16

The Creation of ITCHY Hybrid Protein Libraries

Marc Ostermeier and Stefan Lutz

1. Introduction

Incremental truncation is a method for creating a library of every one base truncation of dsDNA. Incremental truncation libraries can be created by time dependent Exo III digestions ***(1)*** or by the incorporation of α-phosphorothioate dNTPs (αS-dNTPs) ***(2)***. The fusion of two incremental truncation libraries is called an ITCHY library ***(3)***. An ITCHY library created from a single gene consists of genes with internal deletions and duplications. An ITCHY library created between two different genes consists of gene fusions created in a DNA-homology independent fashion. ITCHY libraries, as well as an incremental truncation-like method called SHIPREC ***(4)***, have the potential to create proteins with improved or novel properties as well as to generate artificial families for in vitro recombination in a method called SCRATCHY ***(5)*** (*see* **Chapter 17**).

To create an ITCHY library, the two genes or gene fragments are cloned in tandem. In time-dependent truncation (*see* **Fig. 1**), the DNA is subjected to Exo III digestion starting from a unique restriction enzyme site between the two genes (RE1). During Exo III digestion, whose rate is control by the addition of NaCl, small aliquots are removed frequently and quenched by addition to a low pH, high salt buffer. Blunt ends are prepared by treatment with a single-strand nuclease and a DNA polymerase. This synchronized truncation library is uncoupled by digestion at a second restriction enzyme site (RE2) followed by religation. In truncation using αS-dNTPs (*see* **Fig. 2**), the entire plasmid is amplified by PCR using dNTPs and a small amount of αS-dNTPs. Subsequent digestion with Exo III is prevented from continuing past the randomly incorporated αS-dNMP ***(6)***. Blunt ends are prepared by treatment with a single-strand nuclease and a DNA polymerase followed by unimolecular ligation to recyclize the vector.

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

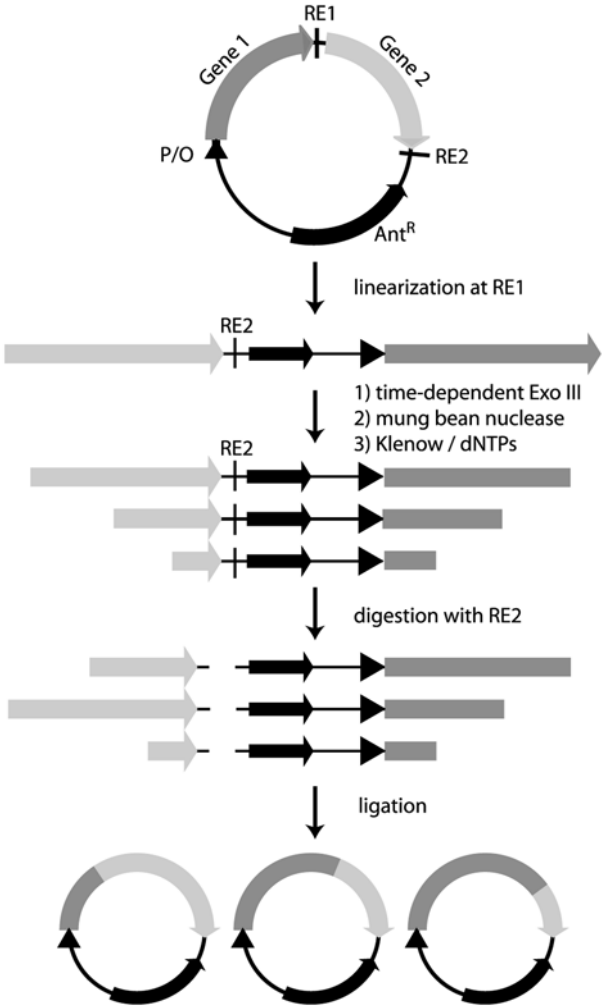

Fig. 1. Schematic of the creation of ITCHY libraries using time-dependent truncation. RE1 and RE2, unique restriction endonuclease sites; P/O, promoter; Ant^R, antibiotic resistance gene.

