

Selection of Horseradish Peroxidase Variants with Enhanced Enantioselectivity by Yeast Surface Display

Daša Lipovšek,^{1,5,6} Eugene Antipov,^{1,5} Kathryn A. Armstrong,¹ Mark J. Olsen,¹ Alexander M. Klibanov,^{1,2} Bruce Tidor,^{1,3,*} and K. Dane Wittrup^{1,4,*}

¹Biological Engineering Division

²Department of Chemistry

³Department of Electrical Engineering and Computer Science

⁴Department of Chemical Engineering

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁵These authors contributed equally to this work.

⁶Present address: Codon Devices, Inc., Cambridge, MA 02139, USA.

*Correspondence: tidor@mit.edu (B.T.), wittrup@mit.edu (K.D.W.)

DOI 10.1016/j.chembiol.2007.09.008

SUMMARY

We report a method for in vitro selection of catalytically active enzymes from large libraries of variants displayed on the surface of the yeast *S. cerevisiae*. Two libraries, each containing $\sim 2 \times 10^6$ variants of horseradish peroxidase (HRP), were constructed; one involved error-prone PCR that sampled mutations throughout the coding sequence, whereas the other involved complete combinatorial enumeration of five positions near the active site to non-cysteine residues. The enzyme variants displayed on the yeast surface were allowed to modify it with a fluorescently labeled substrate. A combination of positive and negative selection applied to the active-site-directed library resulted in variants with up to an 8-fold altered enantioselectivity, including its reversal, toward L/D-tyrosinol. In contrast, the library constructed by using error-prone PCR yielded no HRP variants with a significantly improved enantioselectivity.

INTRODUCTION

Enzymes are attractive catalysts for applications in organic chemistry, primarily due to their exquisite stereospecificity, and especially the ability to recognize and produce particular enantiomers of chiral molecules [1]. Because only a small fraction of reactions of interest to synthetic chemists are catalyzed by naturally evolved enzymes, recent years have seen a major effort to create enzymes with altered activity, usually by reengineering existing enzymes. Directed evolution has proven to be a particularly powerful approach to engineering enzymes [2, 3], as well as other proteins [4], with improved properties, because it does not require a detailed knowledge of protein structure and function. Instead, large libraries of proteins with

varying sequence, structure, and function are created, followed by screening or selecting for variants with desired properties.

The major challenge in the directed evolution of enzymes is the creation of a stable linkage between genotype (the DNA encoding a particular enzyme variant) and phenotype (enzymatic activity). The most direct approach, which separates different enzyme variants in wells of microtiter plates, can be applied to almost any enzyme, but it limits the throughput of screening to 10^3 – 10^4 variants per library [5]. The alternative, in vitro selection by using display technologies, can process libraries of 10^6 – 10^{10} variants, but it limits the types of enzymes and chemical reactions that can be explored [6]. An indirect approach to in vitro selection for enzymatic activity uses binding to transition-state analogs as an indication of potential activity [7]. Selection methods that test the performance of variants in actual enzymatic reactions generally require that the reaction product be trapped on the surface of a phage or bacterial cell or inside bacterial cells. Alternatively, water-in-oil emulsion droplets have been used to colocalize reaction products with the bacteria that produce each enzyme variant [8]. To date, all selection methods applied to enzymes that modify small molecules rely on bacterial or in vitro expression of enzyme variants, thus precluding directed evolution of numerous eukaryotic enzymes with extensive disulfide bonding or posttranslational modification. Whereas the advantages of a selection system based on a eukaryotic organism for the evolution of such enzymes are clear, the only example of selection for enzymatic activity in a eukaryote described so far involves a negative selection of homing endonucleases unable to cleave their DNA targets in yeast [9]. We set out to extend this work by designing a yeast-based selection system that can be applied to enzymes not involved in the biology of the yeast cell.

We report a new, to our knowledge, method for in vitro selection of enzymatic activity from large libraries of variants displayed on the surface of the yeast *S. cerevisiae* and separated by fluorescence-activated cell sorting (FACS). As previously demonstrated for antibody fragments

[10, 11], extracellular receptor domains [12, 13], and a eukaryotic lipase in a high-throughput screen [14], yeast surface display is well suited to eukaryotic proteins. We used the model enzyme horseradish peroxidase (HRP), which contains four disulfide bonds, as well as a heme prosthetic group, and cannot be expressed in a soluble form in bacteria [15].

Due to the possibility of their applications to chemical synthesis, HRP and other peroxidases have been subjected to directed evolution by using random or directed mutagenesis, DNA shuffling, and high-throughput screening to identify mutants with higher stability [16, 17] or altered specificity [18]. In addition, a bacterial catalase has been mutated and screened [19], and a catalytic antibody raised against a transition-state analog has been mutated and selected by phage display [20] for increased peroxidase activity.

We focused on the enantioselectivity of HRP during its catalysis of radical dimerization of two chiral phenolic substrates—tyrosinol supplied in solution and tyrosine found naturally on the yeast surface. Wild-type HRP shows a marginal preference for L-tyrosinol over D-tyrosinol. Using two separate selection strategies, we both augmented and reversed HRP enantioselectivity.

Enzymes with altered enantioselectivity have been engineered previously [21, 22] by screening libraries randomized by error-prone PCR [23], mutagenesis of specific active-site residues [24–27], or a combination of randomization methods [28–30] and DNA shuffling [31–33]. Selected enantioselective variants contained between 1 and 11 mutations per gene; there were critical mutations found in, or near, the active site, as well as at a distance from it [34].

To identify the most efficient randomization strategy to manipulate enantioselectivity of HRP, we started our selections with two different HRP-derived libraries. The first one was constructed by using error-prone PCR and introduced a range of mutations throughout the HRP gene. The second library focused on 5 residues at, or close to, the HRP active site and was designed to sample all possible sequence permutations at those five positions. We found that in selections for selectivity for D- and for L-tyrosinol, only the active-site-directed library yielded variants with a significantly enhanced enantioselectivity. While it is tempting to conclude that the failure of the error-prone PCR-generated library to yield variants with improved enantioselectivity is due to the requirement for simultaneous mutation of multiple residues close to the active site, which would be a rare event in the error-prone PCR library, other possibilities exist; for instance, uneven or insufficient coverage of the error-prone PCR library, or selectivity and detection issues in the screen could also account for the current results.

RESULTS AND DISCUSSION

Yeast Surface Display of Wild-Type Horseradish Peroxidase

To demonstrate that HRP displayed on the yeast surface retains enzymatic activity, the synthetic gene encoding

the wild-type enzyme (Figure 1) was cloned into a pCT-derived yeast surface display vector, pCT2, downstream from the gene for Aga2 protein and upstream from a c-myc tag (Figure 2A) and transformed into *S. cerevisiae*. The presence of HRP on the yeast surface was established by fluorescently labeling the cells with antibody against C-myc, by using Alexa 633 dye. The enzymatic activity of surface-displayed HRP was confirmed by incubating the yeast cells with two substrates, hydrogen peroxide and tyrosinol labeled with a second fluorescent dye, Alexa 488 (Figure 2B). The resulting labeled yeast cells were characterized by using analytical flow cytometry (Figure 2C). The presence of double-labeled cells in the yeast transformed with wild-type HRP, but not in the negative-control yeast transformed with the same plasmid missing the HRP gene, demonstrates that the yeast displaying HRP can incorporate the substrate, presumably by the reaction between enzymatically produced, fluorescently labeled tyrosinol-free radicals and tyrosines on the surface of membrane-associated yeast proteins.

