Video Article

In Vitro Directed Evolution of a Restriction Endonuclease with More Stringent Specificity

Krzysztof J. Skowronek[1], Matthias Bochtler[1]

[1]International Institute of Molecular and Cell Biology, Warsaw (IIMCB)

Correspondence to: Krzysztof J. Skowronek at kskowronek@iimcb.gov.pl, Matthias Bochtler at mbochtler@iimcb.gov.pl

Abstract

Restriction endonuclease (REase) specificity engineering is extremely difficult. Here we describe a multistep protocol that helps to produce REase variants that have more stringent specificity than the parental enzyme. The protocol requires the creation of a library of expression selection cassettes (ESCs) for variants of the REase, ideally with variability in positions likely to affect DNA binding. The ESC is flanked on one side by a sequence for the restriction site activity desired and a biotin tag and on the other side by a restriction site for the undesired activity and a primer annealing site. The ESCs are transcribed and translated in a water-in-oil emulsion, in conditions that make the presence of more than one DNA molecule per droplet unlikely. Therefore, the DNA in each cassette molecule is subjected only to the activity of the translated, encoded enzyme. REase variants of the desired specificity remove the biotin tag but not the primer annealing site. After breaking the emulsion, the DNA molecules are subjected to a biotin pulldown, and only those in the supernatant are retained. This step assures that only ESCs for variants that have not lost the desired activity are retained. These DNA molecules are then subjected to a first PCR reaction. Cleavage in the undesired sequence cuts off the primer binding site for one of the primers. Therefore, PCR amplifies only ESCs from droplets without the undesired activity. A second PCR reaction is then carried out to reintroduce the restriction site for the desired specificity and the biotin tag, so that the selection step can be reiterated. Selected open reading frames can be overexpressed in bacterial cells that also express the cognate methyltransferase of the parental REase, because the newly evolved REase targets only a subset of the methyltransferase target sites.

Video Link

The video component of this article can be found at https://www.jove.com/video/60807/

Introduction

Sequence specificity engineering is extremely challenging for class II REases. In this class of endonucleases, sequence recognition and catalysis are closely intertwined, most probably as an evolutionary safeguard against creation of an endonuclease of broader specificity than its cognate methyltransferase, which would damage host DNA. Directed evolution of new specificities in cells is further complicated by the need to protect host DNA against the newly engineered endonuclease activity. Therefore, there are only a few successful attempts of REase engineering reported and all of them exploit the unique features of a particular enzyme[1,2,3,4,5,6,7].

Here we provide a detailed protocol for specificity engineering that can be used to generate endonuclease variants that have narrower specificity than a parental enzyme that is based on our successful engineering of a NlaIV endonuclease[8]. For any such enzyme with an arbitrary recognition sequence, extra specificity can be introduced for bases in the flanks. For parental enzymes that recognize partially degenerate sequences (such as NlaIV with its GGNNCC target), additional specificity can also be introduced within the recognition sequence. As extra specificity will likely require protein-DNA contacts, the newly recognized bases should lie within the footprint of the parental endonuclease on DNA. In principle, selection schemes can be set up for any desired specialization of the recognition sequence. However, most REases that recognize palindromic and nearly palindromic target sequences are functional dimers that recognize only a half-site of the palindrome. Hence, selection of new specificities that violate the symmetry of protein nucleic interactions is unlikely to work. For the dimeric NlaIV, for example, the GGNNCC sequence can theoretically be narrowed down to GGATCC but narrowing the specificity down to GGAACC is expected to be more difficult. Our scheme involves both positive and negative selection.

The process is more efficient when negative selection is also used to remove the specificities able to cleave all sequences other than the preferred narrower specificity. For example, selection for GGATCC could be combined with antiselection against GGBVCC (where B is any base other than A, and V is any base other than T). When some of the possible target sequences are not covered, the outcome of the selection experiment depends on the effectiveness of positive and negative selection. In our NlaIV work, we selected for GGATCC, and against GGSSCC (where S is G or C), and obtained a specificity that, ignoring symmetry breaking targets, could be described as GGWWCC (where W is A or T), suggesting that in this particular case, negative selection was more important than positive selection.

Our approach starts with the creation of an expression selection cassette (ESC). The ESC is structured in sections. On the inside core section, there are variants of the open reading frame (ORF) of the REase, under T7 promoter control. This core section of the ESC cannot contain any cognate site for the engineered REase. The core is sandwiched between two cognate sites for wild type REase: a cleavage site for the undesired activity (counter selected sequence, GG**SS**CC in this example) and a cleavage site for the desired activity (selected sequence, GG**AT**CC in the example). The final step of the preparation of the ESC in PCR adds biotin close to the desired activity at the 5' end and creates a variety of counter selected sequences (GGSSCC in the example). The selection strategy relies on the use of carefully designed primers at the ESC reamplification protocol after an in vitro transcription/translation/selection protocol (**Figure 1A**). The ESC library is expressed in an in vitro compartmentalized transcription translation water-in-oil emulsion[9,10,11]. Within each droplet, the specificity of the expressed enzyme affects the state of the ESC (**Figure 1B**, step I). For the described arrangement, the desired cleavage activity of the translated protein removes the DNA's biotin tag but does not affect the other ESC end with the counter selected sequence. When the emulsion is broken, biotinylated fragments are removed by streptavidin affinity pulldown, so that only fragments from droplets with the desired activity remain (**Figure 1B**, step II). This step removes inactive REase variants. The supernatant fraction of the pull-down step is then amplified by PCR. In the first PCR reaction primers F2 and R1 are used (**Figure 1A,B**, step III). Primer F2 binds to the ESC section between the counter selected sequence and the molecule end. Therefore, ESCs expressing variants that are capable of cleaving the counter selected sequence (and, therefore, separate the binding sites for primers F2 and R1 into two different DNA molecules) are not amplified and are thus removed from the library. The primer R1 binds between the selected site and the core of the ESC so that it is not affected by the cleavage status of the selected site and restores the cleavage site for the desired activity (GGATCC). The cycle is closed by a second PCR (with primers F1 and R2) that adds biotin at the 5' end close to the selected site and restores designed variation at the counter selected site close to the opposite end of the ESC (**Figure 1B**, step IV). The resulting DNA mixture is ready for another round of selection.

The success of the selection protocol depends strongly on the proper choice of the new, more stringent target recognition sequence and on careful design of the mutagenesis strategy and its effective implementation. Because it is much easier to improve upon slight preexisting preferences of the REase than to overcome them, we recommend starting with a kinetic study of any preexisting preferences. The necessity of careful mutagenesis design results from the limited size of a mutant library that can be processed by the presented protocol (10^9 clones in a single experiment). Therefore all 20 possible amino acid substitutions can be effectively tested in only a few positions (see Discussion). Random mutagenesis, such as error-prone PCR (EP-PCR) presented as an alternative method, will lead to profound undersampling of existing complexity. If any information concerning potential amino acid positions involved in contacts with DNA (or even located in a close proximity to the degenerate nucleotides in a cognate sequence) is available, it certainly should be used to select a few amino acids for oligonucleotide guided saturation mutagenesis (protocol steps 1.6-3.10).