Each method has advantages and disadvantages. Time dependent truncation requires multiple time-point sampling to achieve a comprehensive library. Truncation using αS-dNTPs is less labor-intensive and offers the possibility of incorporating random mutagenesis in the PCR amplification. A theoretical treatment of both methodologies *(7)* predicts that time-dependent truncation offers higher control over the range of truncation and a higher probability of

parental length fusions in the library. However, truncations using αS-dNTPs are predicted to inherently produce a uniform distribution of parental length fusions across the gene regardless of experimental conditions whereas time-dependent truncation only produces such a uniform distribution under optimal experimental conditions.

2. Materials

1. QIAquick PCR and gel purification kits (Qiagen; Valentia, CA) including buffers PB and EB.
2. Restriction endonucleases (with reaction buffers).
3. 100X BSA: 10 mg/mL.
4. NaCl.
5. Exonuclease III (Exo III).
6. 10X Exo III buffer: 660 m*M* Tris-HCl, pH 8.0, 6.6 m*M* $MgCl_2$.
7. Mung Bean Nuclease.
8. 10X Mung bean buffer: 500 m*M* sodium acetate, pH 5.0, 300 m*M* NaCl, 10 m*M* $ZnSO_4$.
9. Klenow fragment DNA polymerase.
10. Klenow mix: 20 m*M* Tris-HCl, pH 8.0, 100 m*M* $MgCl_2$, 0.25 U/μL units Klenow.
11. 10X Klenow buffer: 100 m*M* Tris-HCl, pH 7.5, 50 m*M* $MgCl_2$. Store at room temperature.
12. Deoxynucleotide triphosphates (dNTPs): 2.5 m*M* per dNTP stock solution. Aliquots are stored at –20°C.
13. 100% Ethanol stored at –20°C.
14. 70% Ethanol (v/v) stored at –20°C.
15. Ammonium acetate: 7.5 *M* stored at 4°C.
16. T4 DNA ligase.
17. 10X T4 DNA ligase buffer: 500 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 100 m*M* dithiothreitol, 10 m*M* ATP, 250 μg/mL bovine serum albumin.
18. Agarose gels: SeaKem GTG agarose (BioWhittaker; Rockland, ME) in 1X TAE buffer (40 m*M* Tris-acetate, 1 m*M* EDTA) and 0.5 μg/mL ethidium bromide.
19. Oligonucleotide primers: 100 μ*M* stock solutions.
20. *Taq* DNA polymerase.
21. 10X *Taq* DNA polymerase buffer: 500 m*M* KCl, 100 m*M* Tris-HCl, pH 9.0, 1.0% Triton X-100.
22. α-Phosphorothioate dNTPs: 100 m*M* stock (Promega; Madison, WI) stored at –20°C.
23. 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.
24. Electrocompetent *E. coli* strain DH5α-E (Invitrogen; Carlsbad, CA) stored at –80°C.
25. Luria Bertans 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.

26. 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.
27. 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.

3. Methods

The methods described below outline 1) the construction of ITCHY libraries using time-dependent truncation, 2) the construction of ITCHY libraries using nucleotide analogs and 3) characterization of the ITCHY libraries. Throughout these methods, 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. Preparation of Incremental Truncation Libraries Using Time-Dependent Truncation

1. Construct a vector as shown in **Fig. 1**. The two parental gene sequences are cloned in series, using unique restriction sites in the vector (*see* **Note 1**).
2. Linearize 10 μg of non-nicked plasmid DNA (*see* **Note 2**) by digestion with a restriction enzyme RE1 (*see* **Note 3**) that recognizes a unique cleavage site between the two parental genes (*see* **Fig. 1**). Purify the reaction mixture using Qiagen's QIAquick PCR purification kit, eluting with 100 μL of buffer EB provided by the QIAquick kit. Determine the DNA concentration by absorbance at 260 nm using the relationship that 50 μg/mL gives an A_{260} of 1.0.
3. Equilibrate 300 μL of PB buffer (provided in Qiagen's QIAquick PCR purification kit) on ice in a 1.5-mL tube (tube A).
4. To a second 0.65-mL tube (tube B), add 2 μg of DNA from *step 2*, 6 μL of 10X Exo III buffer, and NaCl to the desired concentration (*see* **Note 4**), water to 60 μL and equilibrate at 22°C. There are useful controls that can be incorporated at this step (*see* **Note 5**).
5. At time = 0, add 100 units of Exo III per μg DNA to tube B and mix immediately and thoroughly.
6. At regular intervals (generally 20–30 s) remove small samples (~0.5–1 μL) and add to tube A (*see* **Note 6**). Mix tube A well. Note that all time points are removed to tube A, which is kept on ice. The rate of Exo III is very temperature dependent; thus, it is preferable to leave tube B open during the sampling to avoid warming the tube by repeated handling.
7. Purify the DNA using Qiagen's QIAquick PCR purification kit following the manufacturer's directions. Elute the DNA from the column with 47 μL of Buffer EB.
8. Add 5 μL of mung bean buffer (10X) and mix thoroughly.

