Construction and flow cytometric screening of targeted enzyme libraries

Navin Varadarajan[1,2], Jason R Cantor[2], George Georgiou[1,2] & Brent L Iverson[1,3]

[1]Institute for Cell and Molecular Biology, University of Texas, Austin, Texas, USA. [2]Department of Chemical Engineering, University of Texas, Austin, Texas, USA. [3]Department of Chemistry and Biochemistry, University of Texas, Austin, Texas, USA. Correspondence should be addressed to B.L.I. (biverson@mail.utexas.edu) or G.G. (gg@che.utexas.edu).

Published online 21 May 2009; doi:10.1038/nprot.2009.60

Herein, we describe a methodology for the construction of targeted libraries intended to modify the substrate specificity of proteases expressed on the cell surface of *Escherichia coli*. The native outer membrane protease, OmpT, is used as a model system. The protocol relies on gene assembly using oligonucleotides and is easily adaptable to any enzyme in which information is available on the putative active site residues. Increasingly complex libraries can be generated in a systematic manner and screened using flow cytometry (fluorescence-activated cell sorting, FACS) for variants displaying altered function. Furthermore, if the substrate-binding pockets have not been elucidated, a protocol for partial multi-site saturation library construction is presented that allows for sampling a large number of residues, while maintaining an appropriate level of protein function. The entire procedure, from start to finish, should take approximately 2–3 weeks.

INTRODUCTION

The engineering and design of enzymes is a powerful process that can be applied not only for the generation of custom catalysts for biotechnological and medical applications but also to investigate protein mechanisms and, perhaps, even natural molecular evolution pathways from a molecular and biological standpoint[1,2]. In spite of the recent progress in the *de novo* design of enzymes[3,4], most protein engineering efforts are based on either rational or combinatorial redesign of existing enzyme scaffolds. Rational redesign, guided by knowledge of three-dimensional structure, homology modeling or experimental data, involves the construction of a small number of designed variants that are experimentally validated for new function[5]. Combinatorial mutagenesis followed by screening/selection, however, is the preferred approach when dramatic increases or changes in activity/reactivity are desired[2,6].

There are two basic strategies for the diversification of protein sequences at the genetic level: single gene manipulation techniques (such as random or targeted mutagenesis)[6] and recombination-based techniques (such as DNA shuffling[7] or homology independent crossovers[8]). Random mutagenesis remains the simplest strategy for the construction of mutant libraries, as it requires no prior structural or mechanistic information[9]. There are many reports of successful enzyme engineering by random mutagenesis, generally using error-prone PCR approaches, and it is often the method of choice for improving preexisting activities or physical properties of enzymes[2,10]. However, the introduction of a new functionality, for example, engineering reactivity toward novel substrates for which the wild-type (WT) enzyme has no latent activity, is a more challenging task that requires major sequence alteration and might necessitate an epistatic combination of multiple mutations among residues residing in or around the active site[11,12]. An epistatic combination refers to mutated residues that function synergistically to improve desired new activity when combined in the same enzyme variant, whereas those same individual mutations by themselves do not confer any advantage, or may even be deleterious. The discovery of useful epistatic combinations through random mutagenesis is improbable given the stochastic rarity of any particular set of mutational groupings within a generated library[13]. In addition, error-prone PCR generally involves changing only a single base per modified codon, that is, there are on average only 4–6 possible changes at the amino-acid level per codon. Thus, vast regions of functional sequence space cannot be accessed using random mutagenesis techniques based on error-prone PCR, reinforcing the need to use a targeted approach when dramatic changes in protein activity are desired.

Library construction through targeted mutagenesis[14] generally uses structural or other information regarding the identity of enzyme active site residues to randomize substrate-binding space in a more systematic manner. When low- or medium-throughput enzyme activity screens are available, it is appropriate to carry out sequential saturation of an individual, or a small collection of, desired residue(s), followed by the combination of beneficial substitutions in an additive manner[15,16]. Second- and third-shell residues, often known to affect enzyme function[15], can then be optimized in a second step using random mutagenesis that targets the entire protein sequence[12]. However, similar to using error-prone PCR alone, such an approach is inherently limited in its ability to identify epistatic combinations of beneficial mutations that may be essential when drastic changes in activity are sought.

When a high-throughput activity screen or a selection such as flow cytometry (fluorescence-activated cell sorting, FACS) is available ($>1 \times 10^6$ clones screened per round), a comprehensive library generation technique is required to take full advantage of the large number of library members being analyzed. This protocol, based on our recently published method[17], describes the construction of enzyme mutant libraries using gene assembly of short 35–50 mer oligonucleotides in which degenerate NNS (N = A/T/G/C, S = G/C) codons can be substituted for specific residues of interest across the entire gene. This library construction technique is easily adaptable toward the evolution of enzymes in which some knowledge of the active site location is available. Note that unlike error-prone PCR-based methods, using degenerate NNS codons provides access to all possible amino acids at the targeted residues of interest.

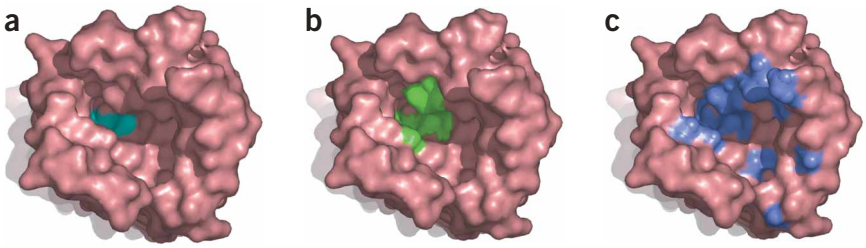

Figure 1 | The concept of expanding zone to create libraries of increasing complexity. An initial focused library is screened for the desired enzyme variant (**a**). In the absence of winners, a library targeting a greater subset of substrate-binding residues is constructed (**b**) and screened. If the second library fails to yield enzyme variants with the desired functionality, the set of residues that comprise the entire substrate-binding sites are targeted using partial site-saturation mutagenesis (**c**)[19]. Shown are the ray trace images of OmpT (PDB:Ii78[21]) generated using PyMol (DeLano, W.L. *The PyMOL Molecular Graphics System*. DeLano Scientific, Palo Alto, CA, USA (2002)).