This approach to detecting HRP activity, first proposed conceptually as “enzyme screening by covalent attachment of products via enzyme display” by Becker and Kolmar [6], is similar to the method used by Yin et al. [20], who detected HRP-catalyzed modification of a phage-displayed antibody fragment with biotinylated tyramine. Whereas both selections rely on the incorporation of a phenolic substrate into protein associated with a display particle, by using a similar chemical reaction, the two approaches differ in two significant ways. First, our use of a eukaryotic display organism allowed us to study HRP, which cannot be expressed in an active form in bacteria used to express protein in phage display. Second, our use of tyrosinol, a chiral substrate, allows us to study enantioselectivity of HRP.

As shown in Figure 2C and Table 1, Alexa 488-L-tyrosinol appears to be incorporated into yeast only slightly more efficiently than Alexa 488-D-tyrosinol, with the enantioselectivity $E_{L/D}$ of 1.2.

The fact that the subpopulation of yeast transformed with wild-type HRP with a low Alexa 633 signal, and thus presumably a low level of HRP expression [35, 36], is still labeled with Alexa 488-tyrosinol (Figure 2C) suggests *trans*-labeling, i.e., that HRP displayed on a yeast cell attaches Alexa 488-labeled substrate onto a different yeast cell. A significant amount of *trans*-labeling would disturb the linkage between genotype and phenotype and thus preclude selection for HRP enantioselectivity. We quantified the amount of *trans*-labeling by labeling a mixture of a yeast strain displaying HRP and a C-terminal C-myc tag and a yeast strain displaying bovine trypsin inhibitor I (BPTI) and a C-terminal Flag tag. Most of the cells (62%) expressing the C-myc tag (and HRP), but only 7% of the cells expressing the Flag tag (and BPTI), were labeled with Alexa 488. In addition, as shown in Figure 2C, yeast transformed with wild-type HRP with a high Alexa 633 signal, and thus a high level of HRP expression, incorporates more Alexa 488-tyrosinol than the yeast with low-level expression of HRP; the amount of Alexa 488 incorporated is

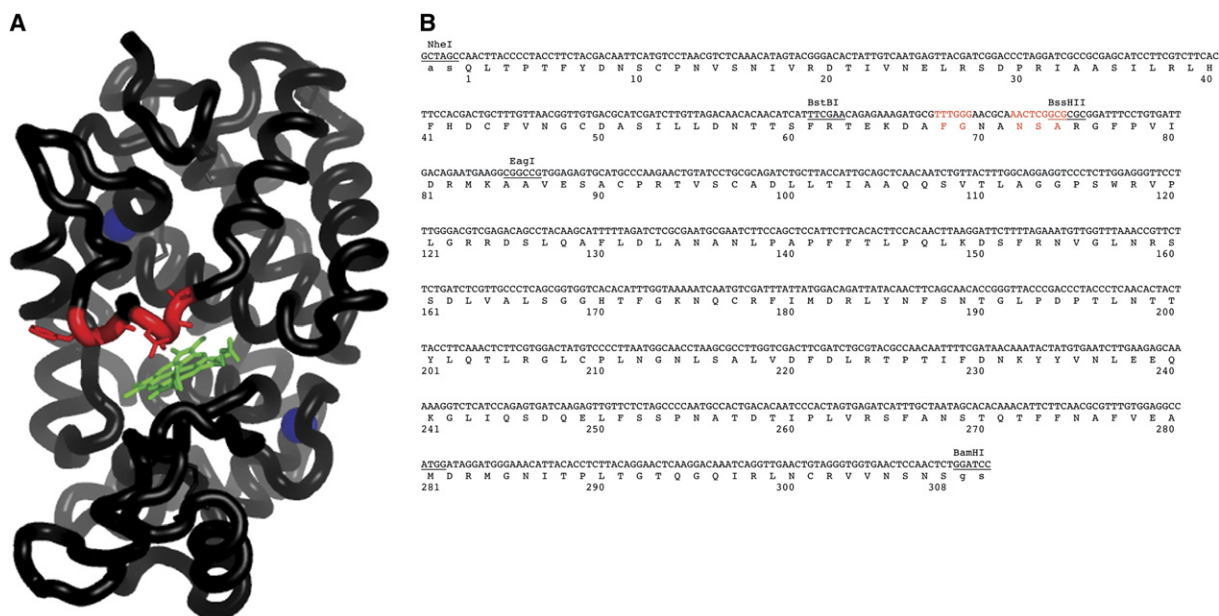


Figure 1. Horseradish Peroxidase

(A) Crystal structure (PDB ID: 1HCH). The HRP main chain and the side chains of disulfide-bonded cysteines are shown in black. Residues 68, 69, 72, 73, and 74, which were randomized to make library HRP-C, are shown in red; the heme prosthetic group is shown in green; and the two calcium ions are shown in blue.

(B) Wild-type protein sequence and DNA sequence used to construct HRP libraries. Unique restriction-enzyme recognition sites used in HRP cloning and library construction are underlined. The wild-type protein sequence is shown in capital letters. Residues 68, 69, 72, 73, and 74, which were randomized to make library HRP-C, are shown in red.

roughly proportional to the level of expression. The combination of a low number of *trans*-labeled cells and the high efficiency of *cis*-labeling by HRP-expressing cells provides a high enough *cis*-to-*trans* (i.e., signal-to-noise) ratio to select new HRP variants based on enzymatic activity by using yeast surface display.

Construction of HRP-Based Libraries

We used two different HRP-based libraries to compare the effectiveness of two common approaches to generating sequence variation in libraries for in vitro evolution—random versus active-site-directed mutagenesis.

The randomly mutagenized library, HRP-E, was generated by error-prone PCR amplification of the entire HRP gene. Mutations could occur anywhere in the gene; between 0 and 17 DNA mutations per clone were observed in the sequences of 24 randomly chosen clones from the unscreened library, with a median of 3. The perceived advantage of this approach is that it samples all possible types of mutations, namely, (i) single, double, and multiple mutations; (ii) those in the active site and distal from it; and (iii) those both affecting substrate binding and catalysis directly and through subtle changes in enzyme structure. The disadvantage of this approach is that any physical library generated by random mutagenesis is only a small subset of all possible libraries generated by this method, because it is impossible to sample all possible permutations of multiple mutations for all but the shortest

sequences. Library HRP-E contains $\sim 1.6 \times 10^6$ different HRP variants.

The active-site-directed library, HRP-C, was generated by exhaustive randomization of five positions at or near the active site: Phe68, Gly69, Asn72, Ser73, and Ala74 (Figure 1), allowing any of the 19 non-Cys amino acid residues to occur at each of the 5 positions. (Cysteine was excluded to avoid possible disruption of folding, dimerization, or aggregation of HRP through unpaired cysteines under the oxidizing conditions found in the yeast secretory apparatus and outside the cell.) The limited number of residues randomized allows an exhaustive sampling of the 2.5×10^6 possible sequence permutations. The proximity of the randomized sites to the active site (Figure 1A) ensures that many of the mutations will have a significant effect on enzyme activity; however, it leads to the risk that the mutations may be too drastic to preserve activity.

