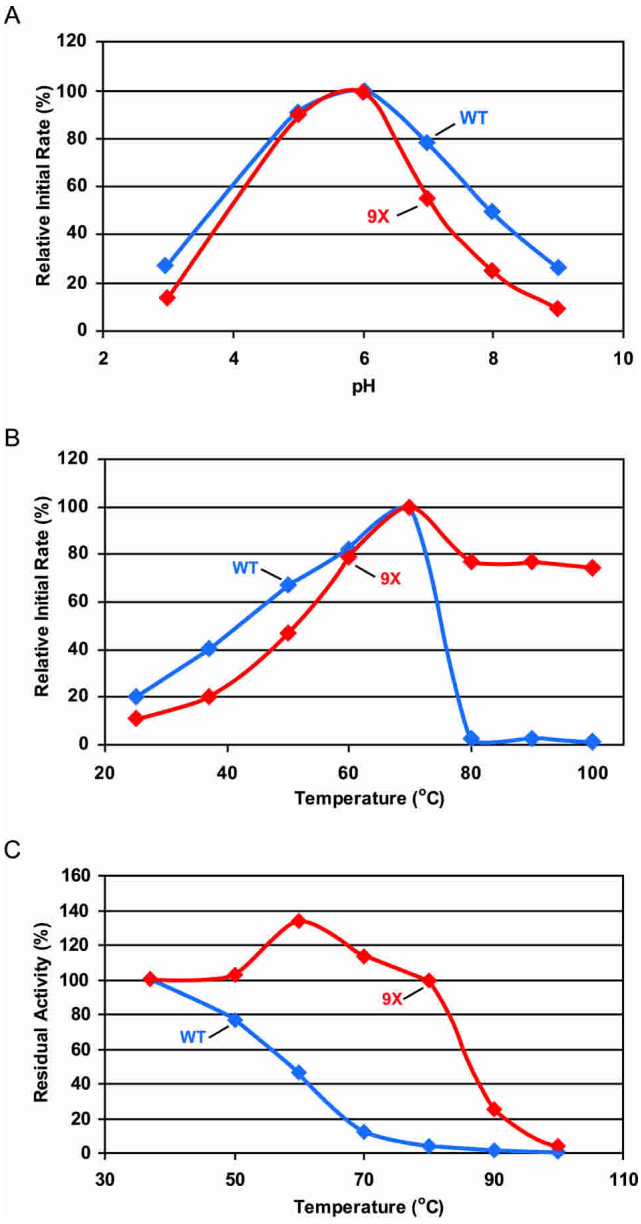

Figure 2. Biochemical characterization of wild-type and evolved 9X mutant enzymes. (*A*) pH dependence of activity for the wild-type and evolved 9X mutant enzymes. Xylanase activity was measured at 37°C at each pH and the initial velocity was plotted against absorbance at 590 nm to determine initial rates. (*B*) Temperature dependence of activity for the wild-type and evolved 9X mutant enzymes. The optimum temperatures of the wild-type and 9X mutant enzymes were measured over a temperature range of 25°C–100°C at pH 6.0, and are based on initial rates measured over 5 min. (*C*) Thermal stability of wild-type and evolved 9X mutant enzymes. Thermal dependence of activity of the wild-type and evolved 9X mutant enzymes was measured by first heating the enzymes at each of the indicated temperatures for 5 min followed by cooling to room temperature and the measurement of residual activity (initial rate at 37°C [pH 6.0]). For all experiments initial rates were measured two or more times, and the variation was less than 10%. For comparison, the data are shown normalized at the wild-type maximum for each condition.

(Fig. 3A) and all of the reassembly variants displayed higher relative activity than the 9X enzyme (data not shown). The activity of one variant in particular, 6X-2, was greater than the wild-type enzyme and significantly better (1.7-fold) than the 9X enzyme (Fig. 3B). Sequence comparison of the reassembly variants identified at least six mutations that were required for the enhanced thermostability (>20 degrees). All 33 unique variants found in the initial thermostability screen contained both Q11H and G17I mutations, demonstrating their importance for thermal tolerance.

Analysis of wild-type and variant polysaccharide product fingerprints

The products generated by the wild-type, 6X-2, and 9X variants were compared by polysaccharide analysis using carbohydrate gel electrophoresis (PACE). Different substrates (oligosaccharides and polysaccharides) were tested for hydrolysis by the xylanases. The digestion products of the three xylanases tested were very similar (Fig. 4). All three enzymes hydrolyzed $(Xyl)_6$ and $(Xyl)_5$, mainly into both $(Xyl)_3$ and $(Xyl)_2$, and $(Xyl)_4$ was hydrolyzed to $(Xyl)_2$ (Fig. 4A). Only a small amount of hydrolysis of $(Xyl)_3$ into $(Xyl)_2$ and Xyl was observed, indicating that $(Xyl)_3$ is a relatively poor substrate for the enzyme. No activity was detected on $(Xyl)_2$ (data not shown). Beechwood xylan, which contains glucuronosyl residues, was hydrolyzed by all three enzymes mainly into $(Xyl)_2$ and $(Xyl)_3$, but other bands were detected that migrated between oligoxylan bands (Fig. 4B). In PACE analysis, each oligosaccharide has a specific migration, depending on the sugar composition and degree of polymerization (Goubet et al. 2002); thus, these bands likely correspond to oligoglucuronoxylans. Therefore, the evolved enzymes retained the substrate specificity of the wild-type enzyme.

