17

Preparation of SCRATCHY Hybrid Protein Libraries

Size- and In-Frame Selection of Nucleic Acid Sequences

Stefan Lutz and Marc Ostermeier

1. Introduction

SCRATCHY is a combination of the incremental truncation for the creation of hybrid enzymes (ITCHY) technology ***(1)*** and DNA shuffling ***(2)***. It generates combinatorial libraries of hybrid proteins consisting of multiple fragments from two or more parental DNA sequences with no restriction to DNA sequence identity between the original sequences ***(3)***. Such multi-crossover hybrids can be of interest to the studies of fundamental questions of protein evolution and folding, as well as to the tailoring of enzymes for therapeutic and industrial applications.

The experimental implementation of SCRATCHY consists of two successive steps, an initial creation of an ITCHY library, followed by a homologous recombination procedure such as DNA shuffling (*see* **Fig. 1**). In the process, the ITCHY library serves as an artificial family of hybrid sequences that, upon fragmentation, provides a variety of crossover-carrying templates. During the random reassembly step, these templates can anneal with one another, leading to the introduction of two or more crossovers per shuffled sequence independent of the sequence homology in any particular region.

Two experimentally critical steps of SCRATCHY lie at the interface of the ITCHY and DNA shuffling protocols. First, a significant fraction of the incremental truncation library members contain large sequence insertions and deletions. Presumably, these deviations from the parental-size gene sequences are likely detrimental to the library and should be removed by size selection. Second, the blunt-end fusion of the *N*- and C-terminal fragment of the truncation library results in the disruption of the correct nucleotide-codon reading frame

From: *Methods in Molecular Biology, vol. 231: Directed Evolution Library Creation: Methods and Protocols*
Edited by: F. H. Arnold and G. Georgiou © Humana Press Inc., Totowa, NJ