To compare the effectiveness of the two strategies for library construction, we performed in vitro selections for substrate enantioselectivity of HRP by using libraries HRP-E and HRP-C in parallel, under the same conditions and selection pressure, and then analyzed the most successful clones selected from each library.

Selection of Enantioselective HRP Variants

Each library underwent two parallel selections by FACS, one for enantioselectivity for D-tyrosinol over L-tyrosinol

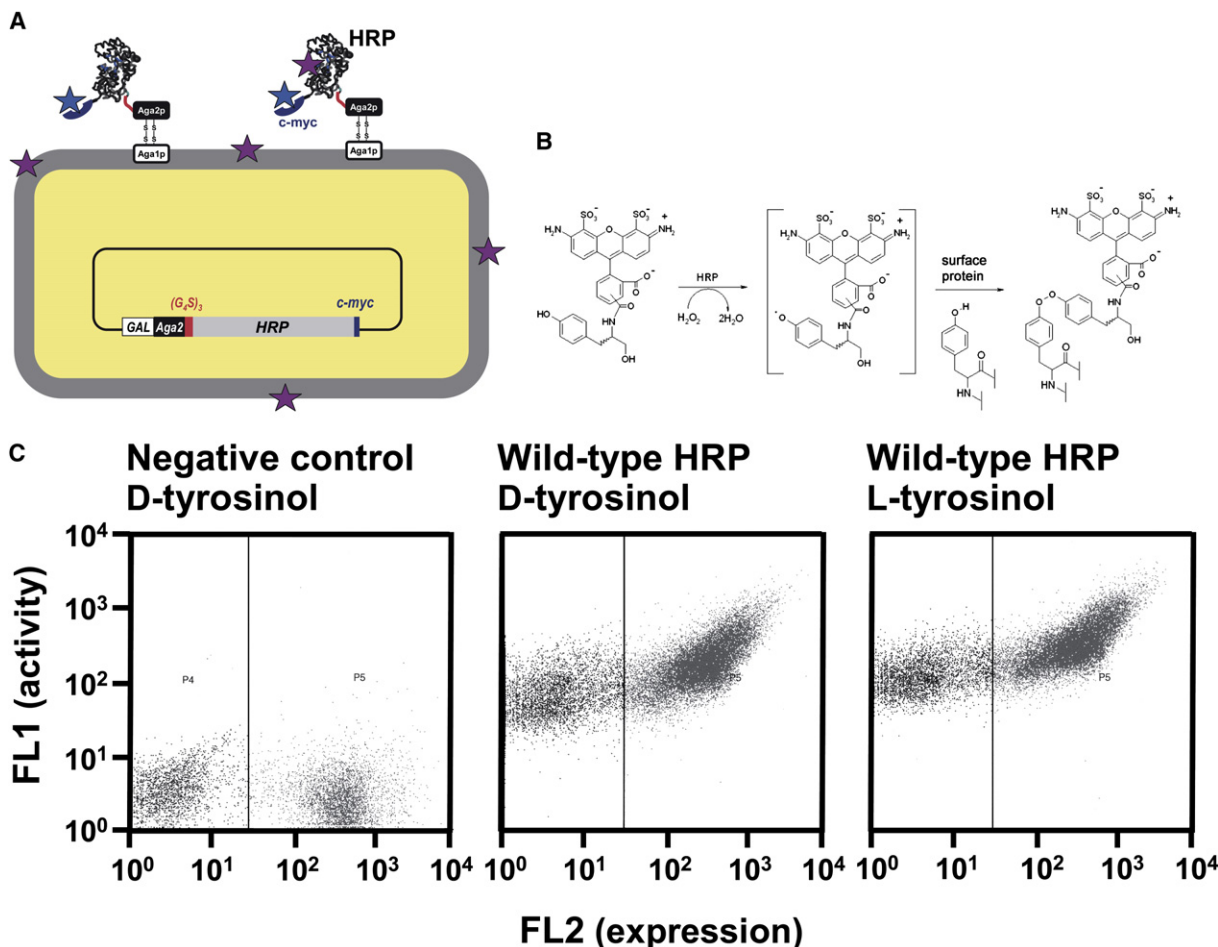


Figure 2. Yeast Surface Display of HRP

(A) The HRP gene fused to Aga2p and the c-myc tag is secreted from yeast and captured on the outside surface of the yeast cell through Aga1p-Aga2p disulfide bonding. Antibodies against the c-myc tag are used to label those yeast cells that display HRP on their surfaces with Alexa 633 (blue star). Active HRP attaches the substrate labeled with Alexa 488 (purple star) to an acceptor substrate (tyrosine) on the yeast surface.

(B) HRP-catalyzed radical polymerization of tyrosine and Alexa 488-tyrosinol.

(C) Analytical flow cytometry of yeast displaying negative-control yeast with no HRP gene and wild-type HRP by using Alexa 488-D- and L-tyrosinol.

($E_{D/L}$), and one for L-tyrosinol over D-tyrosinol ($E_{L/D}$), with alternating rounds of positive and negative selection. For example, the selection for $E_{D/L}$ alternated between FACS of populations with the highest incorporation of Alexa 488-labeled D-tyrosinol (selection rounds 1, 3, 5, and 7; Figure 3A) and FACS of populations with low incorporation of Alexa 488-labeled L-tyrosinol (selection rounds 2, 4, 6, and 8; Figure 3B).

Between 21 and 24 clones from each selected population were sequenced, enantioselectivity of all of the clones that appeared in the selected population more than once was determined (Figure 4 and Tables 1 and 2).

Each selection from the active-site-directed library, HRP-C (Table 1), yielded a single HRP variant that was enriched more than any other selected clone, and whose enantioselectivity exceeded that of all other clones in the selected population and that of wild-type HRP. Variant CD8.02, whose sequence was found in 76% of the clones in the final population selected for $E_{D/L}$, has a 3.4-fold pref-

erence for D- over L-tyrosinol, i.e., a 3.8-fold improvement over wild-type HRP. Similarly, clone CL8.01, whose sequence was found in 52% of the clones in the final population selected for $E_{L/D}$, has a 9-fold preference for L- over D-tyrosinol, which is a 7.5-fold improvement over wild-type HRP.

In contrast, the most highly represented clones selected from the randomly mutagenized library, HRP-E (Table 2), for $E_{D/L}$, were ED8.05 and ED8.02, which represented 19% and 14%, respectively, of selected clones; their enantioselectivity was indistinguishable from that of the wild-type HRP. Similarly, the selection from library HRP-E for $E_{L/D}$ produced no HRP variants with a higher enantioselectivity than the wild-type HRP; the sequence of wild-type HRP was found in 74% of the sequenced clones from that selected population.