Discussion

A combination of laboratory gene evolution strategies was used to rapidly generate a highly active and thermostable xylanase optimized for process compatibility in a number of industrial market applications. GSSM methodology was employed to scan the entire sequence of a xylanase (*XYL7746*) and to identify nine point mutations that improve its thermal tolerance.

Although it had no discernable effect on the hydrolysis product profile of the enzyme (Fig. 4), the addition of the nine mutations to the protein sequence resulted in a moderate reduction in enzymatic specific activity at the wild-type temperature optimum (70°C; Fig. 3B). Using the GeneReassembly method to generate a combinatorial library of the nine single-site amino acid mutants, this reduction in activity was overcome. Ten thermostable variants (T_ms between 89°C and 94°C) with activity better than the 9X variant were obtained from screening the GeneReassembly

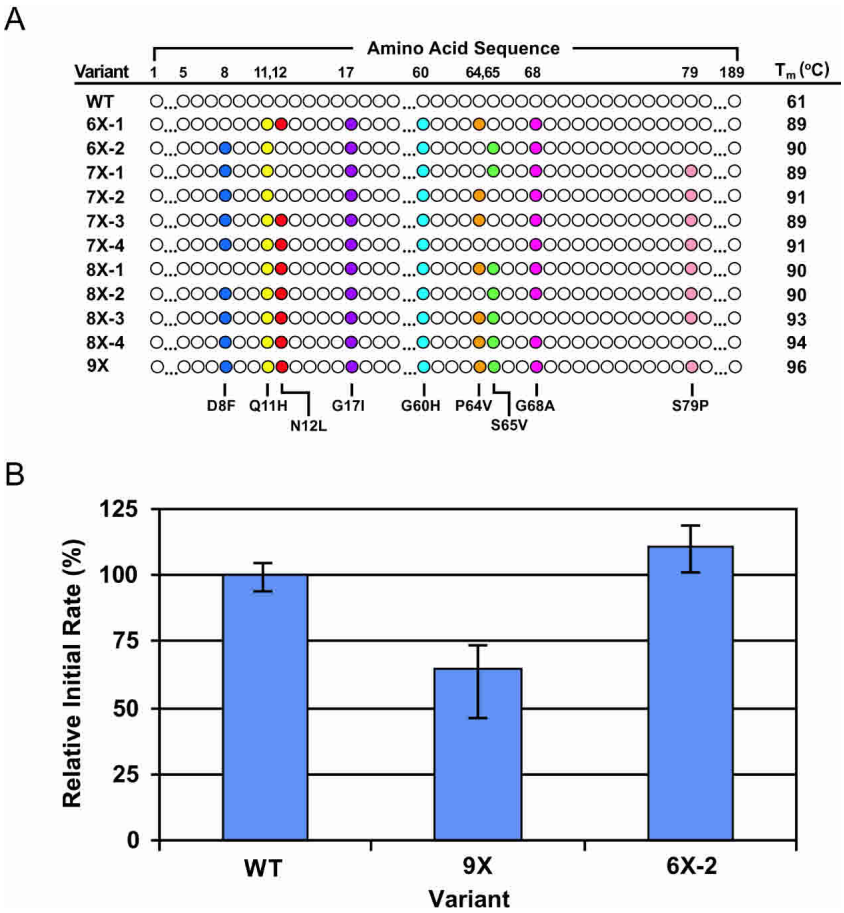

Figure 3. Combinatorial variants identified using GeneReassembly technology. (*A*) A GeneReassembly library of all possible combinations of the nine GSSM point mutations was constructed and screened for variants with improved thermal tolerance and activity. Eleven variants including the 9X variant were obtained. As shown in the figure, the variants possessed six, seven, eight, or nine of the point mutations in various combinations. The corresponding melting temperature transition midpoint (T_m) determined by DSC of each variant is shown on the *right*. (*B*) Relative activity (initial rate measured over a 5-min time period) of the 6X-2 and 9X variants compared to wild-type at the temperature optimum (70°C) and pH 6.0. Error bars show the range in the initial rate for three measurements.

library. With a T_m of 90°C, enzymatic specific activity surpassing the wild type and a product fingerprint unaltered and comparable to the wild type, the 6X-2 variant is particularly notable. To our knowledge the shift in T_m obtained for these variants is the highest increase reported from the application of directed evolution technologies or by rational design.