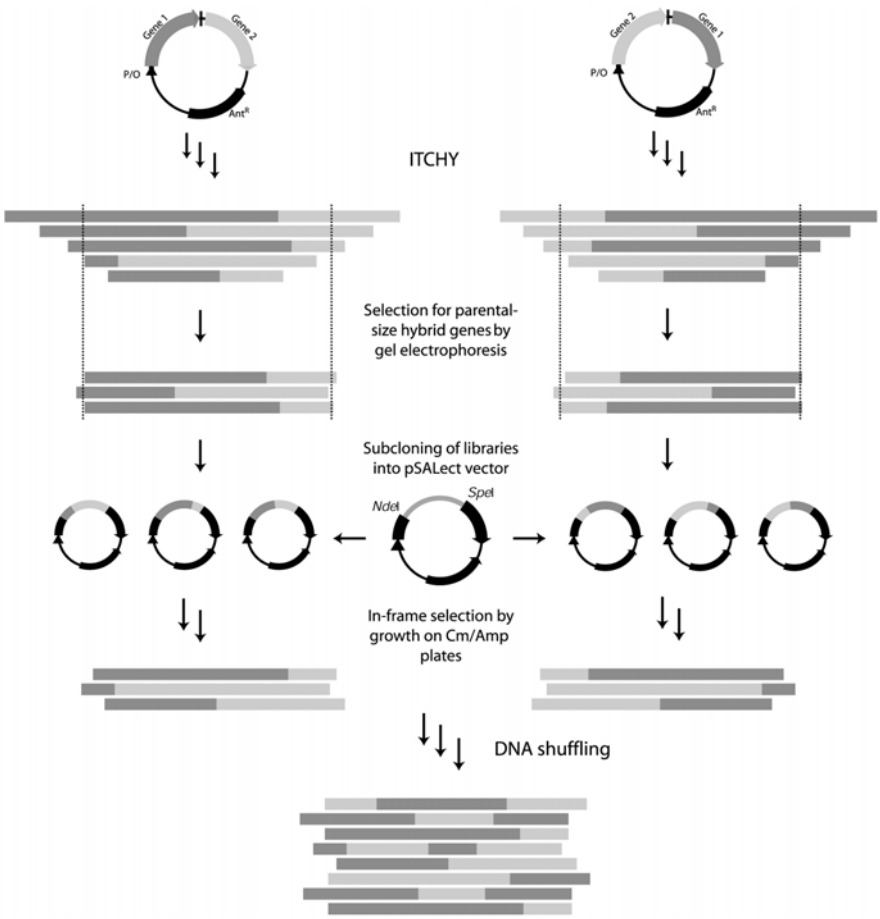

Fig. 1. Schematic view of the SCRATCHY protocol. The method requires two complementary vectors, carrying the genes A and B in alternating order. Following the generation of the ITCHY library, the linearized hybrid genes are selected for parental-size hybrid DNA constructs. After subcloning these DNA fragments into the *Nde*I and *Spe*I sites of pSALect, sequences with the correct reading frame result in the expression of a trifunctional fusion protein which renders the host cells resistant to ampicillin. The plasmid DNA from colonies grown under these conditions is recovered and can be used as starting material for DNA shuffling.

at the crossover in two-thirds of the library members. While not crucial for the preparation of ITCHY libraries, the presence of such out-of-frame fusions will rapidly diminish the fraction of hybrids with the overall correct reading frame upon DNA shuffling. For example, the probability of a hybrid with two fusion

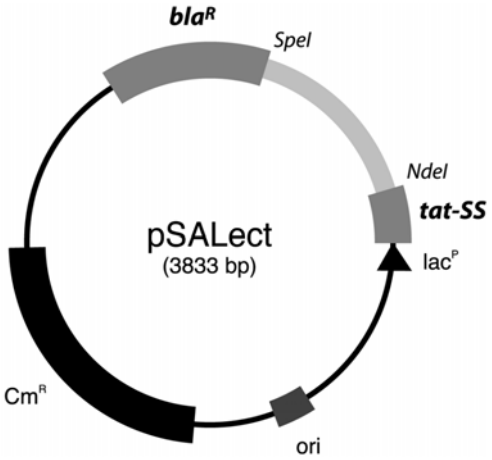

Fig. 2. The pSALect vector is a pBC-SK+ based plasmid for in-frame selection of DNA sequences. The gene of interest is cloned through *Nde*I and *Spe*I sites that are flanked *N*-terminally by the Tat-signal sequence (tat-SS) and C-terminally by the gene encoding for β-lactamase (bla^R) without its natural signal sequence. The expression of the fusion construct is under the control of a *lac*-promoter (lac^P). As a general selection marker, the plasmid carries the gene for chloramphenicol-acetyl-transferase (Cm^R).

points maintaining the correct reading frame is only 1/9 (1/3 × 1/3) while for a hybrid with three crossovers, the chances drop to 1/27 (1/3 × 1/3 × 1/3). Addressing these problems, we have developed and implemented a two-step post-ITCHY protocol to remove fragments with large insertions and deletions and out-of-frame constructs.

This chapter describes the two-step protocol for DNA fragment size- and reading frame-selection of ITCHY libraries prior to DNA shuffling. Initially, the linearized version of the plasmid DNA carrying the ITCHY library is separated by agarose gel electrophoresis. The desired size-range of fragments is excised and recovered. Next, the size-selected sublibrary is cloned into pSALect and in-frame selection is performed. The pSALect vector is specifically designed to allow for the isolation of nucleic acid sequences with the correct reading frame ***(4)***. The target sequence is cloned between an *N*-terminal Tat signal sequence and a C-terminal lactamase gene (*see* **Fig. 2**). Upon expression, the target peptide acts as a linker between the two selection markers. For in-frame fusions, this tri-functional fusion facilitates the successful export of the lactamase into the periplasm, rendering the host cell resistant to antibiotics in the growth media. This strategy eliminates the false positives, presumably arising from internal translational initiation sites, observed in in-frame selection

strategies using *N*-terminal fusions to GFP ***(5)***, chloramphenicol-acetyl-transferase ***(6)*** and aminoglycoside phosphotransferase APH(3')-IIa (Ostermeier, M., Lutz, S., and Benkovic, S.J., unpublished results), since functional expression of the lactamase requires export to the periplasm. After in-frame selection, the plasmid DNA from the selected candidates is recovered and can be used for DNA shuffling.