9. Add 3 U of mung bean nuclease, mix thoroughly and incubate at 30°C for 30 min (*see* **Note 7**).
10. Add 250 µL of QIAquick buffer PB and purify the DNA using Qiagen's QIAquick PCR purification kit following the manufacturer's directions. Elute truncated DNA from QIAquick column with 82 µL of Buffer EB.
11. Equilibrate tube at 37°C.
12. Add 0.5 µL of dNTP stock solution and 10 µL of Klenow mix, mix thoroughly and incubate at 37°C for 5 min.
13. Inactivate Klenow by incubating at 72°C for 20 min.
14. Add 10 µL of the appropriate 10X restriction enzyme buffer for restriction enzyme RE2, 1 µL BSA (100X) and an appropriate number of units of restriction enzyme RE2 (*see* **Note 8**) and incubate at the appropriate temperature for 2 h. Heat-inactivate RE2 under the conditions recommended by the manufacturer.
15. Concentrate by ethanol precipitation: add 50 µL of ammonium acetate followed by 300 µL of 100% ethanol (*see* **Note 9**). Incubate on ice for 30 min. Centrifuge 10 min at 12,000*g*. Wash the pellet with 70% ethanol and centrifuge 2 min at 12,000*g*. Remove liquid by pipet, spin briefly again and remove all traces of liquid. Air-dry the pellet 10 min and resuspend in 17 µL water.
16. Add 2 µL of T4 DNA ligase buffer (10X) and 6 Weiss units of T4 DNA ligase and incubate at 15°C ≥12 h.
17. Add 30 µL of water and ethanol precipitate as in **step 15** into 20 µL of water. The DNA is now ready to be transformed into the desired host.

3.2. Preparation of Incremental Truncation Libraries Using Nucleotide Analogs

1. Construct a template vector as shown in **Fig. 2**. The two parental gene sequences are cloned in series, using unique restriction sites in the vector (*see* **Note 10**).
2. Linearize the template vector (~3 µg) by digestion with a restriction enzyme that recognizes a unique cleavage site between the two parental genes (RE in **Fig. 2**). Purify the reaction mixture by agarose gel electrophoresis (*see* **Note 11**): Pour a 1% agarose gel with sufficiently large well to load entire digest. Load digest in well after mixing it with loading dye and run gel. Visualize the bands of DNA under UV light and excise product band that corresponds in size to the linearized vector with a razor blade. Recover DNA from gel matrix, using Qiagen's QIAquick gel purification kit.
3. Set up 50 µL PCR reactions containing 1X *Taq* DNA polymerase buffer, 1.25 m*M* $MgCl_2$, 0.5–1 µ*M* primer A and B (*see* **Fig. 2**), ~10 ng linearized template vector, 5 units *Taq* DNA polymerase (*see* **Note 12**), and 200 µ*M* dNTP (control experiment), 180 µ*M* dNTP / 20 µ*M* α-phosphorothioate dNTP, or 175 µ*M* dNTP/25 µ*M* α-phosphorothioate dNTP. Amplify template over 30 cycles of PCR (94°C for 30 sec, T_A for 30 sec, 72°C for 1.2 min/kb) (*see* **Note 13**). For all subsequent steps, the control experiment (dNTP only), as well as the individual dNTP/αS-dNTP experiments are treated in parallel.