To gauge the effectiveness of combinatorial mixing versus addition of the point mutants to the desired phenotype, a fitness parameter combining contributions both from changes in enzyme activity and thermal stability was calculated for each mutant. The term fitness as described here is not an objective measure that can be compared to other enzymes, but rather a term that allows the measurement of the success of directed evolution of this particular xylanase. Because enzyme fitness, F, is calculated by equally weighting changes in T_m and enzyme activity (equation 1) for this set of variants, the maximum allowable fitness value is 2 ($F_T \leq 1$ and $F_V \leq 1$, see equations 2 and 3 in Materials and Methods). In other words, if the variant with the best activity also had the highest T_m, its fitness value would be 2. Considering only thermal stability, it is clear that Q11H is the best single site mutation and the 9X variant is the most thermostable combinatorial variant (Fig. 5A). However, with a fitness value near 2 (Fig. 5B), the 6X-2 variant is the closest to possessing the best possible combination of thermal stability and enzyme activity. The single site mutation that confers the highest value of fitness is S65V. Although the T_m of the S65V mutant is lower than that of the Q11H mutant (66°C versus 70°C, respectively), it has a higher fitness value because its specific activity is not reduced relative to the wild type.

The GeneReassembly method also allowed the identification of key residues for improved thermal stability. Two

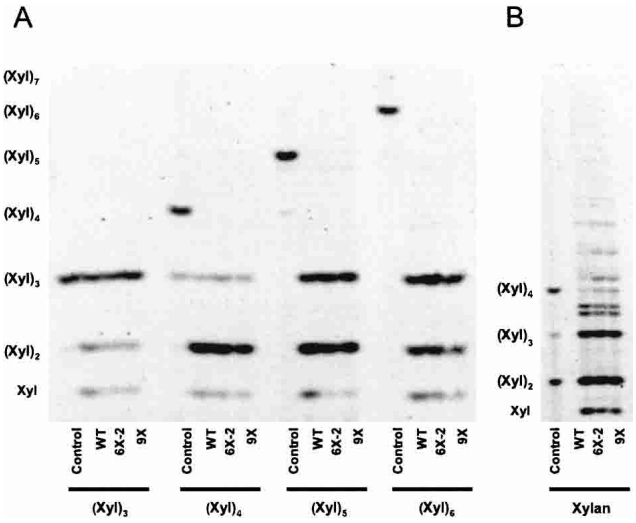

Figure 4. Product fingerprints of wild-type, 6X-2, and 9X variants determined by PACE. (*A*) Fingerprints were obtained after hydrolysis of oligoxylans $(Xyl)_3$, $(Xyl)_4$, $(Xyl)_5$, and $(Xyl)_6$ by wild-type and variant enzymes. Control lanes contain oligosaccharide incubated under the assay conditions in the absence of enzyme. (*B*) Fingerprints obtained after hydrolysis of Beechwood xylan by wild-type and variant enzymes. Standards contained $(Xyl)_2$, $(Xyl)_3$, and $(Xyl)_4$. All assays were performed at 37°C and pH 5.5.

mutations, Q11H and G17I, were present in every GeneReassembly variant identified based on thermal tolerance (Fig. 3A). The structural determinants for thermal stability of proteins have been studied and several theories have been documented (Britton et al. 1995; Chan et al. 1995; Yip et al. 1995; Tanner et al. 1996; Ladenstein and Antranikian 1998; Vetriani et al. 1998; Britton et al. 1999; Kinjo and Nishikawa 2001). Hydrogen bonding patterns, ionic interactions, decreased length of surface loops, and hydrophobic packing are among the key factors contributing to protein thermal stability (Pace et al. 1996; Adams and Kelly 1998; Kumar and Nussinov 2001; Voigt et al. 2001). Even though the contribution of each type of interaction to protein thermal stability is not fully understood (Adams and Kelly 1998; Kumar et al. 2000), many studies have suggested that hydrophobic interactions are perhaps the most significant (Anderson et al. 1993; Britton et al. 1995; Spassov et al. 1995; Witarto and Sode 2001; Machius et al. 2003). Other studies, however, argue that electrostatic interactions and favorable hydrogen bonding patterns contribute as much or more to stability than do hydrophobic interactions (Voorhorst et al. 1997; Wallon et al. 1997; Gruber et al. 1998; Pace 2001; Pace et al. 2001). Our findings corroborate the premise that hydrophobic interactions are key, as all but two of the stabilizing mutations found in this study are replacements of relatively polar, charged or small (glycine) residues for much larger nonpolar residues. It must be noted, however, that the single site mutation that introduces the largest increase in T_m is Q11H. The aromatic nature of the histidine residue could enable hydrophobic contact, but, in addition to the introduction of a favorable hydrogen bond, the histidine residue could also introduce an electrostatic interaction if protonated.

The structure of the *B. circulans* xylanase (Wakarchuk et al. 1994) (PDB ID 1BCX) consists of two β-sheets, where one sheet forms the protein surface and the other lies beneath and forms the active site. A β-strand at the immediate N terminus of the protein lies on the edge of the active site β-sheet. As such, unwinding the structure beginning from the N terminus would initiate the disassembly of the active site. Interestingly, the first four mutations in the sequence that were discovered in this study (D8F, Q11H, N12L, and G17I) are clustered together at the N terminus. Assuming that the structure of *XYL7746* and that of the *B. circulans* enzyme are very similar (*XYL7746* shares 83% sequence identity with the *B. circulans* xylanase), these mutations seem to help anchor the N-terminal β-strand to the active-site β-sheet, making the structure more thermally stable. Another set of four mutations also clusters together in the N-terminal region (G60H, P65V, S65V, and G68A). With the exception of the histidine substitution at position 60, the mutations all introduce hydrophobic side chains and likely stabilize packing through the hydrophobic effect.