2. Materials

1. QIAquick PCR and gel purification kits (Qiagen; Valentia, CA) including buffers PB and EB.
2. Restriction endonucleases (with reaction buffers).
3. Glycerol.
4. Chloramphenicol.
5. Ampicillin.
6. T4 DNA ligase.
7. 10X T4 DNA ligase buffer: 500 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 100 m*M* dithiothreitol, 10 mM ATP, 250 μg/mL bovine serum albumin (BSA).
8. Polyethyleneglycol 45% (w/w): Mix 45 g of polyethylene glycol (MW 6000) and 55 g of water. Autoclave at 121°C for 20 min and let cool to room temperature.
9. Shrimp alkaline phosphatase: USB Corporation (Cleveland, OH).
10. Agarose gels: NuSieve GTG and SeaKem GTG agarose (both from BioWhittaker; Rockland, ME) in 1X TAE buffer (40 m*M* Tris-acetate, 1 m*M* EDTA) and 0.5 μg/mL ethidium bromide.
11. 100 bp-DNA Step Ladder: Promega (Madison, WI) stored at –20°C.
12. Electrocompetent *E. coli* strain DH5α-E (Invitrogen; Carlsbad, CA) stored at –80°C.
13. Luria Bertani Broth (LB) medium: Mix (per liter) 5 g yeast extract, 10 g tryptone, and 10 g NaCl. Autoclave at 121°C for 20 min and let cool to room temperature.
14. LB-agar medium: 5 g yeast extract, 10 g tryptone, 10 g NaCl, and 15 g agar. Autoclave at 121°C for 20 min and let cool to 50°C prior to addition of antibiotic selection marker. Plates can be stored at 4°C for up to 4 wk.
15. SOC medium: Mix (per liter) 5 g yeast extract, 20 g tryptone, 0.5 g NaCl, 2.5 m*M* KCl. Autoclave at 121°C for 20 min and let cool to room temperature. Autoclave separately 1 *M* $MgCl_2$, 1 *M* $MgSO_4$, and 20 % (w/v) glucose. Add 10 m*M* $MgCl_2$, 20 m*M* $MgSO_4$, and 20 m*M* glucose to the media. Store at room temperature.
16. pSALect plasmid DNA vector: Stefan Lutz (Emory University, Atlanta, GA)
17. Deoxynucleoside triphosphates (dNTPs): 2.5 m*M* per dNTP stock solution. Aliquots are stored at –20°C.
18. Oligonucleotide primers: 100 μ*M* stock solutions.
19. *Taq* DNA polymerase.
20. 10X *Taq* DNA polymerase buffer: 500 m*M* KCl, 100 m*M* Tris-HCl, pH 9.0, 1.0% Triton X-100.

3. Methods

In the preparation of an incremental truncation library for DNA shuffling, two additional, consecutive steps are highly recommended: a) the removal of truncated fragments with large insertions and deletions and b) the isolation of single-crossover hybrids that preserve the correct reading frame.

The following protocol builds on the chapter on the creation of ITCHY hybrid protein libraries (*see* **Chapter 16**). The starting material used in the procedure originates from an ITCHY experiment using nucleoside analogs to create the incremental truncation library. The protocol starts after the blunt-ending of the library by Klenow fragment DNA polymerase (**Chapter 16**; **Subheading 3.2.**, **step 15**). Throughout the protocol, we utilize Qiagen's QIAQuick PCR purification and gel extraction kits to purify and desalt DNA solutions. DNA purification kits from other manufacturers presumably work equally well but they have not been tested in our laboratory.

