In vitro DNA recombination

Kentaro Miyazaki and Frances H. Arnold
Division of Chemistry and Chemical Engineering,
Mail Code 210-41, California Institute of Technology,
1200 East California Boulevard, Pasadena, CA 91125, USA.

Introduction
This chapter illustrates in vitro DNA recombination methods frequently used in
directed evolution. Four different methods are presented with full experimental
detail: DNA shuffling (Stemmer method), random priming recombination, Stag-
gered extension process (StEP), and in vitro heteroduplex formation in vivo DNA
repair (heteroduplex recombination). One problem in recombination experi-
ments is unwanted nonrecombinant parent sequences and mis-recombined
sequences. Protocols to reduce these backgrounds are included.

Background to in vitro DNA recombination
Use of in vitro DNA recombination in directed evolution
Directed evolution mimics Darwinian evolution by using iterative mutation and
selection (or screening). This simple algorithm enables us to engineer molecules
with desired properties (1). Sequence diversity is usually created by point muta-
genesis and recombination. Although sequential rounds of random point muta-
genesis can be used, this evolutionary strategy suffers two weaknesses. First, only
one best sequence can parent the next generation even when there are multiple
good sequences. Discarded good mutations have to be rediscovered in future
generations to be incorporated. Depending on the screening strategy, the cost
to rediscover these mutations can be large compared to collecting them by
recombination. Second, the fitness that can be reached by point mutagenesis
is possibly limited because deleterious mutations are often accumulated along
with beneficial ones. Recombination can overcome these drawbacks by allowing
mutations to become associated in different combinations. Multiple sequences
are recombined, and, as a result of screening, beneficial mutations accumulate
while deleterious ones are eliminated.
DNA recombination in vitro as an evolutionary strategy was first reported by
Stemmer (2, 3). Stemmer's method, which he called "DNA shuffling," combines
homologous recombination of related sequences and a low (and controllable) rate of random point mutagenesis in one experimental operation. The term DNA shuffling has now become synonymous with in vitro recombination and its associated point mutation capability. Stemmer's group also demonstrated evolution by recombination of naturally occurring homologous genes ("family shuffling" or "molecular breeding") (4). Gene families, which diverged from a common ancestor and thus share the same structural framework, are considered to be pre-selected for similar folds and function(s). This property—diversity in sequence yet similarity in structure and function—is advantageously used in molecular breeding experiments to minimize the disruption of structure while expanding the sequence space explored for evolution of specific, desirable features. Since Stemmer's shuffling method was first published in 1994, several new methods have been described. These methods are presented and compared in Section 3.

**Applications of in vitro DNA recombination**

The first demonstration of DNA shuffling was to improve drug resistance (2, 3). DNA shuffling in its various forms has since been applied to altering a wide array of protein properties. Folding and solubility were improved for the green fluorescent protein (5) and for single-chain antibody fragments produced in *Escherichia coli* (6). Thermostable enzymes have been generated (7–10). Enzyme substrate specificity was changed (11), and enzyme activity in organic solvent was increased (12). DNA shuffling has also been applied to evolving whole operons (13).

Family shuffling was first employed to recombine members of the cephalosporinase gene family. The activity of the best mutant towards a new antibiotic was far above the level of any parent and also outperformed single-gene shuffling products (4). Similarly, the substrate range of a biphenyl dioxygenase was extended by shuffling a pair of related dioxygenase genes (14, 15). A library prepared by shuffling twenty-six subtilisin sequences contained sequences highly adapted for various properties tested, including activity in acid and alkaline media, stability in dimethylformamide and thermostability (16). For more examples of evolution by in vitro recombination, see (17).

**Recombination statistics**

When recombining multiple sequences to accumulate beneficial mutations, it is useful to consider the recombined library size (12). As the number of sequences and mutations increases, the probability of identifying the rarest recombination species decreases rapidly. Because the least frequent sequences are the ones that contain most mutations from the parent population, those sequences are often the most desired. The experimenter should thus consider the screenable library size when deciding the number of sequences to recombine and the functional improvements that can be obtained.
For the recombination of \( N \) sequences with \( M \) total mutations, the probability of generating progeny sequences containing \( \mu \) mutations \( (P_\mu) \) is equal to the number of ways a \( \mu \)-mutation sequence can be generated \( (\binom{M}{\mu}) \) multiplied by the probability of generating any single \( \mu \)-mutation sequence

\[
P_\mu = \binom{M}{\mu} \left( \frac{1}{N} \right)^\mu \left( \frac{N-1}{N} \right)^{M-\mu}
\]

The probability of generating any specific sequence decreases precipitously with increasing numbers of parents. The rarest sequence will be the one containing all the mutations \( (\mu = M) \), its probability \( (P_M) \) is \( 1/N^M \). In practice, some degree of oversampling is required in order to maximize the chance of discovering a given variant. The sampling \( S \) required to achieve a given level of confidence of having sampled the rarest variant in a library is given by \( (1-P_M)^S < 1 - \text{[confidence limit]} \). Generally, for 95% certainty that a specific clone has been sampled, the oversampling is between 2.6 and 3.0.