Protocol

1. Preparation of ESCs

1. Clone methyltransferase of the restriction-modification system to be engineered in a low copy number plasmid (e.g., pACYC184 or pACYC174 or their derivatives).
   NOTE: The bacterial host strain must be able to tolerate methylation introduced by the cloned enzyme and provide inducible expression of T7 RNA polymerase. Use of the ER2566 strain (carrying McrA, McrBC, and Mrr mutations) is recommended.
2. Confirm that the recombinant plasmid DNA is protected against cleavage by the cognate endonuclease by treating 0.5 µg of plasmid DNA with 10 units of cognate REase in buffer and temperature recommended by the enzyme supplier for 2 h.
3. Prepare competent cells of this strain.
   NOTE: Any method can be used. The NlaIV engineering project used a simple calcium chloride method[12].
4. Construct recombinant plasmid with the ORF for the REase under control of the T7 promoter from a different exclusion group and with a different selection marker than the one containing the methyltrasferase gene in step 1.1. Vectors pET28 and pET30 can be used.
5. **Remove all recognition sites for the engineered enzyme from the section of the recombinant plasmid between the T7 promoter and the stop codon of the enzyme ORF by introducing silent mutations (Figure 2, Table 1A).**
   NOTE: If more than one such site must be removed, multiple mutation rounds will be necessary (steps 1.5.1–1.5.7).
   1. Use an inside-out PCR reaction that amplifies the full-length plasmid with designed variations introduced at the 5' ends of the primers (**Table 2A**).
   2. Remove the template DNA by adding 10 U of DpnI endonuclease to the 50 µL of the PCR reaction, and incubate for 2 h at 37 °C.
   3. Resolve the products by agarose gel electrophoresis. Cut out the band corresponding to the full-length plasmid and purify it with a commercial kit.
   4. Add 10x ligation buffer (to a 1x concentration) and supplement with ATP (to 1 mM). Add 10 U of T4 polynucleotide kinase and incubate for 20 min at 37 °C. Inactivate the enzyme by heating at 70 °C for 10 min.
   5. Add PEG 4000 to 5%, supplement again with ATP (to 1 mM), and add 5 U of T4 DNA ligase. Incubate for 2 h at room temperature (RT).
   6. Transform into a competent bacterial strain carrying cognate methyltransferase (step 1.1).
   7. Isolate the plasmid DNA in small scale and confirm the introduction of sequence changes by dideoxy sequencing.
6. Introduce unique restriction sites close to the sequence(s) targeted by oligonucleotide guided mutagenesis (**Figure 2**, **Table 1B**). Follow steps 1.5.1–1.5.7 for each site.
   NOTE: This step is performed only when a targeted mutagenesis is used. If doing random mutagenesis, skip steps 2-3 and proceed to section 3 instead. In the presented example all sites were introduced upstream of the targeted regions, but they can be introduced downstream as well.
7. **Design primers for the amplification of the ESC (Table 1C).**
   1. Design a reverse primer binding downstream of the endonuclease ORF that will introduce the selected recognition site (R1) and its shorter version (R2) that binds outside the selected NlaIV sequence and contains biotin at the 5' end (see **Figure 1**).

2. Design a forward primer (F1) binding to the ESC upstream of the T7 promoter. This primer should also introduce counterselected variant(s) of the original recognition sequence (i.e., the maximum of sequence variations recognized by the original enzyme with the exception of the selected reverse sequence).
NOTE: A shorter version of this primer (F2) that covers the sequence distal to the counterselected sequence will be used later in the selective PCR (step 5.9).

2. Split-and-mix Synthesis of Mutagenic Primers

NOTE: This step is used only for projects that require subsaturation mutagenesis at more than one site. A synthesizer with multiple synthesis columns is required. Assign columns for synthesis of randomized NNS codon triplets and wild type codon triplets according to the mutagenesis frequencies. For example, if seven equal volume synthesis columns are available, and a mutagenesis rate of 0.3 is desirable at a given site, add randomized NNS codons in ~0.3 x 7 or two columns, and wild type codons in ~0.7 x 7 or five columns (**Figure 3**).

1. Decide about sites for subsaturation mutagenesis. Choose mutagenesis frequencies according to the hypothetical importance of the sites (i.e., the more important the site, the higher the frequency), keeping limits on the overall library complexity in mind (see Discussion).
2. Synthesize oligonucleotides in all columns, up to the triplet immediately preceding the second subsaturation mutagenesis site counting from the 3'-end. Do not remove the 5'-trityl protecting group at the last synthesis cycle (use the trityl-on option on the synthesizer). The protecting group will be removed at the beginning of the next synthesis cycle (step 1 in **Figure 3**).
3. Open the synthesis columns. Collect controlled pore glass (CPG) synthesis support into a dry 1.5 mL tube and mix by vortexing. Repartition the mixed CPG resin into new synthesis columns. Avoid introducing humidity, because it will decrease the overall yield (steps 2 and 4 in **Figure 3**).
4. Continue synthesis, starting from the subsaturation mutagenesis site triplet. Assign columns to randomized NNS triplets or wild type triplets according to the desired mutagenesis frequency (see note above). If additional subsaturation sites are present, proceed only to the triplet preceding the next subsaturation mutagenesis site. Again, leave a 5'-trityl group on at the end (5'-trityl-on option on the synthesizer) (step 3 in **Figure 3**). Then continue with step 2.3.
5. If no more subsaturation sites are present downstream, complete the synthesis, leaving a 5'-trityl group at the end (5'-trityl-on option on the synthesizer) (step 5 in **Figure 3**).
6. Deprotect and purify the oligonucleotide library according to the purification cartridge manufacturer's instructions.
NOTE: Oligonucleotides released by deprotection from the CPG can also be purified in the reverse phase high performance liquid chromatography (HPLC) with trityl-on followed by a manual trityl group removal (1 h treatment with 80% acetic acid at RT) and a second HPLC purification.
7. Check the oligonucleotide library quality in a urea-PAGE gel.

3. Generating Variant Libraries

NOTE: Use the recombinant plasmid from step 1.6.