In summary (Figure 5), selection from the active-site-directed library yielded variants with both a higher $E_{D/L}$ and $E_{L/D}$, but selections from the error-prone PCR-generated

Table 1. HRP Variants from an Active-Site-Directed Library Selected for Enantioselectivity for D- or L-Tyrosinol

| Selection for Enantioselectivity for D- over L-Tyrosinol | | | |
|--|----------------|-------|---------------|
| Clone | Sequence | Freq. | $E_{D/L}$ |
| WT HRP | <u>FGNANSA</u> | | 0.9 ± 0.1 |
| CD8.02 | EP..KA. | 76% | 3.4 ± 0.2 |
| CD8.14 | RP..HWT | 10% | 0.6 ± 0.1 |
| CD8.01 | WV..FWS | 5% | NA |
| CD8.07 | MV..PMG | 5% | NA |
| CD8.11 | HS..GM. | 5% | NA |
| Selection for Enantioselectivity for L- over D-Tyrosinol | | | |
| Clone | Sequence | Freq. | $E_{L/D}$ |
| WT HRP | <u>FGNANSA</u> | | 1.2 ± 0.1 |
| CL8.01 | LA..ELY | 52% | 9 ± 2 |
| CL8.09 | WA..AM. | 17% | 1.9 ± 0.1 |
| CL8.02 | .A..VVT | 13% | 3.3 ± 0.7 |
| CL8.03 | HA..ARD | 13% | 1.4 ± 0.1 |
| CL8.16 | RH..WTT | 4% | NA |

HRP variants from the active-site-directed library, HRP-C, were selected for enantioselectivity for D-tyrosinol ($E_{D/L}$) or L-tyrosinol ($E_{L/D}$). The sequence in the randomized region (68–74) is shown, and the residues randomized to generate library HRP-C are underlined. NA, not active (less than 10% of wild-type HRP activity). Errors were derived from two independent experiments.

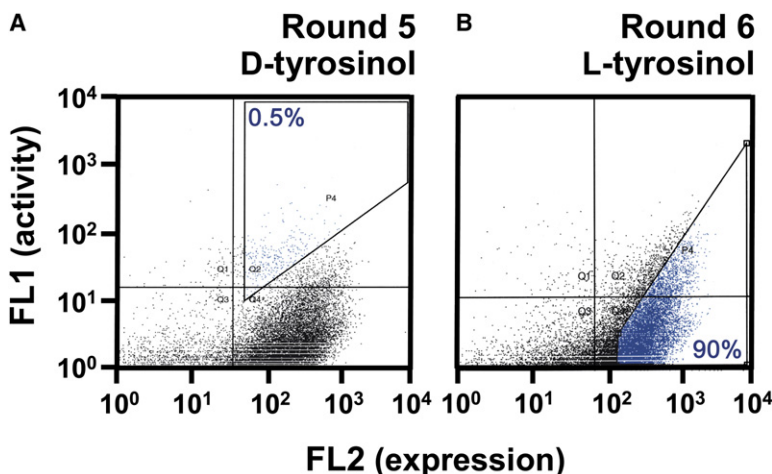
library failed to identify any HRP variants with a significant change in enantioselectivity.

Mutational Analysis of Enantioselective Variant CD8.02

The enantioselectivities of CD8.02-based mutants, in which one of the four mutations at a time was reverted back to the wild-type sequence, are shown in Table 3.

Two of the mutants, Rev69 (CD8.02(P69G)) and Rev72 (CD8.02(K72N)), lost activity against both L- and D-tyrosinol to the degree that precluded accurate determination of their enantioselectivities. Mutant Rev68 (CD8.02(E68F)) remained sufficiently active, but its $E_{D/L}$ was half that of CD8.02 (1.7 compared to 3.4). In contrast, mutant Rev73 (CD8.02(A73S)) had an even higher enantioselectivity than its parent clone, $E_{D/L} = 5.5$, which corresponds to a 6-fold improvement over wild-type HRP.

That a single mutation from the selected CD8.02 sequence back to the wild-type at position 68, 69, or 72 abolishes activity or reduces enantioselectivity suggests that, in the context of the selected CD8.02 sequence (including Ala73), Glu68, Pro69, and Lys72 are all required for catalytic activity and high $E_{D/L}$. Such a requirement for three non-wild-type residues is one possible explanation for the failure of the error-prone PCR-generated library in this selection, but others exist as well. Whereas error-prone PCR is relatively efficient at sampling single and double mutations throughout the HRP gene, the odds of generating a particular combination of three mutations at specific sites are low. In contrast, our active-site-directed library, HRP-C, focused attention on and essentially enumerated all combinations of mutations at positions 68, 69, 72, 73, and 74. This complementary strategy enabled the discovery of the highly enantioselective mutant CD8.02, which requires three simultaneous changes from the wild-type sequence for its favorable phenotype. It is not possible to generalize to other cases from this one example, but the relative effectiveness of whole-gene error-prone methods that provide excellent coverage of single and probably double mutants versus focused enumeration methods of all multiple mutants at a small number of sites remains an open research question. Other strategies not employed here are also possible. The efficiency of constructing multiple mutations nearby in three-dimensional space by using a focused procedure might be particularly useful near the enzyme active and binding sites, where there may be a high level of cooperativity. For identifying more distributed and less cooperative mutations, approaches similar to the error-prone library utilized here

**Figure 3. Selection for HRP Enantioselectivity**

(A and B) Populations shown were selected from the active-site-directed library, HRP-C, for enantioselectivity for D- over L-tyrosinol ($E_{D/L}$). (A) Positive selection in round 5. The 0.5% of HRP-expressing cells with the highest Alexa 488-D-tyrosinol signal was collected. (B) Negative selection in round 6. The 90% of HRP-expressing cells with the lowest Alexa 488-L-tyrosinol signal was collected.