Given these numbers, it is highly unlikely, for example, to find the best recombination of ten mutations from ten separate parents each having a single mutation in a single round of shuffling and screening \( (P_{10} = 0.0000000001) \). However, it is possible to find a good triple mutant \( (P_3 = 0.0574) \).

**Methods for *in vitro* DNA recombination**

**Stemmer method**

Stemmer’s published DNA shuffling method (2) uses enzymatic digestion of the parent genes to generate a pool of random DNA fragments. These fragments can be assembled by iterative cycles of denaturation, annealing and extension with thermostable DNA polymerase. This reaction generates a mixture of products in length and combination, from which full-length genes are amplified in a polymerase chain reaction (PCR) reaction with flanking primers. A low (and controllable) rate of random point mutation accompanies recombination. High-fidelity shuffling is advantageously used to recombine beneficial mutations and minimize the generation of potential deleterious mutations (18, 19).

Fragmentation of the DNA is usually performed using a nonspecific endonuclease such as DNase I (an alternative method for random DNA fragmentation is described in section 3.2). In the reaction, magnesium (2, 3) or manganese (18–20) ion is included. These ions affect the digestion differently: magnesium promotes random nicks, while manganese stimulates cleavage of both strands at approximate the same site (21). The products of reaction with magnesium-DNase I may therefore remain annealed in non-denaturation conditions, and the apparent fragment size will depend on the electrophoresis conditions (e.g. buffers and temperatures). To stop the DNase I digestion, it is
highly recommended to add an excess of ethylene diamine tetraacetic acid (EDTA) over the divalent cation rather than use thermal inactivation. Divalent cations are known to enhance the thermostability of DNase I, and the reaction continues with increasing rate (22). EDTA inactivation is quicker, technically simpler and more reproducible.

**Protocol 1**

**DNA shuffling**

**Equipment and reagents**

- DNase I (Boehringer)
- 0.5 M EDTA, pH 8.0
- Taq DNA polymerase (Boehringer, Promega, or Stratagene)
- 10× Taq buffer (Boehringer, Promega, or Stratagene)
- Deoxyribonucleotide triphosphate (dNTP) mix: 2 mM of each dNTP (Amersham Biosciences)
- QIAEX II gel extraction kit (Qiagen) or Wizard prep (Promega)
- DNA parent samples (2–5 μg/each)
- Forward and reverse PCR primers (50 pmol/each)

**A. Fragmentation**

1. Prepare 10 μl (2–5 μg) of parent DNA samples in water, mixed in equal proportions.\(^b\) Equilibrate at 15 °C.
2. Prepare DNase I solution: 50 mM Tris-HCl, pH 7.4, 10 mM MnCl₂, 0.2 U DNase I in 100 μl. Equilibrate at 15 °C.
3. Add 40 μl of DNase I solution (from step 2) to 10 μl of parent DNA. Perform the digestion at 15 °C.
4. Take 10 μl aliquots after 1, 2, 3, 5, and 10 min of incubation and immediately mix each of them with 1 μl of 0.5 M EDTA.\(^c\)
5. Separate the DNA fragments by electrophoresis in a 2% (w/v) agarose gel\(^d\) containing ethidium bromide (0.5 μg/μl). Visualize the DNA with a preparative UV illuminator (typically at 366 nm)\(^e\) and excise the DNA in the desired size range.\(^f\)
6. Purify the fragments (e.g. using QIAEXII or Wizard prep kit) and elute in 10 μl water.

**B. Assembly**

1. Combine 5 μl of 10× Taq buffer, 5 μl of dNTP mix, 10 μl of the purified fragments (from Protocol 1A, step 6), and 2.5 U Taq polymerase in a total volume of 50 μl.
2. Run the assembly reaction using the following thermocycler program: 3 min at 94 °C followed by 40 cycles of 30 s at 94 °C, 1 min at 55 °C, and 1 min + 5 s/cycle at 72 °C.\(^g\)
3. Check the extent of reaction by analyzing a small proportion on an agarose gel.\(^h\)
C. Amplification

1. Combine 10 µl of 10× Taq buffer, 10 µl of dNTP mix, 0.5 µM (final) forward and reverse primers, 1 µl of assembly reaction, and 5 U Taq polymerase in a total volume of 100 µl.

2. Run the PCR reaction at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Perform a total of 30 cycles.

3. Check the amplification by analyzing a small proportion on an agarose gel.\(^a\)

\(^a\)Modified from the method given in (18).

\(^b\)Parent DNA may be plasmids carrying target sequences, sequences excised by restriction endonucleases or amplified by PCR.

\(^c\)DNase I concentration and/or incubation time/temperature can be adjusted if necessary.

\(^d\)For selection of gel concentration, see Section 3.8.3.

\(^e\)Tracking dyes in standard loading buffer (10× loading buffer: 50% (v/v) glycerol, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol) can mask fluorescence from DNA fragments. For optimal gel transparency, dilute the dyes with glycerol until they are slightly visible. Alternatively, use dye-free loading buffer for the samples and load dyes in empty lanes to monitor the migration.