1. **Generate the libraries by oligonucleotide directed mutagenesis.**
NOTE: Alternatively, use the EP-PCR protocol (step 3.2).
   1. Amplify a section from the T7 promoter to the unique restriction enzyme site flanking the sequence targeted with mutagenesis (in case of NlaIV: SalI, EcoRI, or Eco52I) (**Table 1B-C**, **Table 2B**, **Figure 4**). Amplify the second part from the unique restriction enzyme site to the 3' end of the ESC.
   2. Mix separately 5 µL of the PCR reactions (from step 3.1.1) with 8 µL of water, 1.5 µL of 10x restriction enzyme buffer, and 5 units of the appropriate restriction enzymes (SalI, EcoRI, or Eco52I) and incubate at the appropriate temperature for 2 h.
   3. Resolve the products of both reactions using agarose gel electrophoresis. Cut out the expected size bands and purify with a commercial kit.
   4. Run up to 1/3 of the purified products in an agarose gel and measure the concentration of each purified band by densitometry.
   5. Set up the ligation of two parts of the ESCs in a 1:1 molar ratio with 1x ligase buffer and 1 U of T4 DNA ligase and incubate for 2 h at RT.
   6. Resolve the reaction products in the agarose gel. Cut out the expected size products and purify with a commercial kit.
   7. Amplify the purified ligation products in a PCR reaction with primers F1 and R2 (**Table 1C** and **Table 2A**). Do not run more than 20 amplification cycles.
   8. Fractionate the PCR reactions in an agarose gel. Cut out the products and purify with a commercial kit.
   9. Run a 5 µL aliquot of the purified library from the previous step in the agarose gel and measure the concentration by densitometry.
   10. Clone a small sample of the library (up to 5 µL) and sequence >15 clones to check the mutation frequency and distribution (**Table 3**). Proceed to step 4.
   NOTE: Alternatively, high throughput sequencing of the small sample of the ESCs can be used.

2. **Perform EP-PCR.**
   1. Amplify the ESC from the plasmid obtained in step 1.5.7 with primers F1 and R1. Run 20 cycles with Taq I polymerase (**Table 1B**).
   2. Gel purify the PCR product.
   3. Set up EP-PCR with 2 ng of purified PCR product from the previous step and run 15 cycles of EP-PCR (**Table 1C**) with F1 and R1 primers.
   4. Gel purify the product and quantify it by gel densitometry.
   NOTE: Due to the low concentration of the purified EP-PCR product use about 1/3 for quantification.
   5. Clone a small sample of the library (up to 1/5) and sequence >15 clones to check mutation frequency and distribution (**Table 4**).

NOTE: Alternatively, perform high-throughput sequencing of the small sample of the ESCs.

4. Performing Compartmentalized In Vitro Transcription-translation Reaction

1. **Test endonuclease expression and enzymatic activity in in vitro transcription-translation.**
   1. Prepare a short (200–500 bp) substrate with a single recognition site for the endonuclease located close to the center of the molecule so the cleavage reaction can be easily detected.
      NOTE: The easiest way to prepare the substrate is by PCR amplification of an appropriate fragment of any DNA molecule. The substrate can be radiolabeled or fluorescently labeled to simplify cleavage detection.
   2. Set up 50 µL of a transcription-translation reaction with 0.5 µg of wild type ESC according to manufacturer's recommendations. Add magnesium salt ($MgCl_2$, $MgSO_4$, and magnesium acetate can be tested) to 1.5 mM and the appropriate amount of substrate from the previous step (at least 0.5 µg in case of unlabeled DNA).
      NOTE: Any transcription/translation kit that does not contain nuclease activated by magnesium can be used. Some kit vendors use nucleases to remove DNA contamination during production and then add chelators as nuclease inhibitors. Such kits are not compatible with this method.
   3. Incubate the transcription-translation reaction according to the manufacturer's instructions. Then transfer the reaction mixture to the optimal temperature for the restriction enzyme for 2 h.
   4. Analyze cleavage of the substrate in an agarose gel followed by appropriate detection (e.g., DNA staining, fluorescence visualization, or autoradiography) (**Figure 5**).
      NOTE: At least partial cleavage of the substrate is necessary before proceeding with the compartmentalization. If this is not achieved, further optimization of the magnesium chemical form or its concentration is necessary.

2. Prepare an oil-surfactant mixture by adding 225 µL of Span 80 and 25 µL of Tween 80 to 5 mL of mineral oil in a 15 mL conical tube. Mix thoroughly by gentle inverting the tube 15x.
3. For each library transfer 950 µL of the oil-surfactant mixture to a 2 mL round bottom cryogenic vial, label with a library name, and transfer to ice. Put one small cylindrical stirring bar (5 x 2 mm^2) into each vial.
4. Prepare an in vitro transcription-translation reaction mixture (50 µL for each library) according to the manufacturer's suggestions. Supplement the mixture with magnesium chloride to a final concentration of 1.5 mM (see step 4.1.4).
5. Dispense 50 µL aliquots into 1.5 mL tubes on ice.
6. Add 1.7 fmole of the library (from section 3) to the reaction mixture on ice.
   NOTE: Do not use a higher amount of expression library for selection efficiency. It is crucial to minimize the frequency of aqueous droplets containing more than one DNA molecule.
7. **Prepare water-in-oil emulsion consecutively for each library.**
   1. Put a small beaker (or large bottle cup) filled with ice on a magnetic stirrer with the stirring speed set at 1,150 rpm.
   2. Transfer a cryogenic vial with 950 µL of oil-surfactant mixture and a small stirring bar from step 4.3 to an ice-cold beaker on the magnetic stirrer. Check that the stirring bar is spinning.
   3. Add five 10 µL aliquots of the in vitro library-transcription-translation mixture over a 2 min period in 30 s intervals and continue stirring for an additional minute. Transfer the vial with the emulsion to an ice container. Proceed with the next library starting with step 4.7.2.
   4. After all the libraries are processed start the incubation of all the libraries according to the kit manufacturer's recommendations.

8. Transfer the vials to the temperature optimal for the engineered endonuclease for an additional 2 h and then put them on ice for at least 10 min.

5. Continued Processing of Libraries and Selection

1. Transfer the emulsions from the cryogenic vials into cold 1.5 mL tubes, add 1 µL of 0.5 M EDTA and centrifuge them at 13,000 x *g* for 5 min at room temperature.
2. Remove the upper oil phase with a pipette. If an oil-water interphase is not visible, incubate the tube for at least 5 min at -20 °C to freeze the aqueous phase, then immediately pipet out the liquid oil phase.
3. Add 100 µL of 10 mM Tris HCl pH 8.0 and immediately perform extraction with 150 µL of phenol:chloroform (1:1 v/v) by short vortexing followed by phase separation by 30 s centrifugation at 13,000 x *g*. Collect the upper aqueous phase.
4. Precipitate the DNA by adding 0.1 vol (15 µL) of 3 M sodium acetate (pH = 5.2), 2.5–5 µg of glycogen and 2.5 vol (375 µL) of ethanol. Incubate at -20 °C for 1 h and centrifuge for 15 min at 13,000 x *g*, 4 °C. Discard the supernatant and briefly wash the pellet with 1 mL of cold 70% ethanol.
5. Dry the DNA/glycogen pellet in a speedvac or air dry for >5 min.
6. Dissolve the pellet in 50 µL of 10 mM Tris-HCl (pH = 7.5). Add 5 µL of streptavidin magnetic beads prepared according to the manufacturer's instructions and mix for 1 h at RT, preferably in a carousel mixer or by gentle vortexing.
7. Separate the beads on a magnetic stand and collect the liquid enriched in DNA without biotin.
8. Concentrate the DNA by ethanol precipitation (steps 5.4–5.5).
9. Dissolve the concentrated DNA from the previous step in 5 µL of water and use as a template in a PCR reaction with F2 and R1 primers (**Table 1A**).
   NOTE: To avoid problems with template contamination and minimize PCR artifacts use Taq polymerase (not Pfu or Phusion) and run 18–20 cycles with the extension time proportional to the template size (1 kb = 1 min) (see **Table 2B**).
10. Fractionate the PCR product in an agarose gel and cut out the expected size product. Some smearing indicates that there are products of different sizes (see **Figure 6**). Purify the DNA from the gel slab with a commercial kit.