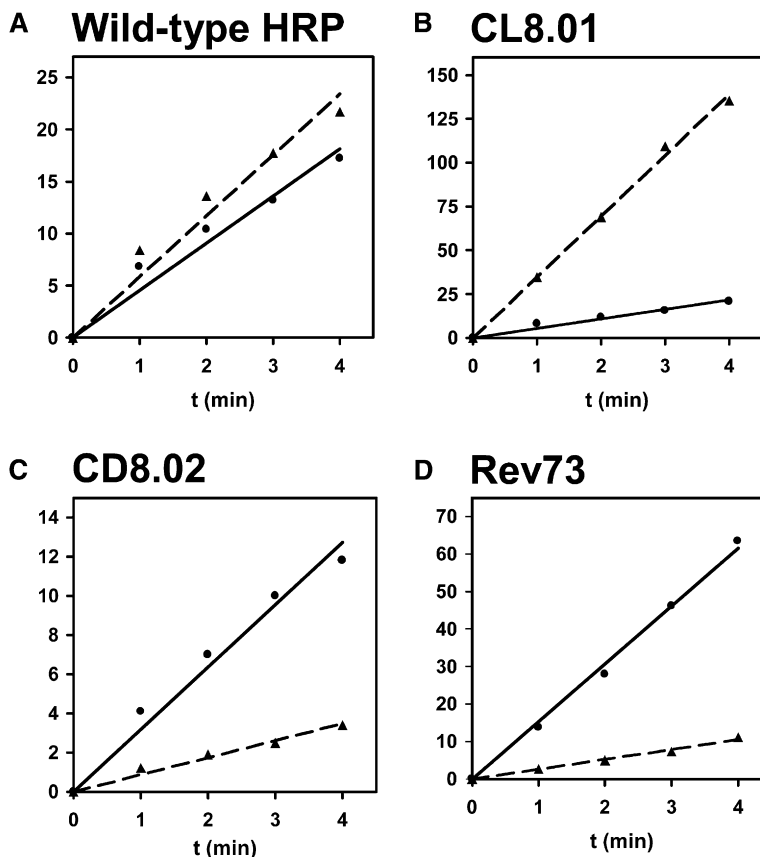


Figure 4. Determination of Enantioselectivity

(A–D) Enantioselectivity of wild-type HRP, as well as those of the most enantioselective selected clones and the most enantioselective site-directed mutant, determined by comparison of the enzymatic initial rates by using Alexa 488-D- versus L-tyrosinol. (A) Wild-type HRP, (B) CL8.01, (C) CD8.02, (D) Rev73. Bullets, solid line, incorporation of Alexa 488-D-tyrosinol; triangles, dashed line, incorporation of Alexa 488-L-tyrosinol; MFU, mean fluorescence units.

might have an advantage, particularly if single and double mutants can be combined to produce further enhancements.

In principle, mutant Rev73, which had a higher enantioselectivity than CD8.02, should have been encoded in the HRP-C library and selected in the selection for $E_{D/L}$. Two possible explanations for not selecting Rev73 from the library are that a 5-fold oversampling of the theoretical sequence space in the physical library was not sufficient to include a copy of each possible sequence, and that Rev73 had other properties that were selected against, such as a lower expression level in yeast. Nevertheless, we expect that thoroughly sampled active-site-directed libraries should provide an advantage in *in vitro* evolution of activity for enzymes whose structure and location of the active site are known. This hypothesis is supported by other enzyme-directed evolution studies [37].

Further Directions for Yeast-Based *In Vitro* Evolution of Enzyme Activity

The use of FACS to capture variants of interest requires physical association of product with the yeast cell that harbors the gene for the enzyme variant. In this study, we ensured this genotype-phenotype linkage by utilizing as one of the substrates Tyr residues that are ubiquitous on the surface of yeast. However, the use of yeast surface display is not limited to the study of enzymes whose substrates are natural components of the yeast cell wall. We

propose that this method can be applied to other bimolecular reactions by tethering one of the synthetic substrates to the surface of yeast (analogously to the tyrosine naturally present on the yeast cell wall in the HRP example), and by adding the second substrate in solution (like tyrosinol in the HRP example). A generalizable method for covalently attaching a small molecule to the yeast surface was recently demonstrated for biotin, which was attached to the yeast surface through a PEG linker with an NHS functional group [38]. Furthermore, the HRP-initiated generation of free radicals captured by a cell might be a generic means for detecting the reaction products of other enzymes that unmask a pro-substrate, which then serves as a substrate for HRP [6].

SIGNIFICANCE

To our knowledge, we present the first application of yeast surface display to *in vitro* selection of altered enzymatic activity. The method immobilizes one of the reaction substrates, as well as a library of enzyme variants, on the surface of live yeast cells. A second, fluorescent substrate is then supplied in solution and is utilized by the active enzyme variants to label those cells that express such active variants. Labeled yeast cells are subsequently captured by fluorescence-activated sorting. The use of a eukaryotic organism to display the enzyme under selection makes possible the in

Table 2. HRP Variants from a Randomly Mutagenized Library Selected for Enantioselectivity for D- or L-Tyrosinol

| Selection for Enantioselectivity for D- over L-Tyrosinol | | | |
|--|---------------------------|-------|----------------|
| Clone | Mutations | Freq. | $E_{D/L}$ |
| WT HRP | | | 0.9 ± 0.1 |
| ED8.05 | S73L | 19% | 0.8 ± 0.1 |
| WT HRP | | 14% | 0.9 ± 0.1 |
| ED8.02 | G69A, S126G, A134T, P146S | 14% | 0.9 ± 0.1 |
| ED8.01 | L215R, T257S, N268S | 5% | ND |
| ED8.04 | N9S, I22L | 5% | ND |
| ED8.06 | S216N | 5% | ND |
| ED8.07 | P261A | 5% | ND |
| ED8.08 | E249G | 5% | ND |
| ED8.10 | R93G | 5% | ND |
| ED8.11 | F221L | 5% | ND |
| ED8.16 | N214S, N307D | 5% | ND |
| ED8.17 | S73T | 5% | ND |
| ED8.18 | S126G | 5% | ND |
| Selection for Enantioselectivity for L- over D-Tyrosinol | | | |
| Clone | Sequence | Freq. | $E_{L/D}$ |
| WT HRP | | | 1.2 ± 0.1 |
| WT HRP | | 74% | 1.2 ± 0.1 |
| EL8.02 | I32V, G213D | 17% | 1.2 ± 0.04 |
| EL8.08 | S216G | 4% | ND |
| EL8.22 | V235M | 4% | ND |

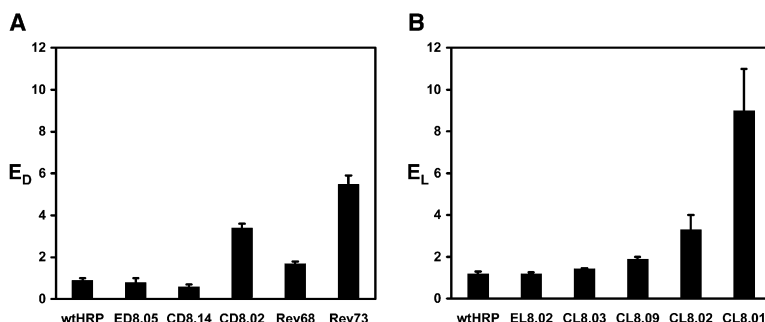
HRP variants from the randomly mutagenized library, HRP-E, were selected for enantioselectivity for D-tyrosinol ($E_{D/L}$) or L-tyrosinol ($E_{L/D}$). Mutations from the wild-type HRP sequence are shown. ND, not determined. Errors were derived from two independent experiments.

vitro evolution of a number of enzymes that cannot be expressed in a soluble and active form in bacteria, such as highly disulfide-crosslinked enzymes.

We used a combination of positive and negative selections to identify variants of HRP that are enantioselective for D- or L-tyrosinol, and we succeeded at enhancing and even reversing the enantioselectivity from that of the slight preference for L-tyrosinol shown by wild-type HRP to a substantial preference for D-tyrosinol, a 4-fold change in enantioselectivity. In a separate selection, we improved the enantioselectivity for

L-tyrosinol compared to wild-type HRP 8-fold. A comparison of selections from two different HRP-based libraries revealed that an active-site-directed library yielded variants with a large change in enantioselectivity, whereas a randomly mutagenized library failed to yield improved clones; this difference could be due to the superior sampling of multiple mutations in the vicinity of the active site by the active-site-directed library.