Using a gene assembly approach to construct variant libraries has important enabling advantages when relatively large libraries can be screened for enhanced activity. First, if highly focused libraries fail to yield improved variants, it is suggested to incorporate additional degenerate codons at any given residue by re-designing and subsequently swapping the appropriate oligonucleotides from the original set. Thus, a systematic increase in complexity of libraries can be achieved in a relatively rapid manner. Furthermore, our protocol allows integration of a higher number of degenerate codons relative to that reported in other targeted mutagenesis studies[14]. As a result, the effects of increasing the number of degenerate codons, and the amino acids they encode, within a given library can be evaluated to determine whether a desired enzymatic activity change can be achieved through expansion of the mutational load around the active site (**Fig. 1**). Our protocol also allows one to create partial saturation libraries by doping in a mixture of oligonucleotides containing both the WT codon and the degenerate codon for a specific residue of interest during gene assembly[18]. In this way, large numbers of residues can be selected for saturation mutagenesis, but for any given library member, only a relatively small number of mutations will be present[19]. This provides for the comprehensive exploration of sequence space while minimizing the mutational load per gene, thereby decreasing the fraction of unfolded/inactive variants.

The following protocol also briefly describes the high-throughput flow cytometric dual substrate sorting/screening procedure for the isolation of rare enzyme variants from randomized libraries. Although a tiered selection/counter-selection system has been reported for other enzyme screens[20], we show the simultaneous application of a selection/counter-selection flow cytometric assay that can be readily expanded to include multiple counter-selections. Using simultaneous selection/counter-selection screening enables the efficient identification of enhanced enzyme variants with high levels of new activity while ensuring overall specificity for this new activity. Such specificity can be a critical enabling feature of enzymes being sought from large libraries; for example, in the case of the *Escherichia coli*, OmpT protease serves as the model system for the following protocols[12,17].

Comparison with other targeted library construction protocols

The protocol outlined here is one of several tools available to protein engineers for the construction and screening of targeted libraries toward improving/modifying protein function. Iterative saturation mutagenesis uses the Stratagene QuikChange strategy to generate small libraries ideally suited for low to medium throughput screening. Briefly, degenerate primers for 2–3 codons are purchased high-performance liquid chromatography (HPLC/PAGE) (polyacrylamide gel electrophoresis)-purified, and are used as templates for PCR in which the entire plasmid is extended using these primers[15]. The advantage of the iterative saturation mutagenesis protocol is that it is technically straightforward to implement, as it does not involve molecular cloning. The drawback, however, is that it is suited only for the construction of small libraries (library sizes 10,000–30,000 members). The library construction protocol outlined in this protocol can routinely generate library sizes of 10^8 members and is more suitable, as described earlier, when large changes in function (including uncovering epistatic interactions) using high-throughput screens is the goal.

A more recent approach for the generation of targeted libraries is ISOR (Incorporating Synthetic Oligonucleotides via Gene Reassembly)[18], which generates WT gene fragments by controlled digestion of a template gene using DNAse, whereas the oligo-mediated gene assembly reported here uses synthetic oligonucleotides at all positions in the gene. Both incorporate library diversity by doping oligonucleotides with degenerate codons at chosen positions. Our oligo-mediated gene assembly approach offers precise control over the ratios of WT to randomized oligos during gene construction, providing for accurate reproducibility and, therefore, efficient optimization. In addition, when the entire gene is derived from synthetic oligonucleotides, all of the codons can be simultaneously optimized for heterologous expression in a bacterial host.

Limitations of the current protocol

The protocol uses HPLC-purified primers for the construction of libraries, which can be costly. However, if it is important to reduce costs, standard desalted primers can be used instead of purified primers in an initial step to construct the codon-optimized gene. (This typically yields 5–20% of the correct sequence without insertions/deletions.) Subsequently, libraries can be constructed by spiking the degenerate oligonucleotides into the template using the codon-optimized gene-based targeted library construction protocols.

Overview of the procedure

The following procedure has been successfully applied to the engineering and isolation of a family of OmpT variants that recognize a diverse array of peptide sequences with superior selectivity as well as with high catalytic efficiency[12,17]. A flowchart depicting a general overview of the procedure is shown in **Figure 2**.

OmpT, an outer membrane protease from *E. coli*, serves as a useful model to explore enzyme library construction and screening

protocols. Proteases are notoriously difficult to engineer, largely because any new substrate specificity needs to be precise, in other words highly selective, to avoid self-cleavage or the cleavage of critical components of host organisms. The following protocol is divided into two steps: library construction and subsequent flow cytometric screening. The initial steps describe the assembly of variant genes designed to randomize three codons using PCR-based gene assembly (Step 1A(i–v)). The initial PCR product is then re-amplified using oligonucleotides containing the appropriate restriction sites to enrich full-length variants (Step 1A(vi–viii)). The purified DNA is then directionally cloned into an appropriate surface display vector (Step 1A (ix–xii)). The ligated DNA is transformed into *E. coli* MC1061, plated, and plasmid is then recovered from scraped cells (Step 1A(xiii–xx)). The plasmid library is then retransformed into the screening strain *E. coli* BL21(DE3) (Step 1A(xxi–xxiii)). The protocol also describes the construction of partial site-saturation libraries (Step 1B(i–vi)). Finally, a procedure for the two-color flow cytometric screening of surface displayed variant libraries is described (Steps 2–15).