The contribution of the ninth mutation, S79P, to improved thermal stability is not as clear. It is interesting, however, that the *B. circulans* xylanase has a proline at the

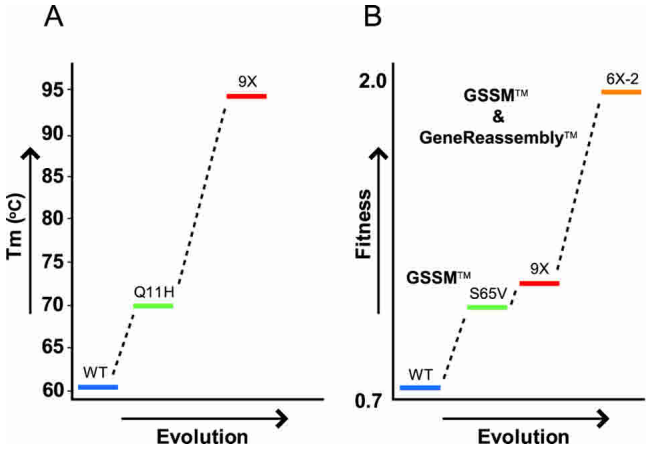

Figure 5. (*A*) Schematic diagram illustrating the level of thermal stability (represented by T_m) improvement over wild-type obtained by GSSM evolution. The single-site amino acid mutant and the combinatorial variant with the highest thermal stability (Q11H and 9X, respectively) are shown in comparison to wild type. (*B*) Fitness diagram illustrating the enzyme improvement obtained by combining GSSM and GeneReassembly technologies. Fitness was determined using the formula $F = F_T + F_V$, where fitness (F) is calculated by equally weighting thermal tolerance fitness (F_T) and relative activity fitness (F_V) as described in Materials and Methods. The point mutation that confers the greatest fitness (S65V) is shown. Combining all nine point mutations also improved fitness (9X); however, the largest improvement in fitness was obtained by combining GSSM and GeneReassembly methods to obtain the best variant (6X-2).

corresponding position (P 76 in the *B. circulans* enzyme). In the primary sequence, this proline is only two residues away from the active site nucleophile, E 78 (Wakarchuk et al. 1994). In the crystal structure, P 76 lies on a turn between two β-strands that form part of the active site, and it may be that the proline at this position helps orient other residues in the β-sheet into conformations that strengthen stabilizing interactions. The T_ms determined by DSC of the *B. circulans* xylanase and xylanase *XYL7746* are very close (~60°C versus 61°C, respectively, at comparable scan rates (Davoodi et al. 1998)). However, in the sequence context of *XYL7746*, a proline at position 79 increases the T_m to 64°C (Fig. 1). It may be that the proline also has a stabilizing effect in the *B. circulans* xylanase, and that a P76S mutation in this enzyme would decrease the T_m of the enzyme.

Even with a good understanding of the optimal interactions to enhance thermal tolerance, the prediction of where to make mutations that introduce such interactions is not straightforward. A nonrational approach using the GSSM method, however, allows rapid sampling of all side chains at all positions within a protein structure. Unlike rational design, or evolution methods that randomize preselected sites, the GSSM method leads to the discovery of amino acid substitutions that introduce functional interactions that could not have been foreseen. The tandem use of GSSM and GeneReassembly technologies is particularly effective, as it capitalizes on the attributes of both methods, enabling the selection of the fittest variant from a combinatorial library of improved enzymes. GeneReassembly technology recombines any number of mutations anywhere in the target gene with the same (high) efficiency in contrast to PCR-based molecular evolution methods where the efficiency drops significantly with decreasing distance between mutations (Volkov and Arnold 2000). In addition, GC-content or secondary structures do not affect the diversity in the resulting chimeric library. The library size can easily be calculated from the number of mutations included. The method is not biased towards any particular crossover, and the resulting pool of gene chimeras contains all possible combinations in equal amounts, reducing screening efforts to a minimum.

The knowledge obtained from directed evolution approaches, combined with structural information, may one day enable the effective prediction of mutations that introduce desirable functional interactions in proteins.

Materials and methods

Genomic DNA libraries and xylanase discovery

The *XYL7746* (internal nomenclature) xylanase gene (GenBank accession number AY394562) was discovered by screening approximately 50,000 plaques from a genomic DNA library (Robertson et al. 1996; Short 1997, 1999, 2001a). Plaques were generated from the λ DNA library on semisolid medium using established procedures (Maniatis et al. 1982), except that top agar containing 0.1%–0.3% Azo-dye-linked birchwood xylan (Azo-xylan, Megazyme) was used. The disappearance of the dye-linked substrate surrounding a plaque was used to identify positive clones. Lambda DNA from positive plaques was recovered, and the sequence of the inserted environmental DNA was determined.