The immobilized substrate used in our selection was tyrosine, present naturally in proteins associated with

**Figure 5. Enantioselectivities of Wild-Type and Variant HRP**

(A and B) Enantioselectivities of wild-type HRP and variants selected for specificity for (A) D- over L-tyrosinol and (B) L- over D-tyrosinol. Errors were derived from two independent experiments.

Table 3. Properties of Single-Site Revertants of Variant CD8.02

| Clone | Sequence | E _{D/L} |
|--------|----------------|------------------|
| WT HRP | <u>FGNANSA</u> | 0.9 ± 0.1 |
| CD8.02 | EP..KA. | 3.4 ± 0.2 |
| Rev68 | .P..KA. | 1.7 ± 0.1 |
| Rev69 | E...KA. | NA |
| Rev72 | EP...A. | NA |
| Rev73 | EP..K.. | 5.5 ± 0.4 |

E_{D/L}, enantioselectivity for D-tyrosinol; NA, not active (less than 10% of wild-type HRP activity). The residues randomized to generate library HRP-C are underlined. Errors were derived from two independent experiments.

the yeast cell wall. Owing to a recent development in derivatizing the yeast surface with a variety of small molecules, the scope of enzyme yeast surface display can be extended by using any nontoxic substrate that can be conjugated to a standard linker.

EXPERIMENTAL PROCEDURES

Synthesis and Cloning of the Wild-Type HRP Gene

The gene for wild-type HRP, redesigned to introduce a number of unique restriction sites without altering the protein sequence (Figure 1B), was synthesized by GenScript (Piscataway, NJ). A new yeast surface display vector, pCTcon2, was derived from plasmid pCTcon [10] by replacing the DNA encoding the (Gly-Gly-Gly-Ser)₃ linker with a less repetitive DNA sequence (5'-GGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTTCG-3'), again without mutating the encoded peptide-linker sequence. The synthetic HRP gene and the pCTcon2 plasmid were digested with NheI and BamHI, and the HRP gene was ligated into the BamHI-NheI backbone of pCTcon2. The resulting plasmid, pCT2-HRP, was transformed into the yeast surface display strain of *S. cerevisiae*, EBY100 [10].

Construction of the HRP-Based Library, HRP-E, by Using Error-Prone PCR

Library HRP-E was made by amplifying the HRP insert in pCT2-HRP in the presence of nucleotide analogs as described previously [10]. Cotransformation of EBY100 with the BamHI-EcoRI backbone of pCT2con and the amplified, mutated HRP gene by following the published method [10] yielded a library of 1.6×10^6 clones in EBY100. DNA sequencing of 24 library clones revealed 0–17 mutations per clone (at the nucleotide level), with a median of 3 mutations per clone. Two of the 24 sequenced clones had the wild-type HRP gene sequence.

Selection of the Five Active-Site Positions for Randomization

Groups of 5 residue positions were chosen based on structural proximity to the active site, and computational protein design techniques (the dead-end elimination [39] and A* algorithms [40]) were used to determine allowed sequences for the wild-type backbone structure within 15 kcal/mol of the wild-type energy, which corresponds to the free energy of unfolding of an extremely stable protein [41]. The pairwise energy function for these calculations was the sum of the van der Waals, solvent-accessible surface area [42] and a Coulombic electrostatic term with a dielectric constant of four times the distance between each pair of atoms [43]. These calculations highlighted multiple sets of candidate positions that were calculated to allow many sequences. Further analysis of the built structures shows that they did not fill the active site with heavy atoms or consist of many charged res-

idues. We then used three other metrics to choose the positions for randomization: mutual information between positions in our sequence alignments, amino acid frequency in known genes, and ease of synthesis. Mutual information was used to seek interactive positions, which are of special interest. The mutual information [44] between positions highlights pairs that might be structurally dependent on each other and therefore might be forced to mutate in unison. A sequence alignment of HRP genes was taken from Pfam [45], and our wild-type HRP gene sequence was aligned by eye with the most highly homologous of the 309 seed alignment sequences to create a sequence alignment of 310 sequences. Highly conserved positions in this alignment were considered to be risky for randomization. We chose to mutate Phe68, Gly69, Asn72, Ser73, and Ala74, which the computational protein design techniques indicated would allow mutation. These 5 residues had low conservation in our sequence alignment and moderate mutual information; in addition, their close proximity in sequence made them easy to modify with a single randomized oligonucleotide.

Construction of the HRP-Based Library, HRP-C, Using Active-Site-Directed Saturation Mutagenesis

Library HRP-C was constructed by replacing the BstBI-EagI fragment of the wild-type HRP gene (Figure 1B) with a synthetic DNA fragment randomized at amino acid positions 68, 69, 72, 73, and 74. The synthetic DNA fragment was assembled from two defined oligonucleotides, c1 (5'-GTTGTGACGCATCGATCTTGTAGACACACACATCA TTTCGAACAGAGAAAGATGCG-3') and c2 (5'-CTGCGCAGGATACAG GTTCTTGGGCATGCACTCTCCACGGCCGCTTCATTCTGTCAATCA CAG-3'GAAATCCGC), and one randomized oligonucleotide, rC (5'-CA TCATTTGGAACAGAGAAAGATGCG11AACGCA111CGCGGATTTCTCT GTGATTGACAGAAATG-3'), where "1" stands for an equimolar mixture of codons encoding for the 19 non-cysteine amino acid residues. The triphosphoramidite codon mixture was purchased from Glen Research (Sterling, VA), and the randomized oligonucleotide was synthesized manually by Trilink (San Diego, CA).

The oligonucleotides were assembled by using KOD Hot Start Polymerase (Novagen, San Diego, CA). First, 20 pmol oligonucleotide c2 and 10 pmol oligonucleotide rC were combined with 1 U KOD Hot Start Polymerase in 50 µl KOD Mix (1 × KOD buffer, 0.2 mM dNTP mix, 1 mM MgSO₄, 1 M betaine, and 3% DMSO). The oligonucleotides were denatured for 2 min at 95°C; subjected to ten cycles of 30 s at 94°C, 30 s at 58°C, and 1 min at 68°C; and, finally, incubated for 10 min at 68°C. A total of 20 pmol of c1 was added to the mixture in 2 µl, and the thermocycling program was repeated as described above. The resulting double-stranded DNA fragment was ethanol precipitated, and 2 µg of the product was amplified 10-fold by limiting the amounts of the PCR primers pc1 (5'-GTTGTGACGCATCGATCTTGTAGAC-3') and pc2 (5'-CTGCGCAGGATACAGTTCTTGGGC-3'), and by using 20 cycles of the program described above with 30 U KOD Hot Start Polymerase in 1.5 ml KOD mix. The amplified DNA fragment ("HRP-C insert") was again ethanol precipitated and resuspended in dH₂O at 0.6 µg/µl.