3.1. Size Selection of DNA Fragments

1. Purify the heat-denatured reaction mixture of the incremental truncation library by agarose gel electrophoresis: Pour a 1% agarose gel with two wells, sufficiently large to load the entire truncation library (*see* **Note 1**), Load aliquots of the reaction mixture in well after mixing them 1:1 with glycerol and run gel at a low voltage, Visualize the DNA library smear under UV light and excise the gel block that contains DNA product, corresponding to the desired size range (*see* **Note 2**). Recover DNA from gel matrix, using Qiagen's QIAquick gel purification kit (for details see manufacturer's protocol).
2. Ligate the product in 400 μL, containing 1X T4 ligase buffer (provided with enzyme), 4.5% PEG-6000, and 5 Weiss-units of T4 DNA ligase per μg of DNA, by incubating the mixture at 16°C overnight (*see* **Note 3**).
3. Quench reaction by addition of 3 volumes of QG buffer (provided in Qiagen's QIAquick gel extraction kit) and recover DNA with spin columns according to the manufacturer's protocol. Elute DNA in a final volume of 50 μL EB buffer (provided in the QIAquick kit).
4. Transform 2-μL aliquots of the library into electrocompetent *E. coli* (*see* **Note 4**):
   a. Thaw one frozen 50-μL aliquot of competent cells for each sample by storing the vials on ice.
   b. Add up to 20 ng of sample DNA per aliquot of competent cells, then transfer the cell suspension into an ice-cold electroporation cuvette, and store the cuvette on ice for 1 min.
   c. Load the cuvet into the electroporator and apply an electrical pulse (for optimal settings of electroporator, contact manufacturer's manual).
   d. Remove cuvet from electroporator and immediately add 1 mL of SOC medium. Incubate the cell suspension at 37°C for 1 h.

e. Dilute a two-microliter culture sample in 1 mL of liquid LB media and plate a 50 μL aliquot on LB-agar plates containing the appropriate antibiotics (*see* **Note 5**).
f. Plate the remaining cell suspension onto LB-agar library plates (245 mm) containing the appropriate antibiotic selection marker and incubate the plates at 37°C overnight (*see* **Note 6**).

5. Carefully suspend the colonies from the agar plate into 2X 10 mL of LB medium. Collect the cell suspension and pellet the culture by centrifugation at 1500*g* at 4°C for 15 min.
6. Resuspend the pellet in a mix containing 1.5 mL of LB medium and 0.5 mL of glycerol.
7. Pipet 100 μL of the cells into an Eppendorf tube and pellet the suspension (*see* **Note 7**).
8. Remove the supernatant and isolate the plasmid DNA from the pellet, using Qiagen's QIAprep Mini Prep kit (for details see manufacturer's protocol).
9. Spectrophotometrically determine the final DNA concentration of all samples (*see* **Note 8**).
10. Set up 50 μL PCR reactions containing 1X *Taq* DNA polymerase buffer, 1.25 m*M* $MgCl_2$, 0.5–1 μ*M* of primer A and B (*see* **Note 9**), ~10 ng of linearized template vector, 5 units of *Taq* DNA polymerase, and 200 μM dNTP.
11. Amplify template over 30 cycles of PCR (94°C for 30 s, T_A for 30 s, 72°C for 1.2 min/kb) (*see* **Note 10**).
12. Check the quality of the PCR products by running 1–2 μL of reaction mixture on an agarose gel next to a suitable size marker. Purify the remaining reactions with Qiagen's QIAquick PCR purification kit.
13. Spectrophotometrically determine the final DNA concentration of all samples (*see* **Note 8**).

