A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display

Classification: Biological Sciences / Biochemistry

A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display

Irwin Chen, Brent M. Dorr, and David R. Liu*

Howard Hughes Medical Institute
Department of Chemistry and Chemical Biology
Harvard University
12 Oxford Street
Cambridge, MA 02138

*Correspondence:
Professor David R. Liu
Howard Hughes Medical Institute
Department of Chemistry and Chemical Biology
Harvard University
Cambridge, MA 02138 USA
drliu@fas.harvard.edu

Abstract

The ability to routinely generate efficient protein catalysts of bond-forming reactions chosen by researchers, rather than nature, is a longstanding goal of the molecular life sciences. Here we describe a directed evolution strategy for enzymes that catalyze, in principle, any bond-forming reaction. The system integrates yeast display, enzyme-catalyzed bioconjugation, and fluorescence-activated cell sorting to isolate cells expressing proteins that catalyze the coupling of two substrates chosen by the researcher. We validated the system using model screens for *Staphylococcus aureus* sortase A-catalyzed transpeptidation activity, resulting in enrichment factors of 6,000-fold after a single round of screening. We applied the system to evolve sortase A for improved catalytic activity. After eight rounds of screening, we isolated variants of sortase A with up to a 140-fold increase in LPETG-coupling activity compared with the starting wild-type enzyme. An evolved sortase variant enabled much more efficient labeling of LPETG-tagged human CD154 expressed on the surface of HeLa cells compared with wild-type sortase. As the method developed here does not rely on any particular screenable or selectable property of the substrates or product, it represents a powerful alternative to existing enzyme evolution methods.

Introduction

Despite the many attractive features of protein enzymes as catalysts for organic synthesis (1), as research tools (2-4), and as an important class of human therapeutics (5, 6), the extent and diversity of their applications remain limited by the difficulty of finding in nature or creating in the laboratory highly active proteins that catalyze chemical reactions of interest. A significant fraction of protein catalysts currently used for research and industrial applications was obtained through the directed evolution of natural enzymes (7). Current methods for the directed evolution of enzymes have resulted in some remarkable successes (8, 9), but generally suffer from limitations in reaction scope. For example, screening enzyme libraries in a multi-well format has proven to be effective for enzymes that

process chromogenic or fluorogenic substrates, and is typically limited to library sizes of ~10^2-10^6 members, depending on the nature of the screen and on available infrastructure (10). Selections of cell-based libraries that couple product formation with auxotrophy complementation (11) or transcription of a reporter gene (12) enable larger library sizes to be processed, but also suffer from limited generality because they rely on specific properties of the substrate or product. Likewise, *in vitro* compartmentalization is a powerful genotype-phenotype co-localization platform that has been used to evolve protein enzymes with improved turnover, but also requires corresponding screening or selection methods which thus far have been substrate- or product-specific (13).

Directed evolution strategies that are general for any bond-forming reaction would complement current methods that rely on screenable reactions or selectable properties of the substrate or product. In principle, chemical complementation using an adapted yeast three-hybrid assay is reaction-independent (14), but requires membrane-permeable substrates and offers limited control over reaction conditions because the bond-forming event must take place intracellularly. Phage-display and mRNA-display systems that are general for any bond-forming reaction have been used to evolve enzymes including DNA polymerases (15) and RNA ligases (16). These approaches also offer advantages of larger library sizes and significant control over reaction conditions because the enzymes are displayed extracellularly or expressed in the absence of a host cell.

Cell surface display (17-20) is an attractive alternative to phage and mRNA display. In contrast with other display methods, the use of bacterial or yeast cells enables up to 100,000 copies of a library member to be linked to one copy of the gene, increasing sensitivity during screening or selection steps. In addition, cell surface-displayed libraries are compatible with powerful fluorescence-activated cell sorting (FACS) that enable very large libraries to be screened efficiently (>10^7 cells per hour) with precise, quantitative control over screening stringency. The multicolor capabilities of FACS also enable normalization for enzyme display level during screening and simultaneous positive and negative screens, capabilities that are difficult to implement in phage and mRNA display.

In this work we integrated yeast display, enzyme-catalyzed small molecule-protein conjugation, and FACS into a general strategy for the evolution of proteins that catalyze bond-forming (coupling) reactions. We applied the system to evolve the bacterial transpeptidase sortase A for improved catalytic activity, resulting in sortase variants with up to 140-fold improvement in activity. In contrast with wild-type sortase, an evolved sortase enabled highly efficient cell-surface labeling of recombinant human CD154 expressed on the surface of live HeLa cells with a biotinylated peptide.

Results

Design and Implementation of a General System for the Evolution of Bond-Forming Enzymes

The enzyme evolution system is overviewed in Figure 1. Yeast cells display the enzyme library extracellularly as a fusion to the Aga2p cell surface mating factor, which is covalently bound to the Aga1p mating factor with a reactive handle that enables covalent attachment of substrate A to cells. We chose the S6 peptide (3) as the reactive handle to link substrate A to cells using Sfp phosphopantetheinyl transferase from *Bacillus subtilis*. Substrate B linked to an affinity handle (e.g. biotin, represented by the gray circle in Figure 1) is added to the substrate A-conjugated yeast display enzyme library. Due to the high effective molarity of substrate A with respect to each cell's displayed library member, both of which are immobilized on the cell surface, active library members will predominantly catalyze the pseudo-intramolecular A–B bond formation between affinity handle-linked substrate B and substrate A molecules on their own host cell. The intermolecular coupling of substrate B with substrate A molecules attached to other cells is entropically much less favorable, and therefore yeast cells displaying inactive enzymes should remain predominantly uncoupled to the affinity handle.