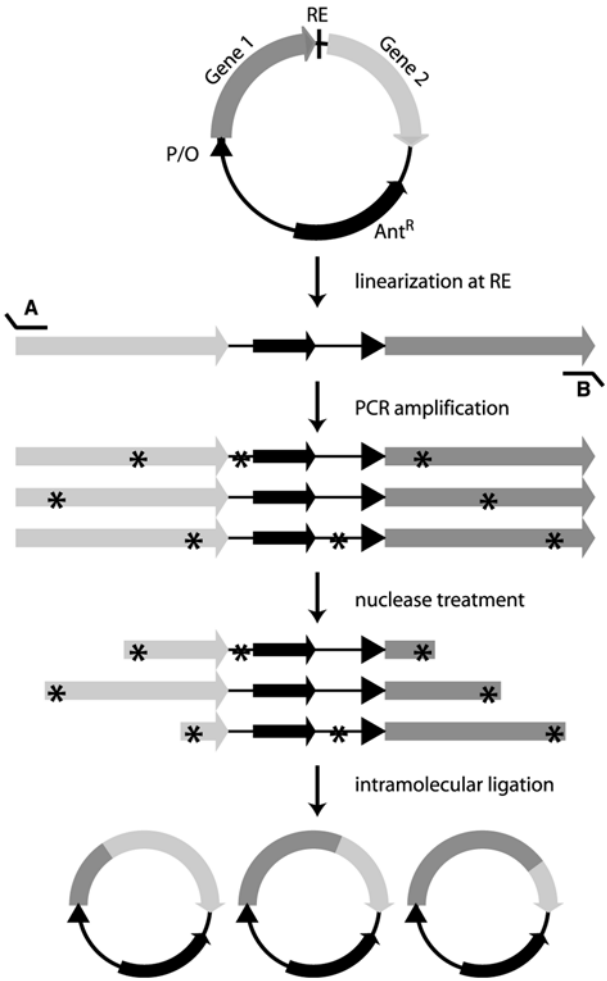

Fig. 2. Schematic of the creation of ITCHY libraries using nucleotide analogs. RE, unique restriction endonuclease sites; P/O, promoter; Ant^R, antibiotic resistance gene; *, site of nucleotide analog incorporation. **A** and **B**, site-specific amplification primers.

4. Check 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 (*see* **Note 14**).
5. Spectrophotometrically determine the final DNA concentration of all samples (*see* **Note 15**).
6. Dilute PCR product to approx 0.1 μg/μL of DNA in 1X Exo III buffer, using water and 10X Exo III buffer. The solution is centrifuged (at 5000*g* at room temperature for 1 min).

7. Add Exo III (120 units per μg of DNA) and digest the reaction mixture at 37°C for 30 min (*see* **Note 16**).
8. Quench reaction by addition of 5 volumes of PB buffer (provided in Qiagen's QIAquick PCR purification kit) and recover the DNA with a spin column according to the manufacturer's protocol (*see* **Note 17**).
9. Dilute DNA solution to approx 0.1 μg/μL of DNA in 1X mung bean nuclease buffer, using water and 10X mung bean nuclease buffer.
10. Add 2.5 units of mung bean nuclease per μg of DNA. Briefly centrifuge the samples and incubate them at 30°C for 30 min.
11. Quench reaction by addition of 5 volumes of PB buffer (provided in Qiagen's QIAquick PCR purification kit) and recover DNA with spin columns according to the manufacturer's protocol (*see* **Note 17**).
12. Mix DNA solution to approx 0.1 μg/μL of DNA in 1X Klenow buffer, using water and 10X Klenow buffer, spin briefly, and incubate 15 min at 37°C (*see* **Note 18**).
13. Add 1 unit of Klenow fragment DNA polymerase per μg of DNA and continue incubation for 5 min at 37°C (*see* **Note 18**).
14. Add dNTP stock solution to a final concentration of 100 μ*M* and continue incubation at 37°C for 10 min (*see* **Note 18**).
15. Quench reaction by heat-denaturation for 20 min at 75°C. After equilibration at room temperature, quality control can be performed by running an aliquot of the reaction mixture on an agarose gel next to a suitable size marker (*see* **Note 19**).
16. Ligate the remaining truncation 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 20**).
17. 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). The DNA is now ready to be transformed into the desired host.