Gene site saturation mutagenesis and activity screening

The *XYL7746* gene was amplified by PCR starting at nucleotide 88 (to remove the 29 amino acid native secretion signal) and cloned into the expression vector pTrcHis2 using the pTrcHis2 TOPO TA Cloning Kit (Invitrogen). The GSSM method was performed as described previously (Gray et al. 2001; Short 2001b) using 64-fold degenerate oligonucleotides to randomize at each codon in the gene so that all possible amino acids would be encoded. The resultant GSSM library was transformed into competent XL1-Blue *E. coli* (Bullock et al. 1987) for screening.

Individual clones were arrayed in 96-well microtiter plates containing 200 μL of LB medium and 100 μg/mL ampicillin using an automated colony picker (AutoGen). Four 96-well plates were screened per codon. The plates were incubated overnight at 37°C. These master plates were replicated, using a 96-well pintool, into fresh medium containing antibiotic. The replica plates were sealed with a gas-permeable adhesive film and incubated overnight at 37°C. After incubation, the seals were removed and the plates centrifuged at approximately 1800*g* for 10 min. The supernatant was removed and the cells resuspended in 45 μL of 100 mM citrate-phosphate buffer (pH 6.0) containing 100 mM KCl (CP buffer). The plates were then covered with an adhesive aluminum seal and incubated at 80°C for 20 min followed by the addition of 30 μL of 2% Azo-xylan prepared in CP buffer and incubation overnight at 37°C. After incubation, 200 μL of 100% ethanol was added to precipitate undigested xylan and the plates were centrifuged at approximately 1800*g* for 10 min. The supernatant was transferred to fresh plates and absorbance measured at 590 nm to quantify residual enzyme activity.

The thermal tolerance of positive clones was verified by reassaying with and without an 80°C heat challenge. The assay was performed as described above, with the following exceptions: cells were resuspended in 45 μL of CP buffer containing 55% BPER lysis reagent (Pierce), 0.23 mg/mL lysozyme and 0.023 mg/mL DNaseI. Duplicate plates were prepared, with one plate incubated at 37°C and the other at 80°C for 20 min each. Following the addition of substrate, the 37°C plate was incubated for 1 h and the 80°C plate overnight at 37°C. Nine verified positive clones were then recovered and sequenced.

Construction and purification of wild-type and mutant enzymes

All nine mutations were combined in one gene using site-directed mutagenesis to generate a 9X mutant enzyme. The *9X* gene, the wild-type gene, and all nine single-site mutant genes were PCR amplified using primers designed to append an N-terminal hexahistidine tag. The PCR products were cloned into pTrcHis2 as described above.

Overnight cultures of *BL21 Star* (Invitrogen) cells carrying the appropriate gene in pTrcHis2 were used to inoculate 1.5 L of Terrific broth (12 g/liter Bactotryptone, 24 g/L yeast extract, 4 ml/L glycerol, 12.5 g/L K_2HPO_4, 2.3 g/L KH_2PO_4) containing 100 μg/mL ampicillin in a six-liter shake flask. The cultures were

incubated at 30°C and shaken at 250 rpm until an absorbance at 600 nm between 0.4 and 0.9 was reached. Cultures were induced with 1 mM isopropyl-1-thio-β-galactopyranoside. Following 18 h of induction at 30°C, the cells were harvested by centrifugation at 10,000*g* for 20 min at 4°C. The pellet was resuspended in lysis buffer (50 mM potassium phosphate [pH 7.0], 100 mM KCl, 7.5 mM β-mercaptoethanol) and cells were lysed by French press (15,000 psi). The cell lysate was centrifuged at 45,000*g* for 30 min to remove cell debris. The clarified supernatant was applied to a Ni-NTA-agarose affinity column and was purified using FPLC exploiting the N-terminal hexahistidine tag. The protein was eluted using a gradient of 5–500 mM imidazole containing 500 mM NaCl and 20 mM sodium phosphate (pH 7.8). Fractions were tested for enzyme activity using Azo-xylan as substrate (see assay below), and SDS-PAGE was used to determine purity. Homogeneous xylanase-containing fractions were combined and dialyzed overnight against 20 mM KPO_4 (pH 7.0) and 100 mM KCl, and were aliquoted and stored at 4°C. Protein concentration was determined spectrophotometrically at 280 nm ($\varepsilon_M = 7.49 \times 10^4\ M^{-1}\cdot cm^{-1}$), and was verified using the Micro BCA Protein Assay reagent kit from Pierce, using bovine serum albumin as a standard.