Following incubation with substrate B for the desired reaction time, cells are stained with a fluorescent molecule that binds the affinity handle (e.g., streptavidin-phycoerythrin (streptavidin-PE)). The most fluorescent cells, which encode the most active catalysts, are isolated by FACS. Up to 10^8 cells can be sorted in a two-hour period using modern FACS equipment. After sorting and growth

amplification, the recovered cells can be enriched through additional FACS steps, or DNA encoding active library members can be harvested and subjected to point mutagenesis or recombination before entering the next round of evolution.

We used a chemoenzymatic approach to link substrate A to cells rather than a non-specific chemical conjugation strategy to more reproducibly array the substrate on the cell surface and to avoid reagents that might alter the activity of library members. The *B. subtilis* Sfp phosphopantetheinyl transferase catalyzes the transfer of phosphopantetheine from coenzyme A (CoA) onto a specific serine side chain within an acyl carrier protein or peptide carrier protein. We chose Sfp to mediate substrate attachment because of its broad small-molecule substrate tolerance (3, 21) and its ability to efficiently conjugate phosphopantetheine derivatives to the 12-residue S6 peptide (22) (Figure S1). We speculated that the small size of the S6 peptide would allow it to be well-tolerated in the context of the Aga1p mating factor. Functionalized CoA derivatives can be readily prepared by reacting the free thiol of commercially available CoA (3, 21) with a commercially available maleimide-containing bifunctional crosslinker, followed by substrate A bearing a compatible functional group.

To integrate Sfp-catalyzed bioconjugation with yeast display required engineering a new yeast display vector and yeast strain (Figure S2). To create a handle for substrate attachment at the cell surface, we fused the S6 peptide onto the N-terminus of Aga1p and integrated this construct under the control of the strong, constitutive GPD promoter in the genome of *Saccharomyces cerevisiae* strain BJ5465 (19). We modified the Aga2p expression construct by inserting the recognition site for tobacco etch virus (TEV) protease between the hemagglutinin (HA) tag and the coding sequence of the protein of interest. Following incubation of the substrate-A-conjugated yeast library with substrate B, TEV protease digestion removes all library members from the surface, including any undesired enzymes that bind or react directly with substrate B, but do not catalyze A–B bond formation, thus removing a potential source of undesired background. The HA tag remains on the cell surface and enables staining

for enzyme display level using an anti-HA antibody. The ability to efficiently cleave enzymes from the yeast cell surface also facilitates enzyme characterization in a cell-free context.

Validation of the Yeast Display System

Sortase A (srtA) is a sequence-specific transpeptidase found in *Staphylococcus aureus* and other Gram positive bacteria. The *S. aureus* enzyme recognizes a LPXTG site (X = any amino acid), cleaves the scissile amide bond between threonine and glycine using a nucleophilic cysteine (C184), and resolves the resulting acyl-enzyme intermediate with oligoglycine-linked molecules to generate the fusion of the LPXT- and oligoglycine-linked peptides or proteins. Sortase A-catalyzed transpeptidation has emerged as a powerful tool for bioconjugation because of the enzyme's high specificity for the LPXTG motif and its extremely broad substrate tolerance outside of the recognition elements described above. Because the LPXTG and oligoglycine motifs can be flanked by virtually any biomolecule, sortase has been used to label proteins, generate nucleic acid-protein conjugates, and immobilize proteins onto solid supports (23). A significant limitation of srtA is the large quantities of the enzyme or long reaction times that are needed to overcome its poor reaction kinetics ($k_{cat}/K_{m\ LPETG}$ = 200 M^{-1} s^{-1} , Table 1). The evolution of a more active *S. aureus* srtA would therefore significantly enhance the utility and scope of this bond-forming reaction.

We first examined if yeast-displayed sortase enzymes in our system could catalyze the reaction between surface-immobilized LPETGG and exogenous biotinylated trigycine peptide (GGGYK-biotin). To conjugate cells to the LPETGG substrate, we incubated yeast displaying wild-type srtA and the S6 peptide with Sfp and coenzyme A-linked LPETGG (CoA-LPETGG, see Figure S3 and Supporting Information for synthesis details). The sortase-catalyzed reactions were initiated with the addition of GGGYK-biotin and 5 mM $CaCl_2$. After washing, the cells were stained with streptavidin-PE and an AlexaFluor488-conjugated anti-HA antibody to analyze the extent of reaction and enzyme display level, respectively, by flow cytometry. When yeast cells displaying wild-type sortase A (wt srtA-yeast)

were analyzed, the majority of the cells exhibited high levels of PE fluorescence, indicating substantial conjugation with GGGYK-biotin (Figure 2A). In contrast, wt srtA-yeast not conjugated to LPETGG, or LPETGG-conjugated yeast cells displaying the inactive C184A sortase mutant, exhibited only background levels of PE fluorescence after incubation with GGGYK-biotin, confirming that biotinylation was dependent both on sortase activity and on the presence of both substrates (Figure 2A).