3.3. Characterization of Incremental Truncation Libraries

1. Transform aliquots of library into electrocompetent *E. coli* (*see* **Note 21**). (a) Thaw one 50 μL-aliquot of competent cells on ice for each sample. (b) Add up to 20 ng of sample DNA per aliquot of competent cells, then transfer the cell suspension into an ice-cold electroporation cuvet and store the cuvettes 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) Plate various volumes of cell suspension (5–100 μL) onto LB-agar plates containing the appropriate antibiotic selection marker and incubate the plates at 37°C overnight.
2. Assess the apparent library size by counting the colony numbers on the culture plates, considering the applied dilutions.

3. Determine the fraction of the library containing inserts in the desired range by colony-PCR (*see* **Note 22**) as follows: Pick 25–50 colonies with sterile toothpicks and suspend cells in 10 μL PCR reaction mixtures (*see* **Note 23**). Initiate amplification with a five-min denaturing/cell lysis step at 95°C, followed by the regular cycling protocol. Mix reaction product with gel loading dye and run an aliquot on an agarose gel next to an appropriate size marker. Count the fraction of samples that generated a PCR product within the desired size range (*see* **Note 24**).
4. Analyze the data from the colony-PCR with respect to actual library size and the fragment size distribution of the library. (*see* **Note 25**) The position of the precise fusion point in selected sequences can be determined by DNA sequencing.

4. Notes

1. There are no restrictions on the vector that are particular to the construction of ITCHY libraries except for the uniqueness of the restriction enzyme site between the cloned genes and that it's digestion produces 3' ends that are susceptible to Exo III digestion (e.g., has blunt or 3' recessed end).
2. Generally we use plasmid DNA that is ~90% or greater supercoiled. We routinely isolate plasmid DNA of such quality from *E. coli* strain DH5α using commercial plasmid-prep kits. We have found that the fraction of nicked molecules can depend on growth conditions ***(8)***.

   Should production of sufficiently pure non-nicked DNA prove difficult, methods to purify non-nicked from nicked plasmid DNA include CsCl-ethidium bromide gradients ***(9)***, acid-phenol extraction ***(10)***, or removal of the nicked DNA by enzymatic digestion ***(11)***. Alternatively, treatment of nicked DNA with T4 DNA ligase should presumably repair the nicks.
3. Another important factor to consider in preparing non-nicked DNA for incremental truncation is the restriction enzyme digestion to linearize the DNA in preparation for truncation. Restriction enzymes from suppliers may have nuclease contamination. In addition, restriction enzymes may have single-stranded nicking activity at high enzyme to DNA ratio. For this reason, we recommend digesting the DNA with the minimum amount of restriction enzyme necessary to fully digest the DNA and avoiding conditions contributing to star activity (relaxed or altered specificity) in restriction enzymes (*see* manufacturer's product specification). On the other hand, it should be emphasized that incomplete digestion at this step could result in a significant fraction of the final library being untruncated. Gel purification of the digested DNA can guard against this (*see* **step 2** in **Subheading 3.2.**).
4. We found the rate of truncation to vary with NaCl at 22°C by the following equation: rate (bp/min)= $47.9 \times 10^{(-0.00644 \times N)}$ where N = concentration of NaCl in m*M* (0–150 m*M*). Using this equation, the rate of Exo III digestion is ~10 bases/min at 22°C in the presence of 100 m*M* salt. This rate expression is valid for a DNA concentration of 33.3 ng/μL and 100 units of Exo III per μg DNA. The rate of Exo III in the presence of NaCl at higher temperatures has been determined elsewhere ***(12)***.