Xylanase activity assays

Enzymatic activities were determined using 400 μL of 2% Azo-xylan as substrate in 550 μL of CP buffer, pH 6.0 at the indicated temperatures. Activity measurements as a function of pH were determined using 50 mM Britton-Robinson buffer solutions (pH 3.0, 5.0, 6.0, 7.0, 8.0, and 9.0) prepared by mixing solutions of 0.1 M phosphoric acid, 0.1 M boric acid, and 0.1 M acetic acid followed by pH adjustment with 1 M sodium hydroxide. Reactions were initiated by adding 50 μL of 0.1 mg/mL of purified enzyme. Time points were taken from 0 to 15 min, where 50 μL of reaction mixture was added to 200 μL of precipitation solution (100% ethanol). When all time points had been taken, the samples were mixed, incubated for 10 min, and centrifuged at 3000*g* for 10 min at 4°C. Supernatant (150 μL) was aliquoted into a fresh 96-well plate, and absorbance was measured at 590 nm. A_{590} values were plotted against time, and the initial rate was determined from the slope of the line.

Differential scanning calorimetry (DSC)

Calorimetry was performed using a Model 6100 Nano II DSC apparatus (Calorimetry Sciences Corporation) using the DSCRun software package for data acquisition, CpCalc for analysis, CpConvert for conversion into molar heat capacity from microwatts, and CpDeconvolute for deconvolution. Analysis was carried out with 1 mg/mL recombinant protein in 20 mM potassium phosphate (pH 7.0) and 100 mM KCl at a scan rate of 1°C/min. A constant pressure of 5 atm was maintained during all DSC experiments to prevent possible degassing of the solution on heating. The instrumental baseline was recorded routinely before the experiments with both cells filled with buffer. Reversibility of the thermally induced transitions was tested by reheating the solution in the calorimeter cell immediately after cooling the first run.

GeneReassembly library construction and screening

Five segments of the 591bp *XYL7746* gene (gene plus codons for hexahistidine tag) were created according to the locations of the mutations in the GSSM clones and then the segments were blended as described (Short 2003) to form a library of all possible combinations of the GSSM mutants. The final library (512 mutants; segments 1–5) was isolated and cloned into pTrcHis2 and transformed into XL1 Blue MRF′ cells (Jerpseth et al. 1992) and was plated on solid LB medium containing 100 μg/mL ampicillin. Approximately 4000 colonies were autopicked (see above) into approximately 40 96-well plates and were incubated at 37°C overnight. The screening assay was performed as described above for the screening of the GSSM mutant library, except that the resuspended cells were incubated for 60 min at 80°C, followed by addition of substrate, before incubation of the plates at 37°C for 20 min.

Thermal tolerance determination

All enzymes were analyzed for thermal tolerance in 20 mM potassium phosphate (pH 7.0) and 100 mM KCl. The enzymes were heated for 5 min at 37°C, 50°C, 60°C, 70°C, 80°C, and 90°C in thin-walled tubes and were cooled on ice. Residual activities were determined with Azo-xylan as substrate using the assay described above for activity measurement.

Polysaccharide fingerprinting

Polysaccharide fingerprints were determined by polysaccharide analysis using carbohydrate gel electrophoresis (PACE). Beechwood xylan (0.1 mg/mL, 100 μL, Sigma) or xylooligosaccharides (1 mM, 20 μL, Megazyme) were treated with enzyme (1–3 μg) in a total volume of 250 μL for 16 h. The reaction was buffered in 0.1 M ammonium acetate pH 5.5. Controls without substrates or enzymes were performed under the same conditions to identify any nonspecific compounds in the enzymes, polysaccharides/oligosaccharides, or labeling reagents. The reactions were stopped by boiling for 20 min. Assays were independently performed at least two times for each condition. Derivatization using ANTS (8-aminonaphthalene-1,3,6-trisulfonic acid; Molecular Probes), electrophoresis, and imaging were carried out as described previously (Goubet et al. 2002).

Fitness calculation

The fitness (F^n), for a given enzyme variant, *n*, was calculated by equally weighting increase in denaturation temperature transition midpoint (T_m) and increase (or decrease) in enzymatic activity relative to the largest difference in each parameter across all variants:

$$F^n = F_T{}^n + F_V{}^n \quad (1)$$

where $F_T{}^n = T_m$ fitness factor of the variant and $F_V{}^n =$ activity fitness factor of the variant. The fitness factors for each (T_m and activity) are relative to the largest difference in T_m, or rate, across all of the variants:

$$F_T{}^n = (T_m{}^n - T_m{}^L)/(T_m{}^H - T_m{}^L) \quad (2)$$

$$F_V{}^n = (V^n - V^L)/(V^H - V^L) \quad (3)$$

where $T_m{}^n$ is the T_m for the given variant, *n*, and $T_m{}^L$ is the lowest T_m across all variants and $T_m{}^H$ the highest T_m across all variants; and where V^n is the relative rate for the given variant, *n*, and V^L is

the lowest rate across all variants and V^H the highest rate across all variants.

Acknowledgments

We thank the Diversa Corp. Sequencing Group as well as Jeff Gonaver, Eileen O'Donoghue, Kelly Mercer, Ryan Mueller, Arne Solbak, Ryan Williams, and Heidi Short. We want to express special thanks to Michael Kulwiec for image design. The PACE studies were supported by grants from the Biotechnology and Biological Sciences Research Council (UK).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Adams, M.W. 1993. Enzymes and proteins from organisms that grow near and above 100°C. *Annu. Rev. Microbiol.* **47:** 627–658.