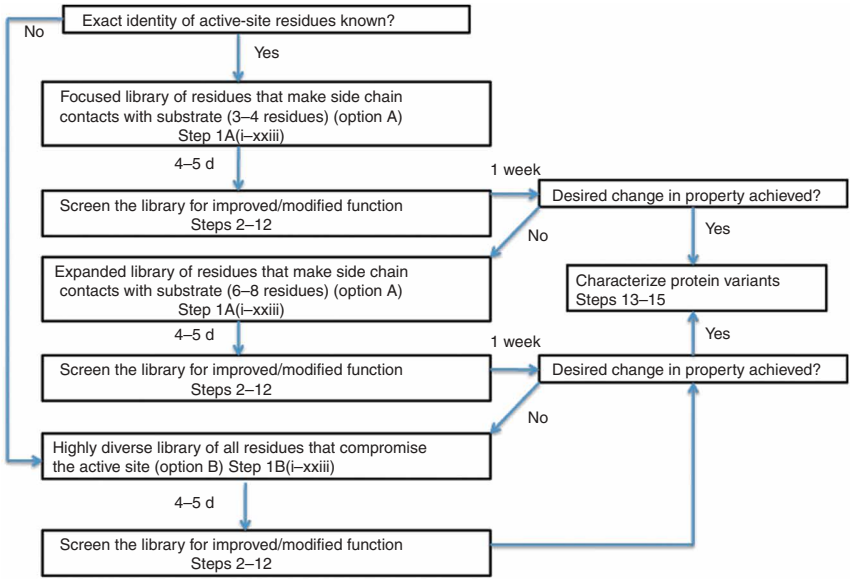

Figure 2 | Flowchart illustrating the typical workflow in the construction and screening of targeted enzyme libraries.

Experimental design

Primer design. Careful design of primers is critical to the success of high-quality library construction. The use of free, online software programs such as DNAWorks (http://mcl1.ncifcrf.gov/dnaworks/) is recommended and allows for codon optimization to facilitate recombinant expression in *E. coli*. The primers are typically 35–45 oligonucleotides long and are designed to minimize coincidental complementarity or hairpins. (Overlaps between any given forward and reverse primer pair is typically 15–20 bases.) Care must be taken to avoid gaps in the sequence, in which every single base of the gene in both forward and reverse directions must be encoded by oligos (example of the fragmentation for OmpT is shown in **Supplementary Figs. 1** and **2** online). The oligos are typically purchased HPLC-purified, as this significantly decreases the number of deletions and consequently increases the percentage of full-length functional variants in the library.

Degenerate oligos. Once the codons to be targeted have been chosen, the oligos that encode the targeted codons (in both forward and reverse directions) are identified. (We typically choose the degenerate codons NNS or NNK (K = G/T), but other triplets can be used to bias the library.) If complete randomization of targeted residues in the library is sought, it is necessary to purchase purified, randomized oligos that comprise the targeted codon(s) in both forward and reverse directions (e.g., refer to **Supplementary Fig. 1**). In the event that partial randomization is desired, a single (per codon) purified, degenerate oligo is sufficient.

Active site residues. The identification of active site residues that can be randomized during library construction is greatly aided by the availability of a three-dimensional structure (X-ray/NMR/homology model). If a structure of the substrate (analog)-bound enzyme is available, an initial focused set of residues (3–4) that make side chain contacts (Van der Waals/ion pair/hydrogen bonds) with substrate are identified using a visualization program such as Swiss-Pdb viewer. If, however, only a structure of the uncomplexed enzyme is available, a combination of the structure and biochemical data is used to define the initial set. For example, in the case of OmpT, the crystal structure (PDB 1i78)[21] was combined with biochemical data and screening experiments to identify putative S1 substrate-binding residues that make ion pair/hydrogen-bonding contacts with the guanidnium group of the substrate arginine (Glu27, Asp208 and Ser223). Screening a focused library that randomized these three positions allowed the isolation of an enzyme that could selectively cleave substrates with Glu at P1 (WT OmpT recognizes Arg at the P1 substrate position). It is important to note that although the correct identification of the substrate-binding residues in the enzymes reduces the number of screening experiments, it is by no means essential for the overall success of the experiment. If the initial screening experiment is not successful for the desired change in function, an expanded library that randomizes 5–7 residues making substrate contacts is targeted, thus improving the chances of success. Should neither of these libraries prove to be fruitful, or in the absence of a 3-D structure, the partial randomization library based on the entire putative active site residues should help identify enzyme variants with modified function. Importantly, enzyme variants identified from the screening of the partial randomization library can help identify residues for creating more focused libraries in a reverse strategy. Depending on the amount of structural/biochemical data available for a particular enzyme, an expanding zone strategy as reported by us for engineering OmpT specificity[17], or a reverse-zone strategy that allows the results of the screening experiment to identify residues, can be used to engineer enzyme variants with modified function.

Mutational load. An important variable in any library construction experiment is the number of mutations per gene construct in

the library. Achieving an optimal balance is crucial to sampling sequence space while still retaining protein folding/function; hence, a library construction protocol must allow precise control over the number of mutations/gene in constructed libraries. In the oligo-mediated gene assembly protocol, the mutational load per gene can be systematically varied by doping in different ratios of the randomized/WT oligos. For example, in the case of OmpT, when 21 active site residues are to be targeted, a 9:1 randomized/WT oligo molar ratio (across all 21 residues) yielded an average of 10 mutations per gene construct, whereas a 1:1 ratio of the same oligos yielded an average of 4 mutations per gene[17]. This fine-tuning of the mutational frequency can be used to adapt the library construction parameters to suit each enzyme being mutagenized.

Selection and counter-selection substrates. Electrostatic interactions between the negatively charged bacterial surface and substrates are exploited to capture the fluorescent and positively charged products of the enzymatic cleavage reaction on the surface of the bacteria[17]. The selection substrate comprises a BODIPY (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) fluorophore (recommended, although alternate fluorophores that show efficient excitation using the 488-nm laser typically found in flow cytometers can be used) and positively charged groups on one side of the putative cleavage site and a quencher (QSY7 recommended) on the other side. Cleavage by surface-displayed enzyme results in release of the quencher moiety, thereby enhancing BODIPY fluorescence. The fluorescent product, which has a +3 overall charge, is electrostatically captured on the negatively charged *E. coli* surface. The counter-selection substrate has a positive charge and a tetramethylrhodamine fluorophore on one side of a putative cleavage site (the activity being screened against, typically parental activity) and only negative charge on the other side (no quencher). Upon cleavage, the negative charge is released, giving rise to an overall positively charged fluorescent product that is again captured on the *E. coli* surface. Flow cytometry is used to isolate variants exhibiting high BODIPY (green) fluorescence and little to no tetramethylrhodamine (red) fluorescence. A number of examples of the design, synthesis and implementation of selection/counter-selection substrates can be found as supporting information online on the *Nature Chemical Biology* website[17].