5. It is useful to include internal controls for testing fidelity and rate of truncation. If using these controls, increase the amount of DNA to 6 μg , the amount of buffer to 18 μL, and the total volume to 180 μL while keeping the desired NaCl concentration. Remove a 60 μL aliquot (a no truncation control) to a tube of 300 μL PB buffer before adding the Exo III. At some point during the truncation (a final time-point works well), remove one 60-μL time point to a tube of 300 μL PB buffer. Process both these tubes in parallel with the truncation library through the mung bean nuclease step. At this point analyze the DNA in these control tubes by agarose gel electrophoresis to determine the experimental truncation rate. To make size determination more accurate, or in cases where the length of truncation is small in comparison to the size of the plasmid, it may be advantageous to digest this DNA with a restriction enzyme prior to electrophoresis.
6. The sample size depends on the maximum truncation length, the truncation rate and the frequency of sampling. For example, to truncate 300 bp, one could digest at 100 m*M* salt (Exo III digestion rate ~ 10 bp/min) and take 1 μL samples every 30 sec for 30 min. Exo III digestion produces a Gaussian distribution of truncation lengths with a standard deviation of 0.075–0.2 times the total truncation length ***(12)***. Thus, one can take successive samples whose average truncation length differs by more than one base and still expect a relatively even distribution of truncation lengths ***(7)***.
7. Occasionally we have had to optimize this step by performing test mung bean nuclease digestions. We use a single truncation time point for the test (i.e., digest 12 μg of DNA with Exo III for the time necessary to digest 100 bp). The single time point is then divided into six tubes of 47 μL each (i.e., 2 μg DNA in each tube, as in the normal truncation protocol) to which is added 5 μL of mung bean buffer (10X) and mung bean nuclease in different amounts (try 12, 6, 3, 1.5, and 0.75, and 0 units). A portion of the DNA is then run on a 0.7% agarose gel. DNA not digested with mung bean nuclease will run in a somewhat diffuse band larger than the expected size owing to the large, single-stranded overhangs. As more mung bean nuclease is added, the DNA will begin to smear between the size obtained without mung bean nuclease treatment and the expected size. At the optimum amount of mung bean nuclease, the DNA will run as a relatively focused band at the expected size. When too much mung bean nuclease is added, the DNA will begin to smear to much smaller sizes as mung bean nuclease begins to make double stranded breaks.
8. It is very important that digestion is complete so as to avoid a significant fraction of the final library being a coordinated truncation library (i.e., both genes truncated approx the same amount).
9. In our experience, ammonium acetate is preferable over sodium chloride or sodium acetate. To help visualize the precipitated pellet in subsequent steps, we usually add 2 μL of pellet paint (Novagen; Madison, WI) after the ammonium acetate and before the 100% ethanol.
10. No particular restrictions apply to what plasmid(s) can or cannot be used as template vector as long as sufficient unique restriction sites are present. However,

minimizing the size of the vector is advantageous for the subsequent PCR amplification step, reducing the required extension time and minimizing the accumulation of point mutations. Template vectors (including parental genes) of up to 5000 basepairs have successfully been used.