To verify that enzymes displayed on the yeast cell surface catalyze pseudo-intramolecular reactions with substrate molecules immobilized on the same cell, we performed one round of model screening on mixtures of wt srtA-yeast and srtA C184A-yeast. Yeast cells were mixed in 1:100 and 1:1000 ratios of wt:C184A sortases. Each mixture of cells was coupled with CoA-LPETGG using Sfp, then incubated with 50 μM GGGYK-biotin for 15 minutes. Because srtA binds weakly to GGG (K_m = 140 μM, Table 1), washing with non-biotinylated GGG was sufficient to remove any background signal and TEV digestion was not performed after the reaction. After fluorophore staining, cells exhibiting both AlexaFluor488 and PE fluorescence were isolated by FACS (Figure 2B) and amplified by culturing to saturation. The plasmid DNA encoding survivors was harvested, and the compositions of the recovered genes were analyzed by restriction digestion with *Hind*III following PCR amplification. The wt srtA gene is distinguishable from C184A by the presence of an additional *Hind*III site (Figure 2C). In both model FACS sort experiments, we observed ≥ 6,000-fold enrichment of the wild-type gene from both mixtures that were predominantly the inactive C184A mutant (Figure 2C). Similarly high enrichment factors were also observed in model sortase screens in which GGG-modified cells were reacted with biotinylated LPETGG peptide, and in model biotin ligase (BirA) screens in which cells displaying a biotinylation substrate peptide and wild-type BirA were enriched in the presence of a large excess of cells displaying a less active BirA mutant (Figure S4). These results collectively suggest that this system can strongly enrich yeast displaying active bond-forming enzymes from mixtures containing predominantly yeast displaying inactive or less active enzyme variants.

Directed Evolution of Sortase A Enzymes with Improved Catalytic Activity

Next we sought to evolve *S. aureus* srtA for improved activity using the enzyme evolution strategy validated above. We focused on improving the poor LPXTG substrate recognition of srtA (K_m = 7.6 mM, Table 1), which limits the usefulness of sortase-catalyzed bioconjugation by requiring the use of high concentrations of enzyme (> 30 μM) or long reaction times to compensate for poor reaction kinetics at the micromolar concentrations of LPXTG substrate that are typically used. To direct evolutionary pressure to improve LPXTG recognition, we formatted the screen such that the triglycine substrate is immobilized on the cell surface along with the enzyme library, and the biotinylated LPETG peptide is added exogenously. This format enables evolutionary pressure for improved LPETG recognition to be increased simply by lowering the concentration of LPETG peptide provided during the sortase-catalyzed bond-forming reaction.

We randomly mutated the wt *S. aureus* srtA gene using PCR with mutagenic dNTP analogs (24) and cloned the resulting genes into the modified yeast display vector using gap repair homologous recombination to yield a library of ~10^8 transformants (round 0, R0). Each library member contained an average of two non-silent mutations. The library was subjected to four rounds of enrichment for sortase activity without any additional diversification between rounds. In each round we subjected control samples— cells displaying wt srtA, or the cells isolated from the previous round— to identical reaction conditions and screening protocols to precisely define FACS gates that captured cells with PE fluorescence corresponding to improved sortase activity (Figure S5). We applied increasing evolutionary pressure for improved LPETG recognition by decreasing the concentration of biotinylated LPETG substrate 10-fold with each successive round, starting from 100 μM in the first round and ending with 100 nM in the fourth round (Figure S6). We also increased evolutionary pressure for overall catalytic activity by accepting a smaller percentage of the most PE-fluorescent cells with each successive round, ranging from 1.4% in R1 to 0.15% in R4, and by shortening the reaction time in R4 from 60 to 15 minutes.

To preclude the evolution of specificity for a particular LPETG-containing sequence, we alternated using biotin-LPETGS (R1 and R3) and biotin-LPETGG (R2 and R4) peptides. After the fourth round of enrichment, surviving genes were subjected to *in vitro* homologous recombination using the NExT procedure (25) and re-cloned into yeast to yield a recombined and diversified library of ~10^8 transformants. The shuffled library (R4Shuf) was subjected to four additional rounds of sorting (resulting in R5, R6, R7, and R8), with the concentration of biotinylated LPETG peptide dropping from 100 nM to 10 nM in the final round (Figure S6).

We developed an assay to rapidly compare the activity of yeast-displayed sortase mutants. Yeast cells were incubated with TEV protease to release the enzymes from the cell surface into the surrounding supernatant. The reaction in the supernatant was initiated by the addition of the two peptide substrates, CoA-LPETGG and GGGYK-biotin. After 30 minutes of reaction, Sfp was added to the same reaction mixture to attach the biotinylated adduct and unreacted CoA-LPETGG onto the cell surface. We verified that the level of cell-surface fluorescence after streptavidin-PE staining is a direct reflection of the relative amount of biotinylated product in solution (Figure S7).