Flow cytometry controls. The following controls should be included, where possible.

- *Negative control.* Cells expressing an inactive form of the enzyme or cells expressing no/irrelevant enzyme are suitable negative controls. These cells are grown under conditions identical to cells expressing the library and are also labeled with both selection and counter-selection substrates. Histograms displaying gated populations of *E. coli* cells (Gate set on forward and side scatter) should be negative in both fluorescent channels (mean: 3–10, depending on the concentration of substrate being used).
- *Positive control.* In typical library-based enzyme screening experiments, a true positive control that displays a functional profile being sought is hard to generate. It is essential to label cells expressing WT enzyme with both selection and counter-selection substrates. Depending on the actual enzyme and the functionality being screened for, histograms for the selection substrate show little to moderate activity (1- to 3-fold over negative control), whereas the histograms for the counter-selection substrate show high activity (5- to 20-fold over negative control).
- *Compensation.* When performing multi-color flow cytometry, it is important to set compensations to account for the spectral overlap of the fluorophores into the different filters being used. An excellent online resource for multi-color flow cytometry and compensations can be found on the BD Biosciences website.

Cell surface display. A recent review covers most of the display technologies available for the surface display of recombinant proteins[22]. The actual choice of plasmid and display system has to be determined and optimized for the protein of interest.

MATERIALS

REAGENTS

- *Escherichia coli strains:*
  - MC1061 (*araD139* Δ (*ara-leu*) 7696 *Δlac174 galU galK hsr⁻hsm⁺strA*^{R}) (available on request) (see **Box 1** for preparation of electrocompetent MC1061)
  - BL21(DE3) (Invitrogen) (available on request)
- Plasmid pDMLE19 for expression of WT OmpT under the control of its native promoter (available on request)
- VentR DNA Polymerase (New England Biolabs, cat. no. M0254S)
- Taq DNA Polymerase with ThermoPol Buffer (New England Biolabs, cat. no. M0267S)
- Deoxynucleotide Solution Mix (dNTPs) (New England Biolabs, cat. no. N0447)
- Restriction enzymes (New England Biolabs): *Eco*RI (cat. no. R0101L), *Hin*dIII (cat. no. R0104S)
- T_4 DNA Ligase (New England Biolabs, cat. no. M0202S)
- Agarose (Sigma, cat. no. A0169)
- Ampicillin (EMD Biosciences, cat. no. 171254; see REAGENT SETUP)
- Glycerol (Sigma-Aldrich, cat. no. G7893-1L; see REAGENT SETUP)
- Difco Luria–Bertani (LB) (Becton Dickinson, cat. no. 244620)
- Difco 2xYT (Becton Dickinson, cat. no. 244020)
- Difco SOB Medium (Becton Dickinson, cat. no. 244310)
- SOC Medium (Invitrogen, cat. no. 15544-034)
- PBS (see REAGENT SETUP)
- Screening substrates (REAGENT SETUP)
- 1% (wt/vol) sucrose solution (see REAGENT SETUP)
- Primers (HPLC-purified; Integrated DNA Technologies): gene-specific—WT and degenerate (see EXPERIMENTAL DESIGN)

EQUIPMENT

- Nitrocellulose Desalting Membrane, 0.025 μm (Millipore, cat. no. VSWP00010)
- Electroporation Cuvettes, 0.1-cm gap (Fisher Scientific, cat. no. FB102)
- BD Falcon Biodish XL 245 mm × 245 mm petri dishes (Becton Dickinson, cat. no. 351040)
- 100 mm × 15 mm Petri dishes (VWR, cat. no. 25384-342)
- Zymoclean Gel DNA Recovery Kit (Zymo Research, cat. no. D4001)
- QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27104)
- Sorting flow cytometer with appropriate lasers and detectors
- Gel electropheresis system
- Incubator 30 °C, 37 °C
- Shaker 25 °C, 30 °C and 37 °C, 250 r.p.m.
- Gene Pulser electroporation apparatus
- Thermal Cycler (MJ Research)
- 12 × 75 mm sterile culture tubes for FACS (VWR, cat. no. 60818-292)

REAGENT SETUP

Ampicillin solution Dissolve ampicillin powder at 200 mg ml^{-1} in distilled deionized H_2O. Filter through a 0.2-μm filter. Aliquot in 1 ml portions and store at −20 °C (can be stored for 1 year). Thawed aliquots should be freshly diluted 1,000-fold into liquid media or media with agar.

BOX 1 | PROTOCOL FOR ELECTROCOMPETENT MC1061

1. Pick a single selected colony and inoculate in 3 ml of LB medium and grow for 12 h at 37 °C at 250 r.p.m. in a shaker.
2. Inoculate 2 × 0.3 liter of LB medium with 300 µl of the fresh overnight culture.
3. Grow cells at 37 °C at 250 r.p.m. to an OD_{600} of ~0.4 (2.5–3 h).
4. Chill the cells on ice for 30 min and then centrifuge the cells at 6,370*g* (4 °C) for 15 min in two separate 500-ml centrifuge bottles.
5. Decant the supernatant and resuspend each pellet separately in 0.4-liter ice-cold sterile distilled deionized H_2O.
6. Centrifuge as in step 4, discard the supernatant and resuspend each pellet separately in 0.4 liter of 10% (vol/vol) ice-cold sterile glycerol.
7. Centrifuge as in step 4, discard the supernatant, combine the pellets and resuspend in 0.4 liter of 10% (vol/vol) ice-cold sterile glycerol.
8. Centrifuge as in step 4, and resuspend the pellet in 1 ml of 10% (vol/vol) ice-cold sterile glycerol.