11. Run a second PCR reaction with up to 50 ng of DNA from step 5.10 and primers F1 and R2 using the same protocol as in step 5.9. Proceed with product purification as described in 5.10. Purified DNA after quantification by agarose gel densitometry (not UV spectroscopy) can be used in the next round of in vitro selection (step 4.6).

6. Screen variants for Altered Sequence Specificity

1. **Clone selected variants.**
   1. Digest the product from step 5.10 for 2 h with 10 U of restriction enzymes appropriate for cloning of the ORF into the expression vector (for NlaIV: NcoI and XhoI) in the temperature and buffer recommended by the enzyme vendor. Resolve the products with agarose gel electrophoresis and isolate the expected size fragment.
   2. Prepare the plasmid vector (e.g., pET28) by double cleavage with the same enzymes as in step 6.1.1 and gel purify the product with a commercial DNA gel purification kit.
   3. Estimate the concentrations of vector and insert by densitometry with agarose gel electrophoresis.
   4. Set up a ligation with 1–5 U of T4 DNA ligase and vector:insert molar ratio 1:3–1:5 in 1x ligase buffer recommended by the enzyme vendor. Incubate for 2 h at RT and introduce into appropriate host bacteria (from step 1.3) by transformation or electroporation[12].
   5. Select extransformants on LB plates containing the appropriate antibiotic (50 µg/mL of kanamycin for pET28 or pET30 vectors) and 1% glucose.

2. **Express protein variants.**
   1. Inoculate single colonies from the transformation (up to 24 clones can be easily processed in a single run) into 2 mL of LB with kanamycin (50 µg/mL) and 1% glucose and grow overnight at 37 °C with shaking.
   2. Inoculate 15 mL of warm (37 °C) LB containing 100 µg kanamycin and no glucose with 0.75 mL of the overnight culture and incubate at 37 °C with vigorous shaking.
   NOTE: Either 50 mL centrifuge tubes or 100 mL Erlenmayer flasks can be used.
   3. Add 176 µL of glycerol to 1 mL of overnight culture (final concentration of glycerol = 15%) mix thoroughly and freeze at -70 °C.
   4. After 2–3 h supplement 15 mL of the culture (from step 6.2.1) with IPTG to 1 mM and culture for an additional 5 h.
   5. Collect the bacterial pellet by centrifugation (10,000 x *g*, 4 °C, 10 min) and freeze at -70 °C.

3. **Purify the protein variants.**
   1. Transfer 20 µL of nickel affinity resin suspension into 200 µL of B1 buffer in a 1.5 mL tube with a wide bore pipette tip, mix gently, and centrifuge (5,000 x *g*, 30 s, 4 °C). Remove the supernatant by pipetting and leave the tube on ice.
   2. Resuspend the bacterial pellet from step 6.2.5 in 300 µL of B1 by vigorous vortexing. Transfer the suspension into a 1.5 mL tube.
   3. Add 3 µL of 100x protease inhibitor cocktail and lysosome solution in B1 (final concentration of 1 mg/mL). Disintegrate the cells by sonication with a tip equipped probe. Use six 10 s bursts per sample with >15 s tip cooling time in ice in between. Keep cell suspensions on ice all the time.
   4. Pellet cell debris by centrifugation (2 min, 12,000 x *g*, 4 °C) and transfer 250 µL of supernatant to the resin aliquot from step 6.3.1.
   5. Mix for 15 min in a cold room, preferably in a carousel mixer or by gentle vortexing.
   6. Centrifuge (5,000 x *g*, 30 s, 4 °C) and aspirate the supernatant with a pipette.
   7. Add 500 µL of W buffer and gently resuspend the resin. Centrifuge (5,000 x *g*, 30 s, 4 °C) and aspirate the supernatant with a pipette.
   8. Repeat step 6.3.7.
   9. Add 20 µL of buffer E, gently resuspend the resin, and leave the sample on ice for 2–5 min. Centrifuge (5,000 x *g*, 30 s, 4 °C) and collect the supernatant.
   10. Repeat step 6.3.9. Pool supernatants.
   11. Analyze protein samples in by SDS-PAGE (5–10 µL) (**Figure 7**).

4. **Screen for variants with the altered specificity.**
   1. Assay cleavage activity on bacteriophage lambda DNA. The protein sample can constitute up to 10% of the final reaction volume. A total of 2 µL of protein sample per 0.5 µg of DNA and 2 h reaction time is a good starting point.
   2. Analyze the reaction products by agarose gel electrophoresis along with the products generated by the wild type enzyme. Select the clones generating cleavage patterns clearly distinguishable from the one generated by the wild type enzyme for further analysis (**Figure 8**).

Representative Results

This protocol is just a tool to increase the frequency of desired variants of an engineered REase by depleting (but not eliminating) two unwanted classes: inactive enzymes and endonucleases with unchanged wild type sequence specificity. On the other hand, because changing REase specificity is extremely difficult, finding even one such variant producing a cleavage pattern that is different from the wild type enzyme in a single screening of 24 clones should be considered a success. In our hands the best screens could identify up to 20% of promising variants (**Figure 8A**).