Adams, M.W. and Kelly, R.M. 1998. Finding and using hyperthermophilic enzymes. *Trends Biotechnol.* **16:** 329–332.

Anderson, D.E., Hurley, J.H., Nicholson, H., Baase, W.A., and Matthews, B.W. 1993. Hydrophobic core repacking and aromatic–aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117→Phe. *Protein Sci.* **2:** 1285–1290.

Beg, Q.K., Kapoor, M., Mahajan, L., and Hoondal, G.S. 2001. Microbial xylanases and their industrial applications: A review. *Appl. Microbiol. Biotechnol.* **56:** 326–338.

Beldarrain, A., Acosta, N., Montesinos, R., Mata, M., and Cremata, J. 2000. Characterization of *Mucor pusillus* rennin expressed in *Pichia pastoris*: Enzymic, spectroscopic and calorimetric studies. *Biotechnol. Appl. Biochem.* **31:** 77–84.

Britton, K.L., Baker, P.J., Borges, K.M., Engel, P.C., Pasquo, A., Rice, D.W., Robb, F.T., Scandurra, R., Stillman, T.J., and Yip, K.S. 1995. Insights into thermal stability from a comparison of the glutamate dehydrogenases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Eur. J. Biochem.* **229:** 688–695.

Britton, K.L., Yip, K.S., Sedelnikova, S.E., Stillman, T.J., Adams, M.W., Ma, K., Maeder, D.L., Robb, F.T., Tolliday, N., Vetriani, C., et al. 1999. Structure determination of the glutamate dehydrogenase from the hyperthermophile *Thermococcus litoralis* and its comparison with that from *Pyrococcus furiosus*. *J. Mol. Biol.* **293:** 1121–1132.

Bullock, W.O., Fernandez, J.M., and Short, J.M. 1987. XLI-Blue: A high efficiency plasmid transforming *RecA Escherichia coli* strain with β-galactosidase selection. *Biotechniques* **5:** 376–378.

Chan, M.K., Mukund, S., Kletzin, A., Adams, M.W., and Rees, D.C. 1995. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267:** 1463–1469.

Chen, C.C., Adolphson, R., Dean, J.F.D., Eriksson, K.E.L., Adams, M.W.W., and Westpheling, J. 1997. Release of lignin from kraft pulp by a hyperthermophilic xylanase from *Thermotoga maritima*. *Enzyme Microb. Technol.* **20:** 39–45.

Clarke, J.H., Davidson, K., Rixon, J.E., Halstead, J.R., Fransen, M.P., Gilbert, H.J., and Hazlewood, G.P. 2000. A comparison of enzyme-aided bleaching of softwood paper pulp using combinations of xylanase, mannanase and α-galactosidase. *Appl. Microbiol. Biotechnol.* **53:** 661–667.

Coutinho, P.M. and Henrissat, B. 1999a. Carbohydrate-active enzymes: An integrated database approach. In *Recent advances in carbohydrate bioengineering* (eds. H.J. Gilbert et al.), pp. 3–12. The Royal Society of Chemistry, Cambridge, UK.

———. 1999b. The modular structure of cellulases and other carbohydrate-active enzymes: An integrated database approach. In *Genetics, biochemistry and ecology of cellulose degradation* (eds. K. Ohmiya et al.), pp. 15–23. Uni Publishers Co., Tokyo.

Davoodi, J., Wakarchuk, W.W., Surewicz, W.K., and Carey, P.R. 1998. Scan-rate dependence in protein calorimetry: The reversible transitions of *Bacillus circulans* xylanase and a disulfide-bridge mutant. *Protein Sci.* **7:** 1538–1544.

Farinas, E.T., Bulter, T., and Arnold, F.H. 2001. Directed enzyme evolution. *Curr. Opin. Biotechnol.* **12:** 545–551.

Goubet, F., Jackson, P., Deery, M.J., and Dupree, P. 2002. Polysaccharide analysis using carbohydrate gel electrophoresis: A method to study plant cell wall polysaccharides and polysaccharide hydrolases. *Anal. Biochem.* **300:** 53–68.

Gray, K.A., Richardson, T.H., Kretz, K., Short, J.M., Bartnek, F., Knowles, R., Kan, L., Swanson, P.E., and Robertson, D.E. 2001. Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.* **343:** 607–617.

Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., and Kratky, C. 1998. Thermophilic xylanase from *Thermomyces lanuginosus*: High-resolution X-ray structure and modeling studies. *Biochemistry* **37:** 13475–13485.

Irwin, D., Leathers, T.D., Greene, R.V., and Wilson, D.B. 2003. Corn fiber hydrolysis by *Thermobifida fusca* extracellular enzymes. *Appl. Microbiol. Biotechnol.* **61:** 352–358.