The efficiency of the cells should be ~5×10^{10} cfu μg^{-1} DNA when 40 µl of electrocompetent cells is transformed using 10 pg of supercoiled pUC18. Aliquot (100 µl) in 1.5-ml sterile cryogenic vials and store at −80 °C until required (can be stored for 1 year).

2xYTAmp agar plates Dissolve 31 g of Difco 2xYT powder and 15 g of agar in distilled H_2O to a volume of 1 liter and autoclave. Cool the agar mixture by stirring until below 50 °C, add 1 ml of 200 mg ml^{-1} ampicillin solution, mix and pour in the plates. Plates can be stored for up to 4 months at 4 °C.

Glycerol solution 50% (vol/vol) of glycerol in sterile distilled deionized H_2O. Autoclave and store at 4 °C indefinitely.

PBS buffer 8.1 mM Na_2HPO_4, 1.47 mM KH_2PO_4, 2.68 mM KCl, 137 mM NaCl in distilled deionized H_2O. Adjust to pH 7.4 and autoclave (can be stored at room temperature (22 °C–25 °C) indefinitely).

1% (wt/vol) sucrose Dissolve 1 g of sucrose in 100 ml of deionized H_2O and sterile filter by a 0.2-µm filter (can be stored at room temperature (22 °C–25 °C) for 1 week).

Substrate synthesis Detailed protocols for the conjugation of dyes/fluorophores to peptides are described elsewhere[12,17]. Briefly, for conjugation to cysteines, a solution of the peptide in Na_2HPO_4 (pH = 8) is reacted with a molar equivalent of the thiol reactive form (maleimide/iodoacetamide) of the dye in DMF (*N*,*N*-dimethyl formamide), reverse-phase fast performance liquid chromatography (FPLC) purified and lyophilized. For the conjugation of lysines, a solution of the peptide/cysteine-conjugated dye in Na_2HPO_4 (pH = 8) is reacted with a molar equivalent of the amine reactive form of the dye (*N*-hydroxy sucinimidyl ester, NHS) in DMF and 25 molar equivalents of 4-(dimethylamino) pyridine. The reaction is quenched, reverse-phase FPLC purified and lyophilized. The end product is resuspended in deionized H_2O and product identity confirmed using electron spray ionization mass spectrometry (ESI-MS).

EQUIPMENT SETUP

Flow cytometer startup procedures Flow cytometer startup procedures must be carried out according to the manufacturer's instructions. Briefly, the steps involved in startup are

- fill sheath tank with sterile PBS (sheath fluid);
- empty waste;
- turn the machine on and login to FACS software;
- wait for instrument to connect;
- run fluidics startup and align stream;
- turn on the stream and wait ~15min for it to stabilize;
- Adjust amplitude and frequency to minimize satellites and ensure satellites have merged with the final visible drop in the stream window.

PROCEDURE

Library construction

1| Libraries can be constructed in one of two ways: use Option A if the exact identity of the desired substrate-binding pocket residues are known (e.g., in the case of OmpT, this would generate a three-member saturation library; Glu27, Asp208 and Ser223) and use Option B to target all of the active site residues using partial site-saturation mutagenesis (e.g., in the case of OmpT, a library partially randomized across 21 residues: 27, 29, 39, 40, 42, 44, 81, 87, 97, 101, 148, 150, 153, 163, 170, 208, 221, 263, 265, 280 and 282).

(A) Multiple-site saturation libraries ● TIMING **4–5 d**

(i) Design a set of overlapping ~40-mer oligonucleotides (oligos), both forward and reverse, either manually using the gene sequence or using a software program such as DNAWorks. The first forward primer includes the 5′ restriction enzyme (*Eco*RI for OmpT) recognition site and the last reverse primer includes the 3′ restriction enzyme recognition site (*Hin*dIII for OmpT); these restriction sites are used for cloning the library into the plasmid vector and suitable enzyme sites should be chosen accordingly. The set of oligos generated for the assembly of OmpT is shown in **Supplementary Figure 1**.
▲ CRITICAL STEP While designing primers, care must be taken to avoid gaps in the sequence to eliminate insertions/deletions during the PCR-based assembly reaction.

(ii) Resuspend the oligos to a final concentration of 50 µM using deionized H_2O.

(iii) To a sterile 0.5-ml eppendorf tube, add 36 µl of deionized autoclaved H_2O. Mix in 0.25 µl of each WT and degenerate primer. Mix thoroughly by pipetting a few times. In the case of OmpT, the WT primers would be 1f–3f, 5f–17f, 20f–24f, 1r–3r, 5r–16r, 18r, 20r–24r and the degenerate primers would be E27Nf, E27Nr, D208Nf, D208Nr, S223Nf and S223Nr (see **Supplementary Fig. 1**).

(iv) Set up a PCR as follows:

Volume	Component	Final concentration
41.5 µl	Distilled deionized H_2O	
5 µl	10× ThermoPol Buffer	1×
1 µl	dNTP mix (10 mM)	0.2 mM
2 µl	Primer mix from Step 3	10 pM (each primer)
0.5 µl	Vent DNA Polymerase (2 U μl^{-1})	1 U

(v) Assemble the oligos in a thermocycler using the following program tabulated below:

Cycle number	Denature	Anneal	Extend
1	95 °C, 2 min	—	—
2–41(−55)	95 °C, 30 s	52 °C, 30 s	72 °C, 30 s

■ **PAUSE POINT** The sample can be stored at 4 °C overnight or −20 °C forever.