The pCT2-HRP plasmid missing the BstBI-EagI fragment (Figure 1B) was prepared by a sequence of restriction digests (EagI, BssHII, and BstBI), followed by purification on QIAGEN PCR-Purification columns and ethanol precipitation. The gapped pCT2-HRP plasmid and HRP-C insert, which overlapped in sequence with the ends of the gapped pCT2-HRP plasmid by 41 nucleotides both upstream of the BstBI restriction site and downstream of the EagI site, were cotransformed into EBY100 by following the established protocol [10]. A total of 20 µg of gapped pCT2-HRP and 30 µg HRP-C insert were transformed into 1 ml electro-competent EBY100, yielding a yeast surface display library of an estimated 9.0×10^7 independent transformants, which is larger than the 2.5×10^6 possible sequence permutations permitted by library design. Of the 24 clones from library HRP-C that were sequenced, 22 conformed to the library design, and 2 showed protein truncations due to frameshift mutations.

Synthesis of Fluorescently Labeled Tyrosinol Substrates

Tyrosinol (Sigma-Aldrich, St. Louis, MO) dissolved in 50 mM sodium borate buffer (pH 8.6) was added in 10-fold molar excess to Alexa Fluor 488 succinimidyl ester (Molecular Probes/Invitrogen, Carlsbad, CA). The mixture was stirred at room temperature for 3 hr. The fluorescently labeled product was then purified by reverse-phase HPLC by using a 9.4 × 250 mm, 5 μ M Zorbax Rx-C8 column (Agilent Technologies, Santa Clara, CA) with 0.1% TFA, by using water as loading buffer and 0.1% TFA acetonitrile as mobile phase. The product was eluted with a 30 min, 4 ml/min gradient of 10%–30% acetonitrile.

Labeling of HRP Libraries Displayed on the Yeast Surface

A yeast culture containing either 10 copies of each clone or 2 × 10⁶ cells, whichever was the larger, was induced at the cell density of 4 × 10⁵/ml by growing the culture in 90% SG-CAA, 10% SD-CAA, 3.6 mM δ -aminolevulinic acid, and 0.2 mM ferric citrate for 18 hr at 30°C. Two million induced yeast cells were washed with 1 ml PBS containing 0.5% BSA, followed by 0.5 ml PBS with 0.1% BSA. The cells were resuspended in 200 μ l PBS containing 0.003% H₂O₂ and 15 μ M L- or D-tyrosinol-Alexa 488. Yeast populations for FACS were incubated for 30 min at 30°C, whereas samples used to determine enantioselectivities of selected clones were incubated for 2–8 min at room temperature. Labeling reactions were stopped by adding a 10× excess of PBS containing 0.5% BSA and 10 mM ascorbic acid and were washed by 0.5 ml PBS with 0.1% BSA. Samples for FACS were then labeled as described previously [10] with anti-*c-myc* monoclonal antibody, 9E10 (Covance, Princeton, NJ), and with goat anti-mouse Alexa-PE polyclonal antibodies, and then were resuspended in 0.5 ml PBS with 0.1% BSA.

Sorting of HRP Libraries Displayed on the Yeast Surface Using Fluorescence-Activated Cell Sorting

Double-labeled yeast cells were sorted on a Becton Dickinson (Franklin Lakes, NJ) Aria high-speed cell sorter with 488 nm and 635 nm lasers at the rates of 6,000–10,000 cells per s. Gates were adjusted to collect the yeast cells positive for Alexa 633 signal that also had the highest Alexa 488 signal (for the positive selection rounds 1, 3, 5, and 7) or the lowest Alexa 488 signal (for the negative selection rounds 2, 4, 6, and 8). Of the Alexa 633-positive cells, the 1% of the cells with the highest Alexa 488 signal was collected in round 1, whereas 0.5% of the cells with the highest Alexa 488 signal was collected in rounds 3, 5, and 7. Conversely, of the Alexa 633-positive cells, 3% of the cells with the highest Alexa 488 signal was excluded in round 2, and 10% of the cells with the highest Alexa 488 signal was excluded in rounds 4, 6, and 8. Selected cells were collected in 0.5 ml SD-CAA (pH 4.5), containing 50 μ g/ml kanamycin, 100 U/ml penicillin G, and 200 U/ml streptomycin, then grown to saturation in 5 ml of the same media by shaking them at 30°C for 2 days before they were induced and labeled for the next round of sorting.

Isolation of Selected HRP Variants

After eight rounds of selection, plasmid DNA was extracted from 1 ml of each saturated culture by using the Zymoprep Yeast Plasmid Miniprep Kit (Zymo Research, Orange, CA) and was transformed into XL1-Blue-competent *E. coli* (Stratagene, La Jolla, CA). Plasmids from 21–23 colonies from each selection were sequenced; those encoding unique HRP variants were retransformed into EBY100 for characterization of enantioselectivity.

Characterization of HRP Variants

To determine the enantioselectivities of selected HRP variants, yeast cells transformed with each variant of interest were labeled in parallel with L- and D-tyrosinol as described above for 0–4 min. Each time point sample was analyzed by using a Coulter (Fullerton, CA) Epics XL flow cytometer. The mean fluorescence of Alexa 488 was plotted against time to determine the initial reaction rates with each substrate (Figure 4), and the enantioselectivity was calculated as $E_{D/L}$ = (initial rate

with D-tyrosinol)/(initial rate with L-tyrosinol) and $E_{L/D}$ = (initial rate with L-tyrosinol)/(initial rate with D-tyrosinol). Clearly, $E_{D/L} \times E_{L/D} = 1$.

Mutagenesis of Enantioselective Variant CD8.02

The four non-wild-type residues in variant CD8.02, Glu68, Pro69, Lys72, and Ala73, were mutated back to wild-type HRP by using the same strategy as in the construction of library HRP-C, except that oligonucleotides with defined sequences were used instead of the randomized oligonucleotide, rC. Mutant Rev68 (CD8.02 E68F) was constructed by using oligonucleotide 5'-CATCATTTTCGAACAGAGAAAGATGCGTTTCCTAACGCAAAGGCAGCGCGGATTTCCTGTGATGACAGAATG-3'; mutant Rev69 (CD8.02 P69G) was constructed by using oligonucleotide 5'-CATCATTTTCGAACAGAGAAAGATGCGGAGGGAACGCAAAGGCAGCGCGGATTTCCTGTGATTGACAGAATG-3'; mutant Rev72 (CD8.02 K72N) was constructed by using oligonucleotide 5'-CATCATTTTCGAACAGAGAAAGATGCGGAGCCTAACGCAACGCGAGCGCGGATTTCCTGTGATTGACAGAATG-3'; and mutant Rev73 (CD8.02 A73S) was constructed by using oligonucleotide 5'-CATATTTCGAACAGAGAAAGATGCGGAGCCTAACGCAAGTCGGCGCGGATTTCCTGTGATTGACAGAATG-3'.

ACKNOWLEDGMENTS

This project was funded by the DuPont-MIT Alliance. We thank the MIT Flow Cytometry Core Facility for technical support, and Dr. Mark J. Nelson and Dr. Aaron Scurto for helpful discussions.