The positive outcome strongly depends on a library quality (i.e., limited frequency of substitutions and their random distribution) and efficient capture of the biotinylated population of library members (steps 3.6–3.7). Both problems can be detected. The library quality should be checked prior to the selection by sequencing as many clones as possible (>15) or by direct sequencing of the library by high throughput sequencing (step 3.10, **Table 3**). If a majority of the selected clones are not active, this is a clear indication of failure of the streptavidin capture selection. A similar effect is observed in the case of libraries that undergo many selection cycles, because such libraries are most probably dominated by inactive variants that escaped the streptavidin capture selection step (**Figure 8B**). Therefore, it is advisable to run screening after every selection cycle and further develop manually selected promising variants rather than to depend on selection iteration.
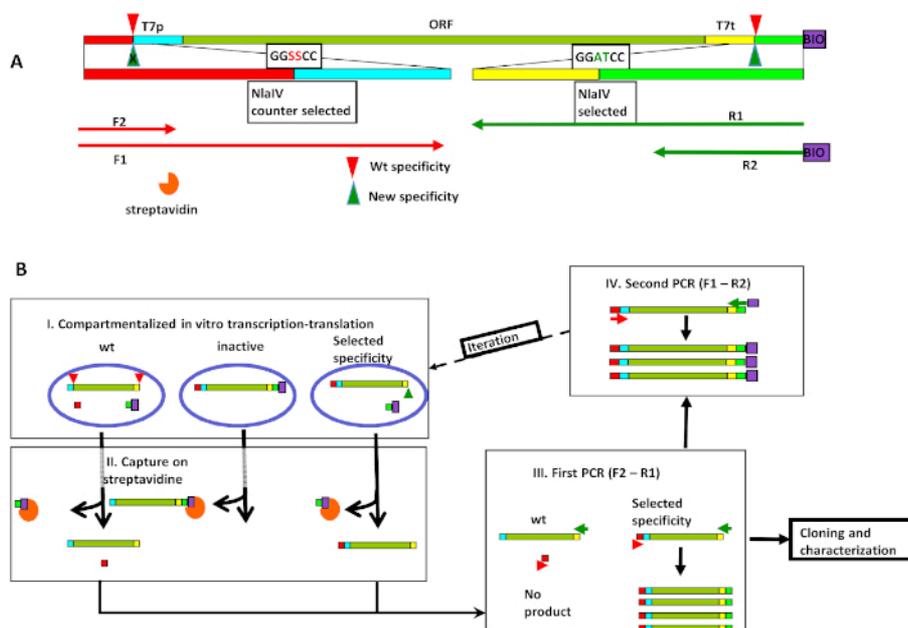
Figure 1: In vitro selection of a new sequence specificity based on NlaIV engineering. (A) The organization of the expression/selection cassette (ESC) includes two recognition sites for REase, 1) the selected sequence (GGATCC) close to the right end and 2) the counter selected sequence (GGSSCC) close to the left end, as well as the T7p and T7t-T7 promoter and T7 terminator. The primer binding sites are shown below. Cleavage by wild type and selected NlaIV variants are shown as red and green triangles respectively. **(B)** Selection cycle steps: I) Emulsification of transcription-translation-cleavage reaction mixes with the ESC library; II) All biotinylated DNA is captured on magnetic particles coated with streptavidin and removed, thus removing encoding inactive variants; III) ESCs encoding REases with wild type activity (i.e., those able to cleave the GGSSCC sequence) are eliminated because cleavage of the sequence separates the binding sites for the forward and reverse primers. Therefore, no amplification of these ESCs occurs; IV) Input for the next selection round is created by addition of biotin on the right end and reintroducing variation of the counter selected sequences on the left end. Reprinted from Czapinska et al.[8] with permission from Elsevier. Please click here to view a larger version of this figure.
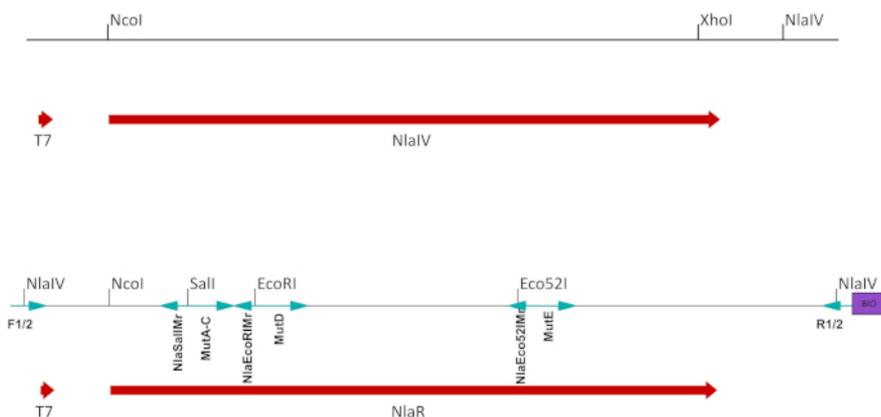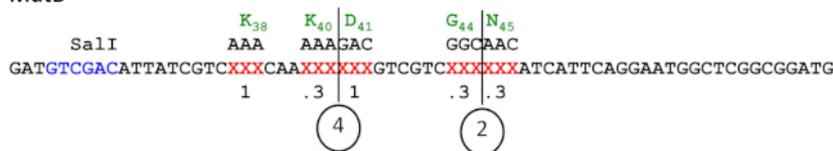

Figure 2: Preparation of ESC. Fragment derived from the original construct in an expression vector containing NlaIV ORF under control of the T7 promoter was modified to be suitable for expression/selection. The NlaIV site downstream from the NlaIV ORF was removed and unique sites (SalI, EcoRI and Eco52I) that were used to mutagenize selected positions were introduced in the NlaIV ORF as silent mutations. The final construct was amplified with flanking primers that introduced two flanking NlaIV sites: The counter selected sequence (GG**SS**CC) on the left and selected sequence (GG**AT**CC) on the right. The reverse primer also introduced biotin. Primers used in creation of mutated ECS are shown as blue arrows and labeled below (see **Table 1B,C**). Please click here to view a larger version of this figure.

MutB

```
                K₃₈    K₄₀ D₄₁        G₄₄ N₄₅
     SalI       AAA    AAAGAC         GGCAAC
GATGTCGACATTATCGTCXXXCAAXXXXXXGTCGTCXXXXXXATCATTCAGGAATGGCTCGGCGGATG
                 1    .3 1         .3 .3
                      (4)              (2)
```

1. Synthesis of **NNS**ATCATTCAGGAATGGCTCGGCGGATG (2 of 7 columns) and **AAC**ATCATTCAGGAATGGCTCGGCGGATG (5 of 7 columns)
2. Mix and spilt
3. Synthesis of **NNS**GTCGTC**NNS** (2 of 7 columns) and **NNS**GTCGTC**GGC** (5 of 7 columns)
4. Mix and split
5. Synthesis of GAT<u>GTCGAC</u>ATTATCGTC**NNS**CAA**NNS** (2 of 7 columns) and GAT<u>GTCGAC</u>ATTATCGTC**NNS**CAA**AAA** (5 of 7 columns)