(vi) Set up a second PCR using the outermost forward and reverse primers and the product of the first assembly PCR as the template as follows:

Volume	Component	Final concentration
40.5 µl	Distilled deionized H_2O	
5 µl	10× ThermoPol Buffer	1×
1 µl	dNTP mix (10 mM)	0.2 mM
0.5 µl	Primer 1f	1 µM
0.5 µl	Primer 24r	1 µM
2 µl	PCR product from after Step 1A(v)	
0.5 µl	Vent DNA Polymerase (2 U μl^{-1})	1 U

(vii) Amplify full-length fragments in a thermocycler using the following program tabulated below:

Cycle number	Denature	Anneal	Extend
1	95 °C, 2 min	—	—
2–30	95 °C, 30 s	58 °C, 30 s	72 °C, 30 s
31			72 °C, 10 min

■ **PAUSE POINT** The sample is left at 4 °C until required (typically overnight). The product can be stored at −20 °C indefinitely.

▲ **CRITICAL STEP** Although Taq can be used instead of Vent and often yields better amplification, the use of Taq increases the fraction of insertions/deletions.

(viii) Gel-purify the PCR product following separation on a 1% (wt/vol) agarose gel for 45 min at 100 V. Excise the band corresponding to the enzyme gene of interest (e.g., the OmpT band is ~1,000 bp in size) using a clean razor blade and extract-purify the DNA from the gel using the Zymoclean Gel DNA Recovery Kit according to the protocol supplied by the manufacturer.

■ **PAUSE POINT** Store the purified PCR product at −20 °C (can be stored indefinitely).

? TROUBLESHOOTING

(ix) Quantitate the amount of DNA recovered by measuring absorbance at 260 nm (A_{260}) using a nanodrop.

(x) Digest the enzyme gene-of-interest fragment (2–3 µg) with unique 5′ and 3′ restriction enzymes (5 U/µg of DNA), whose recognition sequences were incorporated earlier by a gene assembly design (Step 1A(i)) at 37 °C for 3 h according to the manufacturer's instructions. Separate the digested fragment on a 1% agarose gel, excise fragments of the appropriate size and gel-purify the DNA from the gel as described in Step 1A(viii).

(xi) Similarly, digest 3–4 µg of the appropriate expression plasmid (pDMLE19 in the case of OmpT) with the same unique restriction enzymes as used in Step 1A(x) (5 U/µg DNA) at 37 °C for 3 h according to the manufacturer's instructions. Run the digested plasmid on a 1% agarose gel and excise the fragment (pDMLE19 ~3,000 bp) corresponding to the vector backbone. Gel-purify the DNA and quantitate as described in Step 1A(xiii) and (ix).

■ **PAUSE POINT** Store the digested DNA products at –20 °C indefinitely.

(xii) Ligate 50 fmol of the digested and purified plasmid (pDME19 for OmpT) (from Step 1A(xi)) with a 3 molar excess of the digested gene (OmpT) fragment (from Step 1A(x)) (150 fmol) in a total volume of 20 µl at 25 °C for 4 h using T4 DNA Ligase according to the manufacturer's instructions. Heat-inactivate the reaction tube for 10 min at 65 °C and desalt the ligation product on a nitrocellulose membrane for 45 min at room temperature.

(xiii) Transform the ligation product into electrocompetent *E. coli* MC1061 (see **Box 1**). Mix 45 µl of electrocompetent cells in a prechilled 1.5-ml eppendorf tube with 2 µl of ligation product and transfer into a 1-mm gap prechilled cuvette. Apply a pulse at the setting of 25 µF, 1.8 kV, 200 Ω using Gene Pulser. Add 1 ml of SOC immediately, mix well and transfer into a sterile test tube. Use an additional 1 ml of SOC to rinse out the cuvette and combine.

! CAUTION Incomplete dialysis and/or air bubbles can cause arcing of the cuvette. Take care to avoid air bubbles.

(xiv) Recover for 1 h at 37 °C (250 r.p.m.) and spread transformants on 2xYT selective plates (Ampicillin for pDMLE19) and incubate for 8 h at 37 °C or 25 °C overnight (12 h). Depending on the competency of the cells, a typical transformation efficiency based on the ligation should be 3–7 × 10^7 independent colonies. Parafilm the plates and store at 4 °C. (The plates can be stored at 4 °C for 1 week).
? TROUBLESHOOTING

(xv) Pick 10 random colonies, inoculate 2 ml cultures of 2xYT (containing appropriate antibiotic), grow overnight (37 °C) and isolate plasmid from these cultures using the QIAGEN plasmid prep kit according to manufacturer's instructions.

(xvi) DNA—Sequence the gene from the isolated plasmid using plasmid-specific primers to ensure randomization at the desired codons.
? TROUBLESHOOTING

(xvii) Scrape the cells from plates described in Step 1A(xiv) and transfer into a sterile tube with LB medium supplemented with appropriate antibiotics and 2% (wt/vol) glucose and mix thoroughly.

(xviii) Freeze 1 ml aliquots of cells in 15% (wt/vol) final glycerol in a sterile 2-ml cryogenic vials and store at −80 °C (final OD_{600} 150–250 units).

(xix) Add 1 ml of the scraped, pooled cells into 500 ml of LB medium supplemented with appropriate antibiotics and 2% (wt/vol) glucose in a 2-liter shake flask and grow at 25 °C (250 r.p.m.) for 4–6 h until they reach an OD_{600} of 1–2 units.

(xx) Isolate plasmid from 50–100 ml of the culture using the QIAGEN plasmid prep kit according to the manufacturer's instructions.
■ PAUSE POINT Plasmid DNA can be stored indefinitely at −20 °C.

(xxi) Transform the plasmid (50 ng) into electrocompetent *E. coli* BL21(DE3) as described in Step 1A(xiii).

(xxii) Recover for 1 h at 37 °C (250 r.p.m.) and subculture the entire 2 ml into a sterile flask containing 500 ml of LB medium supplemented with appropriate antibiotics and grow for an additional 6 h at 37 °C until the culture reaches an OD_{600} of 2 units.

(xxiii) Freeze 200 µl aliquots of cells (final OD_{600} ~2 units) in 15% (wt/vol) final glycerol in a sterile 2-ml cryogenic vials and store at −80 °C indefinitely.

(B) Partial site-saturation libraries ● TIMING 4–5 d

(i) Resuspend the degenerate primers encoding for NNS to a final concentration of 50 µM as in Step 1A(ii).