11. Gel purification of the digestion product is very important to prevent contamination of the subsequent steps in the protocol and the final library with uncut vector.
12. The preferred polymerase is *Taq* DNA polymerase since it has the lowest error-rate of exonuclease-deficient polymerases. You can use any heat-stable DNA polymerase provided that it is exonuclease-deficient. Upon completion of DNA synthesis, polymerases with exonuclease activity engage in a process known as idling: the continuous removal and resynthesis of the 3'-ends. None of the 3'-5' exonuclease activities of Klenow, T4, *Vent*, and *Pfu* DNA polymerase is capable of hydrolyzing the phosphorothioate linkage. Thus, this idling leads to the unwanted accumulation of phosphorothioate containing nucleotides at the 3'-ends of the resynthesized strands, biasing the resulting library towards full-length fragment sizes.
13. The best ratio of dNTPs to α-phosphorothioate dNTPs depends on the size of the truncation region. The two listed mixtures work well for truncations between 100–1000 basepairs. Larger quantities of PCR product for the following steps may be desired. We recommend to run two or three 50 μL reactions of each dNTP / α-phosphorothioate dNTP mixture in parallel, rather than scaling up the reaction itself. The optimal annealing temperature for primer A and B must be determined prior to the amplification with α-phosphorothioate dNTPs. We recommend primers of 20–25 bp length with T_A of 55–60°C.
14. Each 50 μL PCR reaction generates 1–10 μg of product. The quantity and quality of the product bands of the control reaction and the various dNTP/αS-dNTP ratios should be more or less the same. We usually combine a maximum of two reactions per spin column to minimize product losses.
15. We use the following rule for concentration measurements: 1.0 A_{260} = 50 μg DNA. As a guideline, our experimental yields range from 5–20 μg of PCR product in 50–100 μL of elution buffer (10 m*M* Tris-HCl, pH 8, provided by Qiagen as part of the QIAquick kits).
16. It is very important to carefully but thoroughly mix the reaction upon addition of Exo III. We obtained good, consistent results by using a pipettor for 20–30 s.
17. The final step of the QIAquick purification protocol consists of eluting the DNA with a minimum of 30 μL EB-buffer. We usually adjust this volume to directly obtain DNA concentrations of approx 1 μg of DNA per 10 μL solution. Our calculations are based on the initial absorbance measurements in **step 5** and **not** on measurements after nuclease treatment.
18. During the blunt-ending as described in **steps 10–12**, each consecutive addition of reagent involves a careful mixing of the reaction solution, using the pipettor. Attention should be paid to splashing parts of the reaction mixture. In those cases, a brief spinning of the sample is recommended.
19. We recommend loading 1–4 μg of DNA (again based on the original absorbance measurements) to ensure a sufficient amount of DNA for visualization. A typical

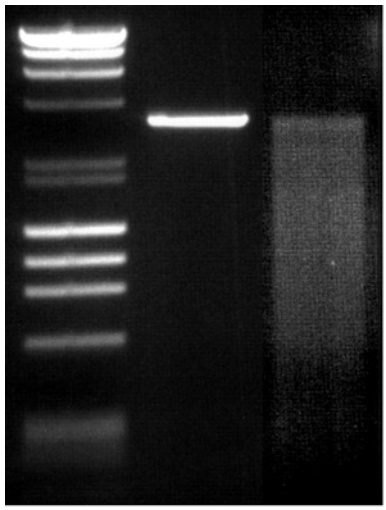

Fig. 3. Exo III treatment of PCR-amplified linearized plasmid containing nucleotide analogs. Size marker is a 1:1 mixture of Lamda DNA (*Hin*dIII digest) and ϕX174 DNA (*Hae*III digest). The center shows the product band after the PCR amplification in the presence of nucleoside analogs. On the right, the same sample after nuclease treatment is a smear of truncated material, reaching from the size of the original construct to very small fragments of a few hundred nucleotides in length.

example for an incremental truncation library at this stage is shown in **Fig. 3**. Note that the control experiment with only dNTPs should have no visible band at all.

20. Crucial to the successful intramolecular ligation is to maintain an overall DNA concentration of <3 ng/μL ***(8)***. Adjust volumes of the ligation reaction for DNA concentration to fall below this threshold.
21. 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.
22. Although the colony PCR method is quick and easy, some samples fail to amplify a band even though they have truncations in the desired range. A more definitive, but time consuming, method to characterize the size distribution of the library is to grow inoculum of randomly selected colonies, miniprep the plasmid DNA, digest the DNA with the appropriate restriction enzymes and analyze the distribution by agarose gel electrophoresis.
23. PCR reaction mixture consists of 1X *Taq* polymerase buffer, 1.25 m*M* $MgCl_2$, 200 μ*M* dNTPs, and 1 μ*M* gene-specific primers.