3.2. Reading Frame Selection of DNA Fragments

1. Digest 10 μg of the PCR product (DNA library) and 1.5 μg of pSALect vector with 20 units of *Nde*I and 20 units of *Spe*I restriction enzymes at 37°C for 6–8 h.
2. Dephosphorylate the pSALect vector by addition of 0.5 μL of shrimp alkaline phosphatase and continue incubation for another hour at 37°C.
3. Purify the reaction mixtures by agarose gel electrophoresis: Pour a 1% agarose gel with sufficiently large wells to load entire digests; Load digests in wells after mixing it with loading dye and run gel; Visualize the bands of DNA under UV light and excise product bands that correspond in size to the expected fragments with a razor blade; Recover DNA from the gel matrix, using Qiagen's QIAquick gel purification kit (for details see manufacturer's protocol) and elute DNA with 50 μL of Qiagen's EB buffer.
4. Ligate 10 μL of vector DNA with 10 μL of hybrid DNA in a 25-μL reaction with 1X T4 DNA ligase buffer and 6 Weiss units of T4 DNA ligase at room temperature overnight.
5. Transform aliquots of the ligation mixture into electrocompetent *E. coli* (*see* **Note 11**): Thaw one frozen 50-μL aliquot of competent cells for each sample by

storing the vials on ice; Add 4 μL of ligation mixture per aliquot of competent cells, then transfer the cell suspension into an ice-cold electroporation cuvet, and store the cuvet on ice for 1 min. Load the cuvet into the electroporator and apply an electrical pulse (for optimal settings of electroporator, contact manufacturer's manual); Remove cuvet from electroporator and immediately add 1 mL of SOC medium. Incubate the cell suspension at 37°C for 1 h. Dilute a two-microliter culture sample in 1 mL of liquid LB media and plate a 50 μL aliquot on LB-agar plates containing 50 μg/mL chloramphenicol (*see* **Note 5**); Plate the remaining cell suspension onto LB-agar library plates (245 × 245 mm) containing 50 μg/mL chloramphenicol and incubate the plates at 37°C overnight.

6. Carefully suspend the colonies from the agar plate into 2X 10 mL of LB medium and collect the cell suspension.
7. Measure the optical density (OD) of the cell suspension at 600 nm.
8. Dilute cell suspension to approx 1×10^6 cells/mL, using LB medium (*see* **Note 12**).
9. Plate 2 mL of cell culture on LB-agar library plates (245 × 245 mm) containing 100 μg/mL ampicillin and incubate the plates at room temperature for 2–4 d (*see* **Note 13**).
10. Dilute 2 μL of cell culture in 200 μL of LB medium, plate 50 μL aliquots on LB-agar plates containing either ampicillin (100 μg/mL) or chloramphenicol (50 μg/mL), and incubate the plates at room temperature for 2–4 d (*see* **Note 14**).
11. Carefully suspend the colonies from the agar plate into 2X 10 mL of LB medium. Collect the cell suspension and pellet the culture by centrifugation at 1500*g* at 4°C for 15 min.
12. Resuspend the pellet in a mix containing 1.5 mL of LB medium and 0.5 mL of glycerol.
13. Pipet 100 μL of the cells in an Eppendorf tube and pellet the suspension (*see* **Note 7**).
14. Remove supernatant and isolate the plasmid DNA from the pellet, using Qiagen's QIAprep Mini Prep kit (for details see manufacturer's protocol).
15. Spectrophotometrically determine the final DNA concentration of all samples (*see* **Note 8**).

At this point, the plasmid DNA can be directly used for DNA shuffling. (*See* related chapters in this volume.) Alternatively, the library of hybrid genes can be amplified with gene-specific primers and the PCR product used for the shuffling protocol (*see* **Note 15**).

4. Notes

1. In a typical experiment, we split the reaction mixture in two equal aliquots. One of the aliquots is mixed with 1 μL of size marker (100 bp-ladder) and glycerol is added to both samples. We prefer glycerol as a loading aid instead of regular loading dye, because of interference of DNA intensity and mobility at the positions of the dye colors in the gel. The two aliquots are then loaded next to one another on the gel and run at 50–80 V for 3–4 h (using BioRad's MiniSub Cell GT system). The endpoint of the electrophoresis is determined by the occasional

monitoring of DNA migration using UV light. An extended running time will result in better separation and will allow more narrow size selection.