Received: January 31, 2007

Revised: August 28, 2007

Accepted: September 18, 2007

Published: October 26, 2007

REFERENCES

1. Drauz, K., and Waldmann, H. (2002). Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook, Second Edition (Weinheim: VCH).
2. Cherry, J.R., and Fidantsef, A.L. (2003). Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* 14, 438–443.
3. Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D., and Arnold, F.H. (2005). Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* 15, 447–452.
4. Schaffitzel, C., and Plückthun, A. (2001). Protein-fold evolution in the test tube. *Trends Biochem. Sci.* 26, 577–579.
5. Aharoni, A., Griffiths, A.D., and Tawfik, D.S. (2005). High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* 9, 210–216.
6. Becker, S., Schmoldt, H.U., Adams, T.M., Wilhelm, S., and Kolmar, H. (2004). Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr. Opin. Biotechnol.* 15, 323–329.
7. Hilvert, D. (2000). Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* 69, 751–793.
8. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* 12, 1281–1289.
9. Chames, P., Epinat, J.C., Guillier, S., Patin, A., Lacroix, E., and Paques, F. (2005). In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res.* 33, e178.
10. Colby, D.W., Kellogg, B.A., Graff, C.P., Yeung, Y.A., Swers, J.S., and Wittrup, K.D. (2004). Engineering antibody affinity by yeast surface display. *Methods Enzymol.* 388, 348–358.

11. Feldhaus, M.J., and Siegel, R.W. (2004). Yeast display of antibody fragments: a discovery and characterization platform. *J. Immunol. Methods* 290, 69–80.
12. Kim, Y.S., Bhandari, R., Cochran, J.R., Kuriyan, J., and Wittrup, K.D. (2006). Directed evolution of the epidermal growth factor receptor extracellular domain for expression in yeast. *Proteins* 62, 1026–1035.
13. Shusta, E.V., Holler, P.D., Kieke, M.C., Kranz, D.M., and Wittrup, K.D. (2000). Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat. Biotechnol.* 18, 754–759.
14. Shiraga, S., Ishiguro, M., Fukami, H., Nakao, M., and Ueda, M. (2005). Creation of *Rhizopus oryzae* lipase having a unique oxyanion hole by combinatorial mutagenesis in the lid domain. *Appl. Microbiol. Biotechnol.* 68, 779–785.
15. Veitch, N.C. (2004). Horseradish peroxidase: a modern view of a classic enzyme. *Phytochemistry* 65, 249–259.
16. Morawski, B., Quan, S., and Arnold, F.H. (2001). Functional expression and stabilization of horseradish peroxidase by directed evolution in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 76, 99–107.
17. Cherry, J.R., Lamsa, M.H., Schneider, P., Vind, J., Svendsen, A., Jones, A., and Pedersen, A.H. (1999). Directed evolution of a fungal peroxidase. *Nat. Biotechnol.* 17, 379–384.
18. Iffland, A., Tafelmeyer, P., Saudan, C., and Johnsson, K. (2000). Directed molecular evolution of cytochrome c peroxidase. *Biochemistry* 39, 10790–10798.
19. Ni, J., Sasaki, Y., Tokuyama, S., Sogabe, A., and Tahara, Y. (2002). Conversion of a typical catalase from *Bacillus* sp. TE124 to a catalase-peroxidase by directed evolution. *J. Biosci. Bioeng.* 93, 31–36.
20. Yin, J., Mills, J.H., and Schultz, P.G. (2004). A catalysis-based selection for peroxidase antibodies with increased activity. *J. Am. Chem. Soc.* 126, 3006–3007.
21. Reetz, M.T. (2004). Changing the enantioselectivity of enzymes by directed evolution. *Methods Enzymol.* 388, 238–256.
22. Reetz, M.T. (2004). Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. *Proc. Natl. Acad. Sci. USA* 101, 5716–5722.
23. Henke, E., and Bornscheuer, U.T. (1999). Directed evolution of an esterase from *Pseudomonas fluorescens*. Random mutagenesis by error-prone PCR or a mutator strain and identification of mutants showing enhanced enantioselectivity by a resorufin-based fluorescence assay. *Biol. Chem.* 380, 1029–1033.
24. Koga, Y., Kato, K., Nakano, H., and Yamane, T. (2003). Inverting enantioselectivity of *Burkholderia cepacia* KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and in vitro expression. *J. Mol. Biol.* 331, 585–592.
25. Park, S., Morley, K.L., Horsman, G.P., Holmquist, M., Hult, K., and Kazlauskas, R.J. (2005). Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations. *Chem. Biol.* 12, 45–54.
26. Funke, S.A., Otte, N., Eggert, T., Bocola, M., Jaeger, K.E., and Thiel, W. (2005). Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Eng. Des. Sel.* 18, 509–514.
27. Reetz, M.T., Wang, L.W., and Bocola, M. (2006). Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int. Ed. Engl.* 45, 1236–1241.
28. May, O., Nguyen, P.T., and Arnold, F.H. (2000). Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nat. Biotechnol.* 18, 317–320.
29. Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T., and Jaeger, K.E. (2000). Directed evolution of an enantioselective lipase. *Chem. Biol.* 7, 709–718.
30. Horsman, G.P., Liu, A.M., Henke, E., Bornscheuer, U.T., and Kazlauskas, R.J. (2003). Mutations in distant residues moderately increase the enantioselectivity of *Pseudomonas fluorescens* esterase towards methyl 3-bromo-2-methylpropanoate and ethyl 3-phenylbutyrate. *Chemistry (Easton)* 9, 1933–1939.
31. Peters, M.W., Meinhold, P., Glieder, A., and Arnold, F.H. (2003). Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J. Am. Chem. Soc.* 125, 13442–13450.
32. van Loo, B., Spelberg, J.H., Kingma, J., Sonke, T., Wubbolts, M.G., and Janssen, D.B. (2004). Directed evolution of epoxide hydrolase from *A. radiobacter* toward higher enantioselectivity by error-prone PCR and DNA shuffling. *Chem. Biol.* 11, 981–990.
33. Kubo, T., Peters, M.W., Meinhold, P., and Arnold, F.H. (2006). Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3. *Chemistry (Easton)* 12, 1216–1220.
34. Morley, K.L., and Kazlauskas, R.J. (2005). Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* 23, 231–237.
35. Shusta, E.V., Kieke, M.C., Parke, E., Kranz, D.M., and Wittrup, K.D. (1999). Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* 292, 949–956.
36. Graff, C.P., Chester, K., Begent, R., and Wittrup, K.D. (2004). Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37°C. *Protein Eng. Des. Sel.* 17, 293–304.
37. Chica, R.A., Doucet, N., and Pelletier, J.N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* 16, 378–384.
38. Rakestraw, J.A., Baskaran, A.R., and Wittrup, K.D. (2006). A flow cytometric assay for screening improved heterologous protein secretion in yeast. *Biotechnol. Prog.* 22, 1200–1208.
39. Desmet, J., De Maeyer, M., Hazes, B., and Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356, 539–542.
40. Leach, A.R., and Lemon, A.P. (1998). Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. *Proteins* 33, 227–239.
41. Creighton, T.E. (1984). *Proteins: Structures and Molecular Properties* (New York: W.H. Freeman and Company).
42. Sitkoff, D., Sharp, K.A., and Honig, B. (1994). Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.* 98, 1978–1988.
43. Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249.
44. MacKay, D.J.C. (2003). *Information Theory, Inference, and Learning Algorithms* (Cambridge: Cambridge University Press).
45. Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000). The Pfam protein families database. *Nucleic Acids Res.* 28, 263–266.