(ii) To a sterile 0.5-ml eppendorf tube add 36 µl of deionized autoclaved H_2O. Mix in 0.25 µl the 24 reverse primers, 1r–24r. Label as PmixR.

(iii) In a separate sterile 0.5-ml eppendorf tube, add 2.25 µl of each of the NNS primers and 0.25 µl of the corresponding WT primers (for OmpT, these are 4f, 5f, 8f, 9f, 10f, 13f, 14f, 15f, 18f, 19f, 22f and 23f). Mix thoroughly and label as PmixN.

(iv) To a sterile 0.5-ml eppendorf tube, add 0.25 µl of the remaining forward primers (for OmpT, these are 1f–3f, 6f, 7f, 11f, 12f, 16f, 17f, 20f, 21f and 24f). Mix thoroughly and label as PmixF.

(v) Add 3 µl of PmixN (Step 1B(ii)) and 3 µl of PmixF (Step 1B(iv)) to 42 µl of PmixR (Step 1B(i)) and mix thoroughly.

(vi) Use 2 µl of the primer mix and assemble, amplify, digest and clone library exactly as described in Step 1A(iv–xxiii).

Flow cytometric screening ● TIMING 7–10 d

2| Thaw an aliquot of the frozen library in BL21(DE3) (Step 1A(xxiii) or 1B(vi)) and inoculate 100 ml of LB medium supplemented with appropriate antibiotics and grow at 37 °C for ~2 h until it reaches an OD_{600} of 2 units.

3| Turn on the flow cytometer and perform all the necessary fluidics/instrument start-up procedures.

4| Transfer a 1-ml aliquot of the cells into a sterile 1.5-ml eppendorf tube and spin at 8,161*g* for 2 min using a VWR Benchtop centrifuge. Discard the supernatant and resuspend in 1 ml of 1% (wt/vol) sucrose.

5| Spin down the cells again (8,161*g* for 2 min). Discard supernatant and resuspend in 1 ml of 1% (wt/vol) sucrose.

6| Add 948 µl of 1% (wt/vol) sucrose to a fresh sterile 1.5-ml eppendorf tube and add a 50-µl aliquot of the cells in sucrose (Step 5). Add 1 µl of the selection substrate (final concentration of 20 nM) and 1 µl of counter-selection substrate (final concentration of 100 nM). Mix the tube by inverting a few times and let the reaction proceed for 9 min at room temperature in the dark.

7| Set up an experiment on the flow cytometer with dot plots to monitor Forward (FSC) versus Side scatter (SSC) and fluorescence, BODIPY 530/30 (FL1) versus tetramethyl rhodamine 570/40 (FL2). Also, set up histograms that show counts of FL1 and FL2.
▲ CRITICAL STEP When performing multicolor sorting, it is important to set compensations to account for the spectral overlap of the fluorophores into the different filters being used. An excellent online resource for multicolor flow cytometry and compensations can be found on the BD Biosciences website.

! CAUTION Modern flow cytometers typically contain Class I lasers and high-voltage deflection plates. The user should have undergone training before using the instrument.

8| At the end of the 9 min, transfer 100 µl of the labeling reaction (from Step 6) to 1 ml of 1% (wt/vol) sucrose in a sterile 12 × 75 mm culture tube and analyze on the flow cytometer at 5–15k events per second.

? TROUBLESHOOTING

9| Set a gate on FL1 vs. FL2 to collect ~1 % of cells displaying high FL1 fluorescence and low FL2 fluorescence (**Fig. 3**) and sort into a 12 × 75 mm culture tube containing 250 µl of 2xYT containing the appropriate antibiotics. Collect 10–30 k events and spread on a 2xYT selective plate and incubate overnight at 37 °C. Typical viability after sorting is 80–90%.

10| Scrape cells after overnight incubation from Step 9, measure OD_{600} and inoculate a 5 ml culture of 2xYT containing the appropriate antibiotics (starting OD_{600} 0.2) and grow at 37 °C for 2 h until it reaches an OD_{600} of 2 units.

11| Perform additional rounds of sorting exactly as described in Steps 2–10.

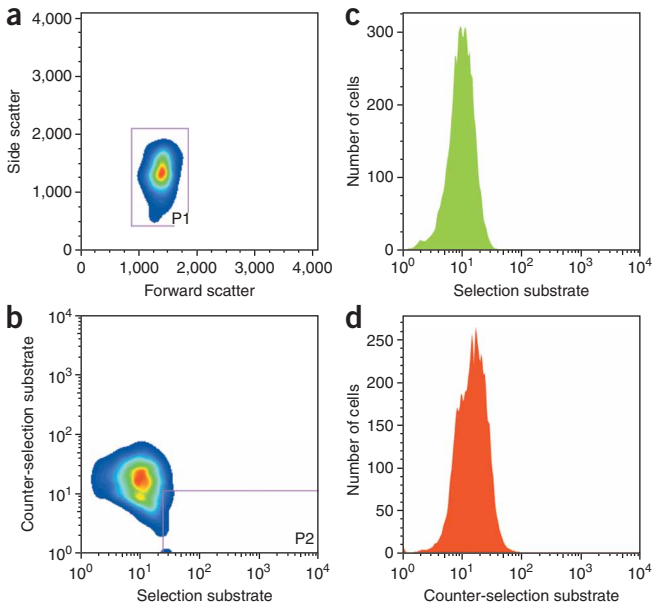

Figure 3 | Two-color sorting strategy. (**a**) A primary gate (P1) is set on FSC versus SSC that identifies *E. coli* cells. (**b**) A secondary sorting gate (P2), gated on P1, is set to collect cells displaying high fluorescence with the selection substrate (**c**) and low fluorescence with the counter-selection substrate (**d**).

12| After 4–6 rounds of sorting, when the FL1 mean stops increasing, pick 10–15 single colonies from the sorted 2xYT selective plate, inoculate 3 ml cultures of 2xYT containing the appropriate antibiotics and grow at 37 °C for 4–6 h until they reach an OD_{600} of 2 units.