We evaluated the mean activity of the yeast pools recovered after each round of sorting using this assay. Over the course of the selections, we observed a steady increase in the extent of product formation catalyzed by the recovered sortase mutants. By the last round (R8) the activity signal was ~130-fold greater than that of the initial, unselected library (R0), and ~40-fold greater than that of wt srtA (Figure 3A, B). These observations suggested that the system had evolved sortase variants with substantially improved activities.

Characterization of Evolved Sortase Mutants

We used the above assay to evaluate the activity of individual clones from R4 and R8 together with wt srtA and the inactive C184A mutant (Figure 3B). All tested mutants from R4 exhibited improved activity relative to wild-type, with the two most active mutants, 4.2 and 4.3, showing ~20-

fold more activity than wt srtA. Mutants isolated from R8 exhibited even greater gains in activity, including four mutants that were ≥ 100-fold more active than wild-type srtA under the assay conditions (Figure 3B).

Sequences of evolved sortase genes revealed the predominance of P94S or P94R, D160N, D165A, and K196T mutations among R8 clones (Figure 4A, Figure S8B). Of the 16 unique sequences we isolated from R8, nine contained all four mutations. Thirteen of the 16 unique sequences contained at least three of the mutations, and all sequences contained at least two of the four mutations. All of these mutations also appeared in clones isolated from R4, but no clone from R4 contained more than two of the mutations, suggesting that recombination following R4 enabled combinations of mutations that persisted in rounds 4-8. Indeed, the highly enriched tetramutant combination appears to have arisen from recombination of two mutations each from clones 4.2 and 4.3, the two most active mutants isolated from R4. Gene shuffling was therefore an important component of the evolutionary strategy to generate genes encoding the most active sortase enzymes tested.

None of these four mutations have been reported in previous mutational studies studying the sortase active site and the molecular basis of LPETG substrate recognition (26, 27). To gain insight into how these mutations improve catalysis, we expressed and purified each sortase single mutant, clones 4.2 and 4.3, and the tetramutant from *E. coli*, and we measured the saturation kinetics of wt srtA and the mutants using an established HPLC assay (28). The observed kinetic parameters for the wild-type enzyme closely match those previously reported (26, 28). Each single mutation in isolation contributed a small beneficial effect on turnover (k_{cat}) and more significant beneficial effects on LPETG substrate recognition, lowering the $K_{m\ LPETG}$ up to three-fold (Table 1). The effects of the mutations in combination were largely additive. Compared to wild-type, 4.2 and 4.3 exhibited a 2.0-2.6-fold improvement in k_{cat} and a 5-7-fold reduction in $K_{m\ LPETG}$, resulting in a ~15-fold enhancement in catalytic efficiency at using the LPETG substrate (Table 1). Combining all four mutations yielded a sortase enzyme with a 140-fold improvement in its ability to convert LPETG ($k_{cat}/K_{m\ LPETG}$). This large

gain in catalytic efficiency is achieved primarily through 45-fold improved LPETG recognition accompanied by a 3-fold gain in k_{cat} (Table 1, Figures S9 and S10).

The effects of the individual mutations on LPETG substrate recognition can be rationalized in light of the reported solution structure of wt *S. aureus* srtA covalently bound to an LPAT peptide substrate (29). The mutated residues are all located at the surface of the enzyme, near the LPAT-binding groove (Figure 4B). P94 lies at the N-terminus of helix 1, and K196 lies at the C-terminus of the β7/β8 loop. Both D160 and D165 lie in the region connecting β6 and β7 that participates in LPETG substrate binding. D165 lies at the N-terminus of a 3_{10} helix that is formed only upon LPAT binding and makes contacts with the leucine residue of LPAT. The localization of the mutations within loops that line the LPAT binding groove suggests that they may be improving binding by altering the conformation of these important loops.

The evolved sortase mutants exhibit decreased GGG substrate binding (Table 1, Figures S9 and S10). Compared to wild-type, we measured a 30-fold increase in $K_{m\,GGG}$ for the sortase A tetramutant. P94S and D165A had larger detrimental effects on $K_{m\,GGG}$ than D160N and K196T. These results are consistent with mapping of the GGG-binding region proposed by NMR amide backbone chemical shift data. The chemical shifts of the visible amide hydrogen resonances for residues 92-97 and 165 were among the most perturbed upon binding of a Gly_3 peptide (29). Due to the absence of a high-resolution structure of the srtA- Gly_3 complex at this time, it is difficult to rationalize in more detail the basis of altered $K_{m\,GGG}$ among evolved mutants.

To recover some of the ability to bind the GGG substrate, we reverted A165 of the tetramutant back to the original aspartic acid residue found in wild-type because our results indicated that the D165A mutation was most detrimental for GGG recognition. Compared to the tetramutant, this P94S/D160N/K196T triple mutant exhibited a 2.6-fold improvement in $K_{m\,GGG}$, accompanied by a three-fold increase in $K_{m\,LPETG}$ and no change in k_{cat} (Table 1, Figures S9 and S10). We also subjected the R8 yeast pool to one additional round of screening (R9), immobilizing LPETGG on the cell surface

before reaction with 100 nM GGGYK-biotin. The P94S/D160N/K196T reversion mutant was recovered in two out of the 24 sequenced clones from R9, but a different triple mutant (P94S/D160N/D165A) dominated the R9 population after screening, representing 14/24 sequenced clones (Figure S8C). Compared to the tetramutant, the $K_{m\ GGG}$ of this mutant improved by 2.7-fold, whereas the k_{cat} and $K_{M\ LPETG}$ were not altered by more than a factor of 3-fold (Table 1).