Figure 3: Scheme of split and mix synthesis. The example refers to MutB primer synthesis where an NNS sequence was introduced at 0.8 frequency at four positions (see also **Table 3**). Note that chemical synthesis is carried out from 3' to 5' but all sequences are shown in canonical 5'-3' orientation (i.e., it proceeds from right to left in this scheme). Wild type sequences at mutagenized positions are shown in green while NNS mutagenic sequences are in red. The SalI recognition site that is later used to introduce mutations in ESCs is underlined. Points of mixing and splitting steps (2 and 4) are indicated. Please click here to view a larger version of this figure.
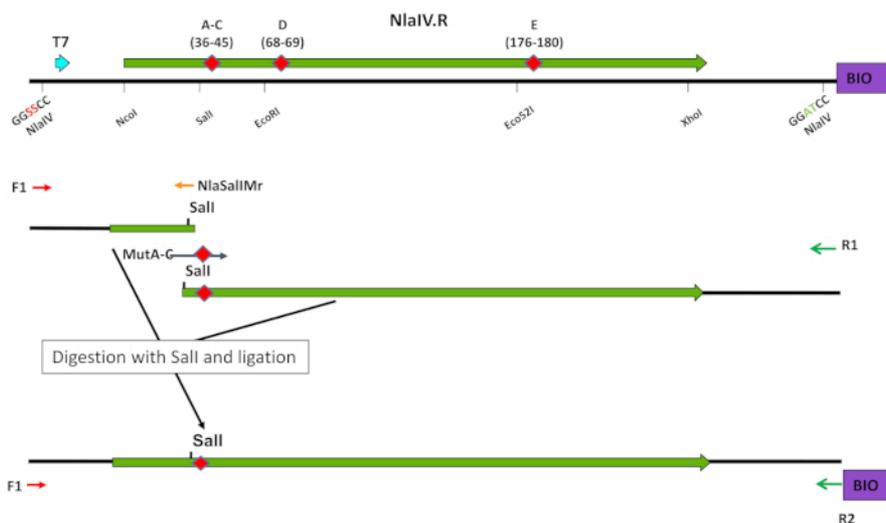

Figure 4: Use of unique restriction enzyme sites in oligonucleotide targeted mutagenesis. The strategy of mutation introduction is shown on an example of the construction of libraries A-C (see steps 3.1–3.7). Reprinted from Czapinska et al.[8] with permission from Elsevier. Please click here to view a larger version of this figure.
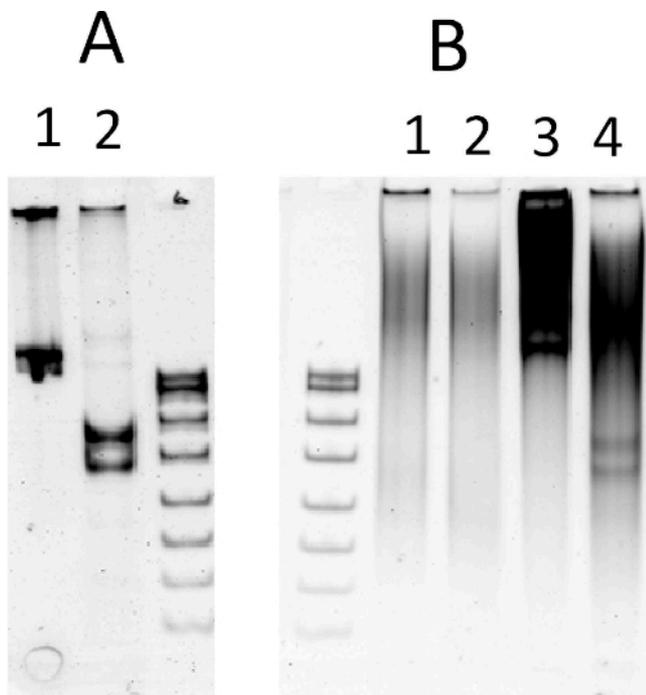
Figure 5: Endonucleolytic cleavage in in vitro transcription-translation. (**A**) Cleavage of a test substrate in optimal REase buffer: 1) Substrate, 612 bp PCR product with a single NlaIV recognition site; 2) Cleavage products, 355 bp and 257 bp. (**B**) Cleavage in an in vitro transcription-translation reaction (containing 0.5 µg of ESC): 1–2) 15 µL aliquots of in vitro transcription translation without substrate (line 2: reaction supplemented with 1.5 mM $MgCl_2$); 3–4) 15 µL aliquots of in vitro transcription-translation with 1 µg of test substrate; (line 4: reaction supplemented with 1.5 mM $MgCl_2$). S-DNA size marker (pBR322 digested with MspI). Samples were resolved in 6% native PAGE. DNA was stained with ethidium bromide. Please click here to view a larger version of this figure.
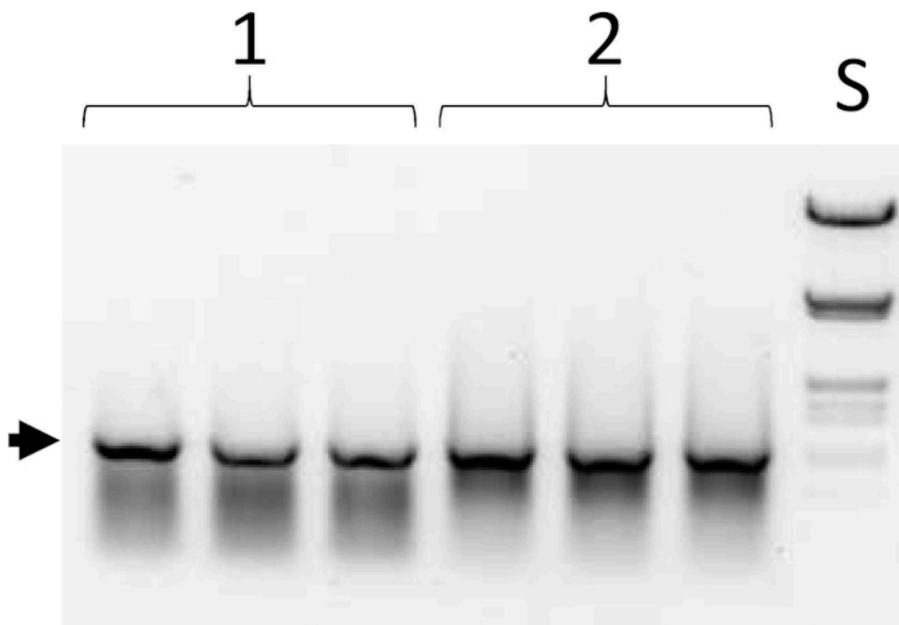

Figure 6: Products of the first PCR in the selection cycle. See **Figure 1B**, step III; protocol step 5.10. Column sets 1 and 2 are aliquots of two different libraries loaded in triplicate. S-DNA size standard (lambda DNA digested with HindIII and EcoRI). Arrow indicates position of the full-length ESC (1,050 bp). Please click here to view a larger version of this figure.