24. In our experience, approx 50% of the colony-PCR reactions generate a PCR product, ranging in size from completely untruncated material to fragments only a few dozen bases in length. The ratio of insert-carrying colonies can be improved by size selection of the incremental truncation library on an agarose gel, subsequent to the blunt-ending step.
25. As a rule-of-thumb, in order for the library to have the vast majority of possible crossovers, the number of transformants with truncations in the correct range should be approx 5–10 times the theoretical degeneracy of the library. The theoretical degeneracy of an ITCHY library is the product of the maximum number bases truncated in each gene. The following more rigorous treatment of library completeness derives from a treatment of genomic libraries ***(13)***. The probability P_C that a library is complete can be calculated as follows. We will treat the ITCHY library as ideal in that only truncations in the desired range are created and all members are equally represented. Deviations from ideality are discussed below. Let T_{max1} and T_{max2} be the maximum desired truncations (in basepairs) of genes 1 and 2 respectively. Then the degeneracy D of the library (the number of possible variants) is the product of T_{max1} and T_{max2}. The frequency, f, of a particular fusion in this library is the inverse of the degeneracy ($1/D$). The probability of not picking a particular sequence if one picks a random library member is $1-f$. Thus, given a number of transformants N, the probability P_i that sequence i is in the library is given by **Equation 1**.

$$P_i = 1 - [1 - f]^N \qquad \text{[Eq. 1]}$$

Using **Equation 1**, it can be shown that one needs to have the number of transformants be 4.6-fold more than the degeneracy D in order to have a 99% probability of having a particular sequence in the library. This factor is essentially independent of D for $D > 100$. The probability that a library is complete, P_C, is the product of all P_i's for the entire degeneracy given by **Equation 2**.

$$P_C = [1 - (1 - f)^N]^D \qquad \text{[Eq. 2]}$$

Using **Equation 2**, it can be shown that the factor $F_{0.99}$ by which number of transformants N must exceed the degeneracy D in order that there is a 99% probability that the library is complete ranges from about 10 to 20 for typical library sizes and can be given by **Equation 3**.

$$F_{0.99} = 4.4049 + 2.3348\log(D) \qquad \text{[Eq. 3]}$$

Equations 1–3 are not particular to ITCHY libraries and can be applied to any library by appropriately calculating its degeneracy. Finally, the above discussion of probabilities should be tempered by three factors. First, libraries are never ideal, and thus the true f is the product of the inverse of the degeneracy and the fraction of the library that has truncations in the desired range. Second, the above discussion assumes that the all members of the library are equally represented. Modeling indicates that even under the best of experimental conditions, this will not be true for either method, particularly truncations using nucleotide analogs ***(6)***. Third, while a complete library is a good goal to have, libraries do not have to be complete to be useful.

Acknowledgments

We thank Stephen J. Benkovic for his support and guidance. We also thank Dave Paschon for optimizing the mung bean nuclease step in the time-dependent truncation protocol.

References

1. Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999) Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. USA* **96,** 3562–3567.
2. Lutz, S., Ostermeier, M., and Benkovic, S. J. (2001) Rapid generation of incremental truncation libraries for protein engineering using α-phosphothioate nucleotides. *Nucleic Acids Res.* **29,** e16.
3. Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17(12),** 1205–1209.
4. Sieber, V., Martinez, C. A., and Arnold, F. H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19(5),** 456–60.
5. 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.
6. Putney, S. D., Benkovic, S. J., and Schimmel, P. R. (1981) A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated *in vivo*. *Proc. Natl. Acad. Aci. USA* **78,** 7350–7354.
7. Ostermeier, M. (2003) Theoretical distribution of truncation lengths in incremental truncation libraries. *Biotechnol. Bioeng.*, in press.
8. Ostermeier, M., Lutz, S., and Benkovic, S.J. (2002) Generation of protein fragment libraries by incremental truncation, in *Protein-Protein Interactions: A Molecular Cloning Manual*, (Golemis, E. A., ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd ed*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. Zasloff, M., Ginder, G. D., and Felsenfeld, G. (1978) A new method for the purification and identification of covalently closed circular DNA molecules. *Nucleic Acids Res.* **5,** 1139–1152.
11. Gaubatz, J. W. and Flores, S. C. (1990) Purification of eucaryotic extrachromosomal circular DNAs using exonuclease III. *Anal. Biochem.* **184,** 305–310.
12. Hoheisel, J. D. (1993) On the activities of *Escherichia coli* exonuclease III. *Anal. Biochem.* **209,** 238–246.
13. Zilsel, J., Ma, P. H., and Beatty, J. T. (1992) Derivation of a mathematical expression useful for the construction of complete genomic libraries. *Gene* **120,** 89–92.