We also performed mutagenesis and enrichment to identify additional mutations that improve GGG recognition in the tetramutant context. We combined four R8 clones as templates for additional diversification by PCR, and subjected the resulting yeast library (R8mut) to two rounds of screening, immobilizing LPETGG on the cell surface before reaction with 100-1000 nM GGGYK-biotin. After two rounds of enrichment, the K190E mutation originally observed in clone 4.2 was found in 56% of the unique sequenced clones in R10mut, and 33% of the clones possessed P94R in place of P94S (Figure S8D). The other three mutations of the tetramutant motif were found intact in 89% of the unique R10mut clones. We constructed the P94R/D160N/D165A/K190E/K196T pentamutant and assayed its activity. Compared to the tetramutant, the $K_{m\ GGG}$ of this mutant improved by 1.8-fold, whereas the k_{cat} and $K_{M\ LPETG}$ were not altered by more than a factor of 1.3-fold. Compared with wt srtA, this pentamutant has a 120-fold higher $k_{cat}/K_{m\ LPETG}$ and a 20-fold higher $K_{m\ GGG}$ (Table 1, Figures S9 and S10). To validate our enzyme kinetics measurements, we followed product formation over one hour and observed turnover numbers of greater than 10,000 per hour. The resulting data (Figure S11) yielded k_{cat} and $K_{M\ LPETG}$ values that closely agree with our kinetics measurements (Table 1). Collectively, these results indicate that relatives of the evolved tetramutant can exhibit partially restored GGG binding and therefore provide alternative enzymes for applications in which the GGG-linked substrate is available only in limited quantities.

Cell-Surface Labeling With an Evolved Sortase

The improved activities of the evolved sortase enzymes may enhance their utility in bioconjugation applications such as the site-specific labeling of LPETG-tagged proteins expressed on the surface of living cells. In these applications, the effective concentration of the LPETG peptide is typically limited to micromolar or lower levels by endogenous expression levels, and therefore the high $K_{M\ LPETG}$ of wt srtA ($K_{M\ LPETG}$ = 7.6 mM, Table 1) necessitates the use of a large excess of coupling partner and enzyme to drive the reaction to a reasonable yield. As it is typically straightforward to synthesize milligram quantities of short oligoglycine-linked probes using solid-phase peptide chemistry, we hypothesized that the much higher $k_{cat}/K_{m\ LPETG}$ of the evolved sortase enzymes might enable them to mediate cell-surfacing reactions that would be inefficient using the wild-type enzyme.

We expressed human CD154 tagged with the LPETG sequence at its C-terminus on the surface of HeLa cells and compared the labeling of the live cells with GGGYK-biotin using wt srtA and the evolved P94S/D160N/K196T mutant. After staining with a streptavidin-AlexaFluor594 conjugate, flow cytometry analysis revealed that the evolved sortase yielded ≥ 30-fold higher median fluorescence than the wild-type enzyme (Figure 5A). Although we used conditions similar to those used to label HEK293 cells using wt srtA for fluorescence microscopy (4), over four independent replicates, the wild-type enzyme did not result in fluorescence more than 2.8-fold higher than the background fluorescence of cells incubated in the absence of enzyme (Figure 5A). Consistent with the flow cytometry data, live-cell fluorescence microscopy confirmed very weak labeling by wt srtA and much more efficient labeling by the evolved sortase mutant (Figure 5B). Cells expressing CD154 without the LPETG tag were not labeled to a significant extent by the evolved sortase, indicating that the site-specificity of the enzyme has not been significantly compromised. Under the conditions tested, the evolved sortase triple, tetra-, and pentamutants all exhibit comparable and efficient cell-surface labeling, despite their differences in $K_{m\ GGG}$ (Figure S12). Collectively, our results suggest that the sortase variants evolved using the enzyme evolution system developed in this work are substantially

more effective than the wild-type enzyme at labeling LPETG-tagged proteins on the surface of live mammalian cells.

Discussion

We integrated yeast display, Sfp-catalyzed bioconjugation, and cell sorting into a general directed evolution strategy for enzymes that catalyze bond-forming reactions. We validated the system through model selections enriching for *S. aureus* sortase A-catalyzed transpeptidation activity, attaining enrichment factors greater than 6,000 after a single round of sorting. We applied this system to evolve sortase A for improved catalytic activity. After eight rounds of sorting with one intermediate gene shuffling step, we isolated variants of sortase A that contained four mutations that together resulted in a 140-fold increase in LPETG-coupling activity compared with the wild-type enzyme. An evolved sortase enabled much more efficient labeling of LPETG-tagged human CD154 expressed on the surface of HeLa cells compared with wild-type sortase.