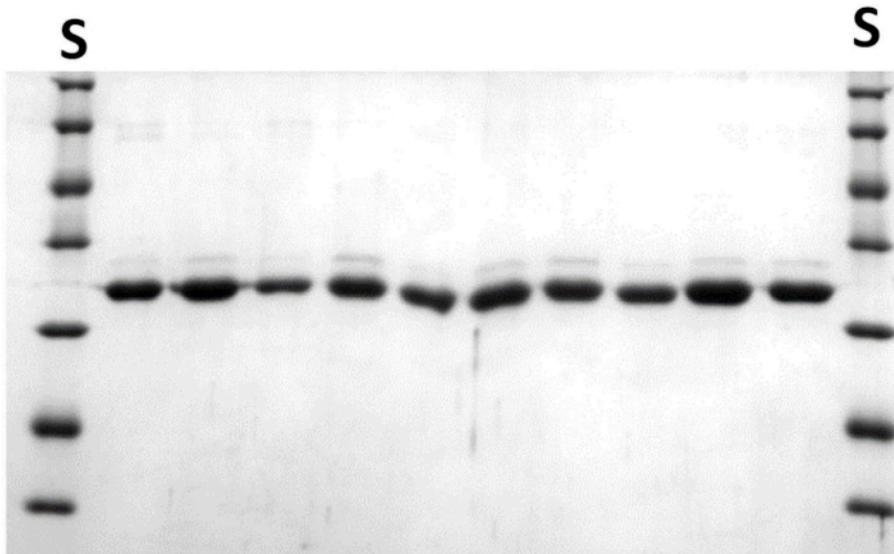**Figure 7: NlaIV variants purified for further screening in mini scale.** See step 6.3.11. Each line contains a 10 µL aliquot of a different variant. S-protein molecular weight standard. Molecular mass of NaIV REase subunit is 29.9 kDa. Please click here to view a larger version of this figure.
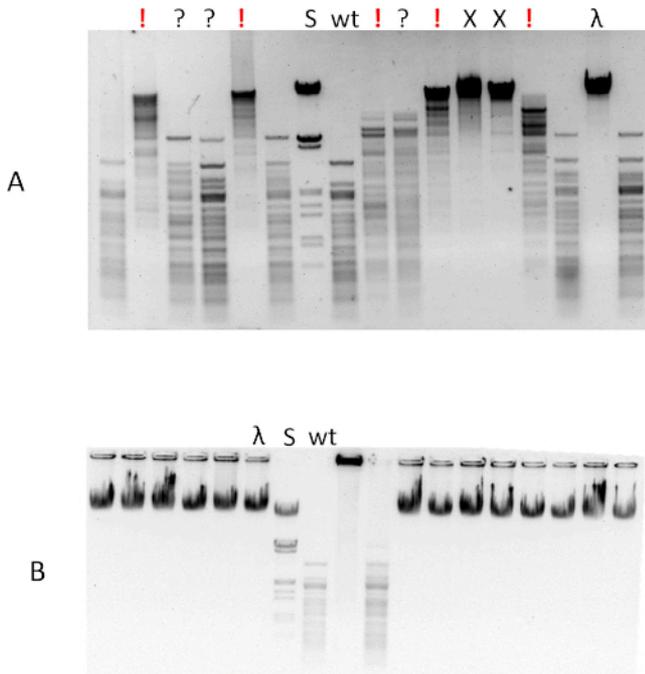

Figure 8: Examples of screening of NlaIV variants for sequence specificity alteration. See step 6.4.2. (**A**) Successful screening with high frequency of promising variants. S = DNA size marker, lambda DNA cleaved with HindIII and EcoRI; wild type (wt) = lambda DNA cleaved with wild type NlaIV; λ = lambda DNA substrate, not cleaved; other columns=variants with very low activity. Variants are labeled ! = promising variants that produce a cleavage pattern distinct from the wild type enzyme; ? = variants that also might have altered sequence preference. (**B**) Unsuccessful screening, with a majority of variants inactive and one variant with apparently unaltered cleavage pattern. Please click here to view a larger version of this figure.

Figure 9: Alternative selection by ligation. This alternative can be used for all REases generating sticky ends. Here we present an example protocol for a selection scheme for MwoI enzyme (unpublished). I) Selected sequence (located at the right end of the ESC) with defined residues shown in red and selected variation of the cognate sequence shown in blue. In parentheses below the counter selected sequence to be placed at the left end of the ESC is shown; II) Product of MwoI cleavage; III) After terminating in vitro transcription/translation, products are purified and ligation is performed with excess adapter. Only the cleavage products that were cleaved in the selected sequence can participate in ligation. Therefore, inactive variants are eliminated, and the pulldown step is unnecessary. The cleavage product in the counter selected sequence (on left end of the ESC, not shown) cannot participate in this ligation because the protruding end of the adaptor is not complementary to the counter selected sequence; IV) Selective PCR uses the same strategy as in the main protocol to eliminate variants with the wild type degenerate sequence specificity (F1 primer binding distal to the counter selected site) whereas inactive variants are eliminated by the selective reverse primer that cannot bind to the uncleaved (and therefore not modified by adapter ligation) right end. In the next cycle the process can be iterated by using adaptor that is identical to the cleavage product of the preceding step (i.e., the "cleaved cassette" in panel III), and an appropriate selective reverse primer. Please click here to view a larger version of this figure.

Table 1: Primers used in NlaIV engineering. Sequences of the restriction sites mentioned in the comments are underlined. Small letters indicate sequences that do not have complements in the DNA templates. Please click here to view this table (Right click to download).

Table 2: Conditions of PCR reactions to be used in the protocol. T_m = primer melting temperature (if T_m is different for the primers, the lower T_m should be used). Please click here to view this table (Right click to download).

Table 3: Results of quality check of two mutagenic primers synthesized with split-and-mix strategy. Mutagenized codons are indicated with [XXX]. A lower index number indicates the position of an encoded amino acid. Adapted from Czapinska et al.[8] with permission from Elsevier. Please click here to view this table (Right click to download).

Table 4: Results of EP-PCR. Main parameters derived from sequence analysis of 22 clones of ECS. Please click here to view this table (Right click to download).

Discussion

The selection protocol described here was tested for NlaIV[8], a dimeric PD-(D/E)XK fold recognition sequence that recognizes a palindromic target site with central NN bases and catalyzes a blunt end cut between the NN bases. NlaIV was picked because cleavage between the NN bases suggests that these bases are close to the protein in the complex. In principle, the protocol could be used for any sequence specific restriction endonuclease, monomeric or dimeric, of any fold group, catalyzing double strand breaks of any stagger, irrespective of whether catalytic and specificity domains coincide (as in the NlaIV example) or are separate (e.g., FokI). Moreover, the protocol in principle is useful not only for the generation of new, more narrow enzyme specificity, but could also be used to eliminate star activities, or to create high fidelity endonucleases. However, all this has not been tested yet. In particular, targeted elimination of star activity may be complicated, because the same amino acid residues could be involved in binding to the desired and undesired bases. The in vitro steps described in this protocol are not limited to the selection of narrowed down specificities but could also be used to select otherwise altered specificities. However, there is then a problem with variant endonucleases: if the spectrum of substrates includes novel targets not cleaved by the parental endonuclease, there is in general no good way to protect cells from the harmful effects of this activity. In contrast, if endonuclease specificity is only narrowed down, the targets are a subset of the wild type targets, and hence the already available cognate methyltransferase should be fully protective.