2. The range of fragment sizes in the final product can be controlled by the length of time to run the electrophoresis and the size of the excised fragment. When selecting a narrow band width, we suggest to excise multiple bands above and below the target size. These fragments are then purified individually and their size distributions analyzed by subsequent ligation and DNA sequencing. It is our experience that size ranges of ±30 bp can be routinely achieved with a 4 kb vector, with a feasible limit of ±10 bp.
3. Crucial to the successful intramolecular ligation is to maintain an overall DNA concentration of <3 ng/μL *(7)*. Adjust the volume of the ligation reaction so that the DNA concentration falls below this threshold.
4. We have had consistently good results with strain DH5α-E. Transformation by electroporation is generally more efficient, generating larger libraries than chemical transformation, however either method works for transforming incremental truncation libraries.
5. These plates serve as controls for the library size. The number of colonies on the library plate(s) can be estimated by counting the control plates and multiplying the number of colonies by a factor of 10,000.
6. Library sizes can range from a few hundred colonies to several million members. We recommend to estimate the theoretical library size (theoretical size = $N*n$; maximum = N^2) whereby N= the number of nucleotides overlapping between the two parental genes or gene fragments, and n= the base pair range after size selection. For example, consider an incremental truncation library between two genes with a 150 bp overlapping region (N= 150) size-selected for fragments of parental size plus-minus 20 nucleotides (n= 40). The resulting theoretical library size would thus be 6000 members. For a maximal probability of representation of each crossover, the number of transformants should be approx 5–10 times the theoretical library size. With respect to the above example, the library plate should consist of at least 30,000 colonies.
7. We recommend to flash-freeze and store the remaining cell suspension in 100 to 200-μL aliquots at –80°C.
8. We use the following rule for concentration measurements: 1 A_{260} = 50 μg DNA.
9. Primer A and B are gene-specific forward- and reverse-primers. The PCR step is important to prepare the library for the subcloning into the pSALect vector. First, the primer design must mutate the 3'-terminal Stop-codon of the hybrid gene to a glycine codon (GGA). Simultaneously, the necessary *Nde*I site at the 5'-end of the sequence and the *Spe*I site on the 3'-end can be introduced with these primers.
10. The optimal annealing temperature for primer A and B must be determined prior to the amplification. We recommend primers of 20–25 bp length with T_A of 55–60°C.
11. We have had consistently good results with *E. coli* strain DH5α-E. Transformation by electroporation is generally more efficient, generating larger libraries than chemical transformation, however either method works.

12. We use the following rule as a guideline to calculate the cell density in cultures; 1 OD_{600} = 1 × 10^8 cells/mL.
13. The rate of cell growth varies widely. We obtained the best libraries by incubation of the library plates at room temperature. The lower temperature accommodates the expression of slow-folding proteins.
14. The diluted plates serve as controls. Plating approximately 500 colony forming units (cfu) per plate, all of them should grow in the presence of chloramphenicol but only a fraction will appear on the ampicillin-containing plate. For quality control, we recommend selecting random colonies from the ampicillin plate, growing them in liquid medium overnight, and isolating the plasmid DNA for sequence analysis to confirm that the crossovers they contain are in-frame.
15. Remember that the Stop codon needs to be reintroduced. This can be done during the PCR step prior to DNA shuffling. Alternatively, it can be reintroduced after the primerless reassembly step when amplifying with outside primers.

Acknowledgments

The described protocols were developed and implemented as part of the authors' postdoctoral work in Stephen J. Benkovic's laboratory at the Pennsylvania State University. We would like to acknowledge and thank him for his invaluable support and guidance.

References

1. Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17,** 1205–1209.
2. Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370,** 389–391.
3. Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S. J. (2001) Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Aci. USA* **98,** 11,248–11,253.
4. Lutz, S., Fast, W., and Benkovic, S. J. (2003) A universal, vector-based system for nucleic acid reading frame selection. *Protein Eng.* in press.
5. Waldo, S., Standish, B. M., Berendzen, J., and Terwilliger, T.C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17,** 456–460.
6. Sieber, V., Martinez, C. A., and Arnold, F. H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19,** 456–460.
7. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.