The kinetic properties of the mutant sortases accurately reflect our screening strategy. The 50-fold decrease in $K_{m\ LPETG}$ of the tetramutant compared to wild-type is consistent with lowering the concentration of free biotinylated LPETG peptide during the reaction in successive rounds. Meanwhile, this screening format ensured that a high effective molarity of GGG was presented to each enzyme candidate over eight rounds of enrichment, which we estimated to be ~950 μM (see Supporting Information, Figure S13). It is therefore unsurprising that GGG recognition among evolved sortases drifted during evolution. Likewise, the three-fold increase in k_{cat} of the tetramutant compared to that of the wild-type enzyme may have resulted from screening pressures arising from shortening the reaction time in later rounds. Larger increases in k_{cat} may require modified selection or screening strategies that explicitly couple survival with multiple turnover kinetics, perhaps by integrating our system with *in vitro* compartmentalization.

Despite the widespread use of yeast display in the evolution of binding interactions (18), to the best of our knowledge, sortase A is only the third enzyme to be evolved using yeast display, in addition to horseradish peroxidase (30, 31) and an esterase catalytic antibody (32). Our results highlight the attractive features of yeast display that offer significant advantages for enzyme evolution, including quality control mechanisms within the secretory pathway that ensure display of properly folded proteins and compatibility with FACS (18). For these reasons, we used yeast as the vehicle for display instead of an M13 phage simultaneously displaying an Sfp peptide substrate and an enzyme library (33). As the method developed here does not rely on any particular screenable or selectable property of the substrates or product, it is in principle compatible with any bond-forming enzyme that can be expressed in yeast, including glycosylated proteins that are likely incompatible with phage and mRNA display, provided that linkage of the substrates to CoA and to the affinity handle is possible and tolerated by the enzyme or its evolved variants. In cases in which the enzyme accepts only one of these modifications, product-specific antibodies in principle could be used to detect bond formation. Furthermore, we note that integrating our yeast display system with the multicolor capabilities of FACS should enable the evolution of enzyme substrate specificity.

Beyond improving existing activities of natural proteins for research, industrial, and medicinal use, we speculate that the enzyme evolution strategy presented here will be valuable in the engineering of artificial proteins with new, tailor-made catalytic activities. The reactions catalyzed by natural enzymes are only a small subset of the diverse array of reactions known in organic chemistry, and a promising route to generating artificial enzymes is the computational design of a protein catalyst with arbitrary activity followed by optimization of its catalytic activity through directed evolution. Indeed, recent advances in computational protein design have created *de novo* catalysts for the retroaldol (34), Kemp elimination (35), and Diels-Alder reactions (36), and these successes demonstrate the feasibility of designing weakly active proteins that are ideal starting points for directed evolution. The integration

of computational design and a general enzyme evolution scheme such as the one presented here represents a promising strategy for creating highly active proteins with tailor-made catalytic activities.

Materials and Methods

See the Supporting Information for additional experimental methods.

Sortase evolution. A library of 7.8×10^7 mutant sortase genes containing an average of 2.0 amino acid changes per gene was introduced into yeast cells using gap repair homologous recombination (see the Supporting Information for details on library construction). In Round 1, 6×10^8 sortase library-expressing cells were conjugated to GGGK-CoA, incubated with 100 μM biotin-LPETGS for 60 minutes, and stained with streptavidin-PE and an AlexaFluor488-conjugated anti-HA antibody (Invitrogen). The top 1.4% of the PE/AlexaFluor488 double-positive population were isolated and grown to saturation. At least a ten-fold excess of cells relative to the number of cells recovered from sorting were removed, pelleted, and induced to display enzymes at the cell surface with galactose before entering the subsequent round of sorting. See Figure S6 for details on screening stringency. Following round 4, the surviving sortase genes were amplified by PCR and shuffled using the NeXT method (25) (see Supporting Information for details). The diversified gene library was introduced into yeast to generate a library of 6.9×10^7 transformants (see Supporting Information for details). Four additional rounds of enrichment were performed with GGG immobilized on the surface and biotinylated LPETG peptide provided exogenously. For rounds 9, 9mut, and 10mut, the cells from the previous round were modified with CoA-LPETGG in TBS-B with 5 mM $MgCl_2$ and 5 mM $CaCl_2$ for 30 minutes to facilitate formation of the acyl-enzyme intermediate, before washing and initiating the reaction with 0.1-1.0 μM GGGYK-biotin.