? TROUBLESHOOTING

13| Use 1 ml of each of the single colony cultures from Step 12 to label and analyze single clones as described in Steps 2–8.

14| From the single colony cultures showing a selective fluorescence profile in Step 13, use the remainder of the corresponding cultures (Step 12) to isolate plasmid using a QIAGEN plasmid prep kit according to the manufacturer recommendations.

15| Sequence the gene using plasmid specific primers to determine the specific changes made to the WT sequence. Retransform the plasmid into electrocompetent BL21(DE3) for enzyme purification and further characterization.

● TIMING

Step 1A(i–xii) should take ~3 d

Step 1A(xiii–xxiii) should take an additional 3 d

Since each round of sorting (Steps 2–10) typically takes a day, the 4–6 rounds of sorting required to isolate enzyme variants with altered function should take approximately a week. The entire procedure, start to finish, should take 2–3 weeks.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
1A(viii)	No amplification or the presence of a smear	Incorrect annealing temperature	The annealing temperature will have to be optimized for the primer/gene pair. Another approach is to use 'Touchdown PCR'[23,24]
1A(xiv)	Small library size	Inefficient ligation	It is best to make single use aliquots of the ligation buffer. Repeated freeze-thaw is not recommended
		Poor transformation efficiency	The competency of the electrocompetent cells should be at least 3 × 10^{10} cfu/µg

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
1A(xvi)	Insertions/deletions in library	Use of Taq instead of Vent in PCR	High-fidelity polymerase such as pfu per Vent is recommended for assembly and amplification PCR
		Desalted oligos	HPLC/PAGE purified oligos are recommended
8	The entire population being analyzed is positive	Incubation times	Longer incubation ($>$15 min) leads to nonspecific labeling of the entire population
12	No enrichment	Event/sort rate	For most common benchtop flow cytometers event rates of 3–5 K/s are recommended
	No enrichment	Library	As smaller libraries might not yield variants with the desired selectivity profiles, it might be necessary to screen a larger and more complex library

ANTICIPATED RESULTS

Following the protocol outlined here, the user should be able to generate custom libraries of increasing complexity using PCR-mediated gene assembly. The advantage of using gene assembly is that if a given library fails to yield the desired enzyme variant, randomization at other potentially key residue(s) can be achieved in a straightforward manner. Following library construction and 4–6 rounds of flow cytometric screening (starting from $\sim 3 \times 10^7$ independent clones), one should expect to isolate cells (1–5 unique variants) displaying high FL1 (5- to 20-fold over negative control) and low FL2 (one- to two-fold over negative control). (Typical results for OmpT are shown in **Supplementary Fig. 3** online.) Note that a selective fluorescence profile does not automatically imply the presence of an active enzyme; further catalytic characterization (usually by HPLC) of purified variants using substrates that lack fluorescent tags is essential to the confirmation of initial sorting results.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS We thank Mark Pogson for assistance in preparing the figures.

Published online at http://www.natureprotocols.com/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Yep, A., Kenyon, G.L. & McLeish, M.J. Saturation mutagenesis of putative catalytic residues of benzoylformate decarboxylase provides a challenge to the accepted mechanism. *Proc. Natl. Acad. Sci. USA* **105**, 5733–5738 (2008).
2. Kaur, J. & Sharma, R. Directed evolution: an approach to engineer enzymes. *Crit. Rev. Biotechnol.* **26**, 165–199 (2006).
3. Park, H.S. *et al.* Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**, 535–538 (2006).
4. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
5. Ballinger, M.D., Tom, J. & Wells, J.A. Furilisin: a variant of subtilisin BPN' engineered for cleaving tribasic substrates. *Biochemistry* **35**, 13579–13585 (1996).
6. Bloom, J.D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447–452 (2005).
7. Stemmer, W.P. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
8. Griswold, K.E. *et al.* Evolution of highly active enzymes by homology-independent recombination. *Proc. Natl. Acad. Sci. USA* **102**, 10082–10087 (2005).
9. Martineau, P. Error-prone polymerase chain reaction for modification of scFvs. *Methods Mol. Biol.* **178**, 287–294 (2002).
10. Bornscheuer, U.T. & Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5**, 137–143 (2001).
11. Khersonsky, O., Roodveldt, C. & Tawfik, D.S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
12. Varadarajan, N., Rodriguez, S., Hwang, B.Y., Georgiou, G. & Iverson, B.L. Highly active and selective endopeptidases with programmed substrate specificities. *Nat. Chem. Biol.* **4**, 290–294 (2008).
13. Morley, K.L. & Kazlauskas, R.J. Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **23**, 231–237 (2005).
14. Cobaugh, C.W., Almagro, J.C., Pogson, M., Iverson, B. & Georgiou, G. Synthetic antibody libraries focused towards peptide ligands. *J. Mol. Biol.* **378**, 622–633 (2008).
15. Reetz, M.T. & Carballeira, J.D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).
16. Robertson, D.E. & Steer, B.A. Recent progress in biocatalyst discovery and optimization. *Curr. Opin. Chem. Biol.* **8**, 141–149 (2004).
17. Varadarajan, N., Gam, J., Olsen, M.J., Georgiou, G. & Iverson, B.L. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl. Acad. Sci. USA* **102**, 6855–6860 (2005).
18. Herman, A. & Tawfik, D.S. Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20**, 219–226 (2007).
19. Balint, R.F. & Larrick, J.W. Antibody engineering by parsimonious mutagenesis. *Gene* **137**, 109–118 (1993).
20. Santoro, S.W., Wang, L., Herberich, B., King, D.S. & Schultz, P.G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20**, 1044–1048 (2002).
21. Vandeputte-Rutten, L. *et al.* Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *EMBO J.* **20**, 5033–5039 (2001).
22. Daugherty, P.S. Protein engineering with bacterial display. *Curr. Opin. Struct. Biol.* **17**, 474–480 (2007).
23. Korbie, D.J. & Mattick, J.S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat. Protoc.* **3**, 1452–1456 (2008).
24. Abecassis, V. *et al.* High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **28**, E88 (2000).