Jerpseth, B., Greener, A., Short, J.M., Viola, J., and Kretz, P.L. 1992. XL1-Blue MRF′ *E. coli* cells: McrA$^-$, McrCB$^-$, McrF$^-$, Mrr$^-$, HsdR$^-$ derivative of XL1-Blue cells. *Strategies* **5:** 81–83.

Kinjo, A.R. and Nishikawa, K. 2001. Comparison of energy components of proteins from thermophilic and mesophilic organisms. *Eur. Biophys. J.* **30:** 378–384.

Kulkarni, N., Shendye, A., and Rao, M. 1999. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol. Rev.* **23:** 411–456.

Kumar, S. and Nussinov, R. 2001. How do thermophilic proteins deal with heat? *Cell Mol. Life Sci.* **58:** 1216–1233.

Kumar, S., Tsai, C.J., and Nussinov, R. 2000. Factors enhancing protein thermostability. *Protein Eng.* **13:** 179–191.

Ladenstein, R. and Antranikian, G. 1998. Proteins from hyperthermophiles: Stability and enzymatic catalysis close to the boiling point of water. *Adv. Biochem. Eng. Biotechnol.* **61:** 37–85.

Machius, M., Declerck, N., Huber, R., and Wiegand, G. 2003. Kinetic stabilization of *Bacillus licheniformis* α-amylase through introduction of hydrophobic residues at the surface. *J. Biol. Chem.* **278:** 11546–11553.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Pace, C.N. 2001. Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* **40:** 310–313.

Pace, C.N., Shirley, B.A., McNutt, M., and Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. *FASEB J.* **10:** 75–83.

Pace, C.N., Horn, G., Hebert, E.J., Bechert, J., Shaw, K., Urbanikova, L., Scholtz, J.M., and Sevcik, J. 2001. Tyrosine hydrogen bonds make a large contribution to protein stability. *J. Mol. Biol.* **312:** 393–404.

Robertson, D.E., Mathur, E.J., Swanson, R.V., Marrs, B.L., and Short, J.M. 1996. The discovery of new biocatalysts from microbial diversity. *SIM News* **46:** 3–8.

Sanchez-Ruiz, J.M. 1992. Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophys. J.* **61:** 921–935.

Saul, D.J., Williams, L.C., Reeves, R.A., Gibbs, M.D., and Bergquist, P.L. 1995. Sequence and expression of a xylanase gene from the hyperthermophile Thermotoga sp. strain FjSS3-B.1 and characterization of the recombinant enzyme and its activity on kraft pulp. *Appl. Environ. Microbiol.* **61:** 4110–4113.

Short, J.M. 1997. Recombinant approaches for accessing biodiversity. *Nat. Biotechnol.* **15:** 1322–1323.

———. 1999. Protein activity screening of clones having DNA from uncultured microorganisms. U.S. Patent No. 5,958,672.

———. 2001a. Gene expression library produced from DNA from uncultivated microorganisms and methods for making the same. U.S. Patent No. 6,280,926.

——— 2001b. Saturation mutagenesis in directed evolution. U.S. Patent No. 6,171,820.

———. 2003. Synthetic ligation reassembly in directed evolution. U.S. Patent No. 6,537,776.

Spassov, V.Z., Karshikoff, A.D., and Ladenstein, R. 1995. The optimization of protein–solvent interactions: Thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci.* **4:** 1516–1527.

Tanner, J.J., Hecht, R.M., and Krause, K.L. 1996. Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å resolution. *Biochemistry* **35:** 2597–2609.

van den Burg, B., Vriend, G., Veltman, O.R., Venema, G., and Eijsink, V.G. 1998. Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci.* **95:** 2056–2060.

Vetriani, C., Maeder, D.L., Tolliday, N., Yip, K.S., Stillman, T.J., Britton, K.L., Rice, D.W., Klump, H.H., and Robb, F.T. 1998. Protein thermostability

above 100°C: A key role for ionic interactions. *Proc. Natl. Acad. Sci.***95:** 12300–12305.

Voigt, C.A., Mayo, S.L., Arnold, F.H., and Wang, Z.G. 2001. Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci.* **98:** 3778–3783.

Volkov, A.A. and Arnold, F.H. 2000. Methods for in vitro DNA recombination and random chimeragenesis. *Methods Enzymol.* **328:** 447–456.

Voorhorst, W.G., Warner, A., de Vos, W.M., and Siezen, R.J. 1997. Homology modelling of two subtilisin-like proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. *Protein Eng.* **10:** 905–914.

Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J., and Yaguchi, M. 1994. Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. *Protein Sci.* **3:** 467–475.

Wallon, G., Kryger, G., Lovett, S.T., Oshima, T., Ringe, D., and Petsko, G.A. 1997. Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*. *J. Mol. Biol.* **266:** 1016–1031.

Witarto, A.B. and Sode, K. 2001. Increasing the hydrophobic interaction between terminal W-motifs enhances the stability of *Salmonella typhimurium* sialidase. A general strategy for the stabilization of β-propeller protein fold. *Protein Eng.* **14:** 891–896.

Yip, K.S., Stillman, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiaraluce, R., and Consalvi, V. 1995. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3:** 1147–1158.