Mammalian cell labeling. HeLa cells were cultured at 37 °C in DMEM supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin under an atmosphere containing 5% CO_2. The

cells were transfected with a 9:1 ratio of plasmid pCDNA3-CD154-LPETG:cytoplasmic YFP expression plasmid (as a transfection marker). After 24 hours, the transfected cells were trypsinized, re-plated onto glass coverslips, and incubated overnight at 37 °C. Each coverslip was washed twice with Hank's balanced salt solution (HBSS) and immersed into HBSS supplemented with 1 mM GGGYK-biotin, 5 mM $CaCl_2$, and 100 μM enzyme. After 5 to 10 minutes, the coverslips were washed twice with PBS supplemented with 1% bovine serum albumin (BSA), 1 mM unmodified GGG, and 5 mM $MgSO_4$ before immersion into a solution of streptavidin-AlexaFluor594 (1:200, Invitrogen) in PBS with 1% BSA and 5 mM $MgSO_4$. For flow cytometry analysis, the coverslips were washed twice with PBS before incubation in PBS on ice for 30 minutes. Cells were resuspended and analyzed using a BD Fortessa flow cytometer. The AlexaFluor594 fluorescence of the top 16-25% most YFP-positive cells was recorded. For imaging, the coverslips were washed twice with PBS containing 5 mM $MgSO_4$ before analysis on a Perkin Elmer spinning disk confocal microscope (Harvard Center for Biological Imaging). Images were recorded using the DIC, YFP, and Alexa channels.

Acknowledgements

Research support was provided by NIH R01 GM065400 and the Howard Hughes Medical Institute. I.C. was supported by an NIH/NRSA Postdoctoral Fellowship and by DARPA HR0011-08-0085. B.M.D. is a Hertz Foundation Fellow. We thank Professor Hidde Ploegh for plasmid pCDNA3-CD154-LPETG, Professor Dane Wittrup for plasmid pCTCon2, and Professor Christopher T. Walsh for the bacterial expression plasmid for Sfp.

References Cited

1. Savile CK*, et al.* (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. (Translated from eng) *Science* 329(5989):305-309 (in eng).
2. Uttamapinant C*, et al.* (2010) A fluorophore ligase for site-specific protein labeling inside living cells. (Translated from eng) *Proc Natl Acad Sci U S A* 107(24):10914-10919 (in eng).
3. Yin J*, et al.* (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc.Natl.Acad.Sci.U.S.A* 102(44):15815-15820.
4. Popp MW, Antos JM, Grotenbreg GM, Spooner E, & Ploegh HL (2007) Sortagging: a versatile method for protein labeling. (Translated from eng) *Nat Chem Biol* 3(11):707-708 (in eng).
5. Walsh G (2006) Biopharmaceutical benchmarks 2006. (Translated from eng) *Nat Biotechnol* 24(7):769-776 (in eng).
6. Vellard M (2003) The enzyme as drug: application of enzymes as pharmaceuticals. (Translated from eng) *Curr Opin Biotechnol* 14(4):444-450 (in eng).
7. Cherry JR & Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. (Translated from eng) *Curr Opin Biotechnol* 14(4):438-443 (in eng).
8. Bershtein S & Tawfik DS (2008) Advances in laboratory evolution of enzymes. (Translated from eng) *Curr Opin Chem Biol* 12(2):151-158 (in eng).
9. Bloom JD*, et al.* (2005) Evolving strategies for enzyme engineering. (Translated from eng) *Curr Opin Struct Biol* 15(4):447-452 (in eng).
10. Turner NJ (2003) Directed evolution of enzymes for applied biocatalysis. (Translated from eng) *Trends Biotechnol* 21(11):474-478 (in eng).
11. Neuenschwander M, Butz M, Heintz C, Kast P, & Hilvert D (2007) A simple selection strategy for evolving highly efficient enzymes. (Translated from eng) *Nat Biotechnol* 25(10):1145-1147 (in eng).
12. van Sint Fiet S, van Beilen JB, & Witholt B (2006) Selection of biocatalysts for chemical synthesis. (Translated from eng) *Proc Natl Acad Sci U S A* 103(6):1693-1698 (in eng).
13. Kelly BT, Baret JC, Taly V, & Griffiths AD (2007) Miniaturizing chemistry and biology in microdroplets. (Translated from eng) *Chem Commun (Camb)* (18):1773-1788 (in eng).
14. Lin H, Tao H, & Cornish VW (2004) Directed evolution of a glycosynthase via chemical complementation. (Translated from eng) *J Am Chem Soc* 126(46):15051-15059 (in eng).
15. Leconte AM, Chen L, & Romesberg FE (2005) Polymerase evolution: efforts toward expansion of the genetic code. (Translated from eng) *J Am Chem Soc* 127(36):12470-12471 (in eng).
16. Seelig B & Szostak JW (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* 448(7155):828-831.
17. Olsen MJ*, et al.* (2000) Function-based isolation of novel enzymes from a large library. *Nat.Biotechnol.* 18(10):1071-1074.
18. Gai SA & Wittrup KD (2007) Yeast surface display for protein engineering and characterization. *Curr.Opin.Struct.Biol.* 17(4):467-473.
19. Boder ET & Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat.Biotechnol.* 15(6):553-557.
20. Varadarajan N, Rodriguez S, Hwang BY, Georgiou G, & Iverson BL (2008) Highly active and selective endopeptidases with programmed substrate specificities. (Translated from eng) *Nat Chem Biol* 4(5):290-294 (in eng).
21. Yin J, Liu F, Li X, & Walsh CT (2004) Labeling proteins with small molecules by site-specific posttranslational modification. (Translated from eng) *J Am Chem Soc* 126(25):7754-7755 (in eng).