Our protocol differs in several respects from many directed evolution protocols. Open reading frame diversity is generated once at the beginning of the experiment, not in every iteration. Moreover, it is created by split-and-mix synthesis, rather than by EP-PCR. For NNS substitutions of codons, as used in this work, there are $(4 \times 4 \times 2)^6 \sim 1.07 \times 10^9$ combinations for six positions. Therefore, any given variant is present on average once in 1.7 fmoles of ESC. This capacity can be increased to seven positions by using synthesis with a mixture of 20 trinucleotide precursors that is offered by Glen Research or by decreasing mutation frequency in less promising positions with split-and-mix oligonucleotide synthesis.

If possible, it is recommended to limit the extent of variation to six positions. Obviously, such mutagenesis targeting requires some preexisting knowledge about at least the regions of the REase involved in substrate binding. The split-and-mix protocol to generate diversity has clear advantages in comparison to EP-PCR. Using EP-PCR, we obtained unchanged variants and sequences carrying eight substitutions for NlaIV ESCs in the same EP-PCR (**Table 4**). The library from EP-PCR contains a substantial fraction of clones that should be avoided (wild type sequences, multiple substitutions, frameshift and nonsense mutations, and mutations in places unlikely to affect sequence specificity).

Our protocol also differs from many other directed evolution protocols by the presence of two sequential selection steps. Positive selection makes sure that the desired activity is retained, otherwise the biotin tag is not removed, and the coding sequence can be removed by pull-down. It is technically possible that the fortuitous emergence of a novel, non-overlapping specificity (e.g., GCATGC) could lead to severing of the biotin tag as well, if a suitable cleavage site is present near the desired cleavage, but not elsewhere. However, this should be highly unlikely. Negative selection removes open reading frames that code for enzymes that still have the undesired activity. This step is not strictly mandatory, because the protocol will still enrich the output library with variants that are able to cleave the selection sequence but not able to cleave elsewhere in the ESC, therefore rendering it unsuitable for PCR amplification. However, selection effectiveness is expected to be lower because enzymes with the original sequence specificity will not be removed from the output and will outcompete promising variants with altered specificity but also decreased enzymatic activity. Note that at the population level, both desired and undesired target sequences can, but need not be, degenerate. In the NlaIV example, the anti-target was degenerate and the target non-degenerate. Even when there is degeneracy at the population level, in a single droplet only one (non-degenerate) target or anti-target is present. In our protocol, target and anti-target sequences are reintroduced at every repetition of the selection steps. Therefore, an open reading frame must encode an enzyme capable of cleaving all possible targets, and unable to cleave any of the anti-targets, to survive multiple selection rounds. Notice that the need to reintroduce the antiselection target at each iteration of the protocol enforces two sequential PCRs. The first PCR uses a primer that anneals outside the anti-target, so that cleavage of the anti-target prevents the PCR reaction. The second PCR requires a primer that reaches beyond the anti-target, and reintroduces anti-target, to make sure that during multiple rounds of selection, each open reading frame is tested against all variants of the anti-target.

For enzymes that generate sticky ends, a related alternative protocol based on a previously described method for isolation of REase ORF[10] can be used. The depletion of inactive variants by biotin capture that is used in our experiments is replaced in the alternative protocol by ligation of the compatible adapter with a sequence that is used as a primer binding site in a selective PCR (**Figure 9**). Only ESCs that produce enzymes with the selected specificity generate ligation-capable ends and will therefore be selected. The sequence of the sticky end of the counter eselected sequence must be designed in such a way that it cannot participate in ligation with adapters. Iteration of the selection process can be easily achieved by switching between two different adapters and consequently two different reverse primers in selective PCR.

Even with new protocols, the task of engineering novel specificities in vitro is still very challenging. For typical type II REases, sequence specificity and endonucleolytic activity depend on the same protein regions. It is therefore difficult to alter one without affecting the other. Success is made more likely by a strategy that takes into account the footprint of the enzyme, respects the symmetry of protein-DNA interactions, and builds on preexisting enzymatic preferences, which should be determined upfront in biochemical experiments, as was done for the NlaIV example[8].

Disclosures

The authors have nothing to disclose.

Acknowledgments

References

1. Schöttler, S., Wenz, C., Lanio, T., Jeltsch, A., Pingoud, A. Protein engineering of the restriction endonuclease EcoRV--structure-guided design of enzyme variants that recognize the base pairs flanking the recognition site. *European Journal of Biochemistry.* **258** (1), 184-191 (1998).
2. Wenz, C., Hahn, M., Pingoud, A. Engineering of variants of the restriction endonuclease EcoRV that depend in their cleavage activity on the flexibility of sequences flanking the recognition site. *Biochemistry.* **37** (8), 2234-2242 (1998).
3. Samuelson, J. C., Xu, S. Y. Directed evolution of restriction endonuclease BstYI to achieve increased substrate specificity. *Journal of Molecular Biology.* **319** (3), 673-683 (2002).
4. Samuelson, J. C. et al. Engineering a rare-cutting restriction enzyme: genetic screening and selection of NotI variants. *Nucleic Acids Research.* **34** (3), 796-805 (2006).
5. Rimseliene, R., Maneliene, Z., Lubys, A., Janulaitis, A. Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco57I to select the mutant with a novel sequence specificity. *Journal of Molecular Biology.* **327** (2), 383-391 (2003).
6. Morgan, R. D., Luyten, Y. A. Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity. *Nucleic Acids Research.* **37** (15), 5222-5233 (2009).
7. Skowronek, K., Boniecki, M. J., Kluge, B., Bujnicki, J. M. Rational engineering of sequence specificity in R.MwoI restriction endonuclease. *Nucleic Acids Research.* **40** (17), 8579-8592 (2012).
8. Czapinska, H. et al. Crystal Structure and Directed Evolution of Specificity of NlaIV Restriction Endonuclease. *Journal of Molecular Biology.* **431** (11), 2082-2094 (2019).
9. Miller, O. J. et al. Directed evolution by in vitro compartmentalization. *Nature Methods.* **3** (7), 561-570 (2006).
10. Zheng, Y., Roberts, R. J. Selection of restriction endonucleases using artificial cells. *Nucleic Acids Research.* **35** (11), e83 (2007).

11. Takeuchi, R., Choi, M., Stoddard, B. L. Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. *Proceedings of the National Academy of Science U. S. A.* **111** (11), 4061-4066 (2014).

12. Howland, J. L. *Short Protocols in Molecular Biology.*, Third edition: Edited by F. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith. K. Struhl. John Wiley & Sons, New York (1995).

13. Wilson, D. S., Keefe, A. D. Random mutagenesis by PCR. In *Current Protocols in Molecular Biology.* Chapter 8 Unit 8.3. John Wiley & Sons, New York (2001).