22. Zhou Z*, et al.* (2007) Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. (Translated from eng) *ACS Chem Biol* 2(5):337-346 (in eng).
23. Tsukiji S & Nagamune T (2009) Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. (Translated from eng) *Chembiochem* 10(5):787-798 (in eng).
24. Zaccolo M, Williams DM, Brown DM, & Gherardi E (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. (Translated from eng) *J Mol Biol* 255(4):589-603 (in eng).
25. Muller KM*, et al.* (2005) Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. (Translated from eng) *Nucleic Acids Res* 33(13):e117 (in eng).
26. Bentley ML, Lamb EC, & McCafferty DG (2008) Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from Staphylococcus aureus. (Translated from eng) *J Biol Chem* 283(21):14762-14771 (in eng).
27. Frankel BA, Tong Y, Bentley ML, Fitzgerald MC, & McCafferty DG (2007) Mutational analysis of active site residues in the Staphylococcus aureus transpeptidase SrtA. *Biochemistry* 46(24):7269-7278.
28. Kruger RG, Dostal P, & McCafferty DG (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the Staphylococcus aureus sortase transpeptidase SrtA. (Translated from eng) *Anal Biochem* 326(1):42-48 (in eng).
29. Suree N*, et al.* (2009) The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. (Translated from eng) *J Biol Chem* 284(36):24465-24477 (in eng).
30. Agresti JJ*, et al.* (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. (Translated from eng) *Proc Natl Acad Sci U S A* 107(9):4004-4009 (in eng).
31. Antipov E, Cho AE, Wittrup KD, & Klibanov AM (2008) Highly L and D enantioselective variants of horseradish peroxidase discovered by an ultrahigh-throughput selection method. (Translated from eng) *Proc Natl Acad Sci U S A* 105(46):17694-17699 (in eng).
32. Yang G & Withers SG (2009) Ultrahigh-throughput FACS-based screening for directed enzyme evolution. (Translated from eng) *Chembiochem* 10(17):2704-2715 (in eng).
33. Sunbul M, Marshall NJ, Zou Y, Zhang K, & Yin J (2009) Catalytic turnover-based phage selection for engineering the substrate specificity of Sfp phosphopantetheinyl transferase. (Translated from eng) *J Mol Biol* 387(4):883-898 (in eng).
34. Jiang L*, et al.* (2008) De novo computational design of retro-aldol enzymes. *Science* 319(5868):1387-1391.
35. Rothlisberger D*, et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190-195.
36. Siegel JB*, et al.* (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. (Translated from eng) *Science* 329(5989):309-313 (in eng).

Figure 1. A general strategy for the evolution of bond-forming catalysts using yeast display.

Figure 2. Validation of the enzyme evolution strategy. (A) FACS histogram of the reaction between cell surface-conjugated LPETGG and free GGGYK-biotin catalyzed by yeast-displayed wild-type *S. aureus* sortase A (wt srtA). Cells were stained with streptavidin-PE and an AlexaFluor488-anti-HA antibody. Negative control reactions with either the inactive C184A srtA mutant or without LPETGG are shown. (B) Dot plots comparing PE fluorescence (extent of reaction) vs. AlexaFluor488 fluorescence (display level) for two model screens. Mixtures of cells displaying either wt srtA or the inactive C184A srtA (1:1000 and 1:100 wt:C184A) were processed as in (A), then analyzed by FACS. Cells within the specified gate (black polygon) were collected. (C) Model screening results. Gene compositions before and after sorting were compared following *Hind*III digestion, revealing strong enrichment for active sortase.

Figure 3. Activity assays of mutant sortases. (A) Yeast pools recovered from the sorts were treated with TEV protease, and the cleaved enzymes were assayed for their ability to catalyze coupling between 5 μM CoA-LPETGG and 25 μM GGGYK-biotin. (B) Yeast cells expressing select individual clones were treated as described above. Error bars represent the standard deviation of three independent experiments.

Figure 4. Mutations in evolved sortases. (A) Highly enriched mutations are highlighted in black; other mutations are shown in blue. (B) Mapping evolved mutations on the solution structure of wild-type *S. aureus* sortase A covalently bound to its Cbz-LPAT substrate. The calcium ion is shown in blue, the LPAT peptide is colored cyan with red labels, and the side chains of amino acids that are mutated are in orange. The N-terminal Cbz group is shown in stick form in cyan.

Table 1. Kinetic characterization of mutant sortases. Kinetic parameters k_{cat} and K_m were obtained from fitting initial reaction rates at 22.5 °C to the Michaelis-Menten equation. Errors represent the standard deviation of three independent experiments.

Figure 5. Cell-surface labeling with wild-type and mutant sortases. Live HeLa cells expressing human CD154 conjugated at its extracellular C-terminus to LPETG were incubated with 1 mM GGGYK-biotin and no sortase A (srtA), 100 μM wild-type srtA, or 100 μM P94S/D160N/K196T srtA. The cells were stained with AlexaFluor-conjugated streptavidin. (A) Flow cytometry analysis comparing cell labeling with wild-type sortase (blue) and the mutant sortase (red). Negative control reactions omitting sortase (black) or LPETG (green) are shown. (B) Live-cell confocal fluorescence microscopy images of cells. The YFP (transfection marker) and Alexa (cell labeling) channels are shown.