\(^f\)Contamination with full-length parent sequence sometimes happens, especially when a PCR product is used. Even after digestion for extended periods of time, a trace of full-length parent may remain. Using the unfraccionated mixture for the assembly reaction would generate an unacceptably large fraction of parent molecules. Visual control by gel electrophoresis is recommended for reliable separation of fragments in the desired size range.

\(^g\)The PCR program depends on the fragment size. Assembly from small fragments may require more cycles and longer extension time than from large fragments. Small fragments may also require lower annealing temperature, at least during the initial cycles.

\(^h\)You should see a smear extending through the size of your full-length product. Use more cycles, or lower the annealing temperature, if a smear does not appear.

\(^i\)If the amplification is dominated by a smear rather than full-length sequence, repeat the amplification with a lower concentration of assembly products. Another solution to the smear problem is the use of nested primers; see Section 3.7.3.

Random DNA fragmentation with endonuclease V from E. coli

DNase I has most frequently been used for random DNA fragmentation. However, the length of fragments varies greatly with minor changes in conditions, including the amount of nuclease, the source (supplier) or lot of nuclease, the reaction temperature, and the purity of DNA substrates. The experiment requires very careful control of the partial digestion reaction and, hence, is labor-intensive and time-consuming. To overcome these drawbacks, an alternative approach was developed, in which random fragmentation is achieved in
complete digestion (23). In this method, uracil-containing recombination templates are prepared by PCR in the presence of dUTP and dNTP. The full-length DNA is then digested with endonuclease V from *E. coli*, which is known to cleave uracil-containing DNA at the second or third phosphodiester bond 3' to uracil sites. By adjusting the concentrations of dUTP and dTTP, one can obtain the desired fragment length.

**Protocol 2**

**Random DNA fragmentation**

**Equipment and reagents**

- Taq DNA polymerase (Boehringer, Promega, or Stratagene)
- 10 × Taq buffer (Boehringer, Promega, or Stratagene)
- 2 mM each of dATP, dGTP, dCTP (Amersham Biosciences)
- dTTP/dUTP mix: 2 mM/0 mM, 1.5 mM/0.5 mM, 1 mM/1 mM, 0.5 mM/1.5 mM, 0 mM/2 mM (Amersham Biosciences)
- Forward and reverse primers (25 pmol each)
- Template DNA samples (1-50 ng each)a
- QIAquick spin column (Qiagen) or QIAEX II gel extraction kit (Qiagen)
- Endonuclease V from *E. coli* (Trevigen)
- 10 × endonuclease V buffer (Trevigen)

**Method**

**A. Preparation of uracil-containing recombination templates**

1. Combine 5µl of 10 × Taq buffer, 0.2 mM each of dATP, dGTP and dCTP, and 0.2 mM of a dTTP/dUTP mixture, 1-50 ng of template DNA, 25 pmol of each forward and reverse primers, and 1.25 U Taq DNA polymerase in a total of 50 µl.

2. Heat the PCR mixture at 95°C for 1 min, followed by 25-30 cycles of incubation at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a final incubation at 72°C for 7 min.

3. Separate the DNA fragments by electrophoresis in a 1% (w/v) agarose gel containing ethidium bromide (0.5 µg/ml). Visualize the DNA with a preparative UV illuminator and excise the DNA in the desired size range.

4. Purify the fragments (e.g. using a QIAspin column or QIAEX II kit) and elute in 27 µl water.

**B. Endonuclease V digestion of the recombination templates**

1. Add 3 µl of 10 x endonuclease buffer and 1 U endonuclease V.

2. Incubate the solution at 37°C for 12 h, followed by heating at 95°C for 10 min.

3. Separate the DNA fragments by electrophoresis in a 2% (w/v) agarose gel containing ethidium bromide (0.5 µg/ml). Visualize the DNA with a preparative UV illuminator and excise the DNA in the desired size range.b
4. Purify the fragments (e.g. using a QIAspin column or QIAEX II kit) and elute in 30 μl water. Use 3-10 μl of the purified fragments to continue with the assembly reaction (section B of Protocol 1).

\[ \text{Plasmid DNA is preferred.} \]

\[ \text{You should see not only smear DNA fragments but also full-length recombination templates. This is because some recombination templates lack uracil in their sequence.} \]

Random priming recombination

In this method, conventional random priming DNA synthesis (24) is used to generate fragments for assembly (25). DNA synthesis is carried out using Klenow fragment of E. coli DNA polymerase I which lacks 3'→5' exonuclease activity (26). The random priming reaction can be carried out using a “wild type” Klenow fragment under slightly acidic conditions (pH 6.6) where problematic 3'→5' exonuclease activity is reduced (25, 27). However, we recommend using a “variant” Klenow fragment, whose 3'→5' exonuclease activity is abolished by genetic engineering (D355A, E357A mutations). Random hexanucleotides are used for priming: they are long enough to form stable duplexes with template and short enough to ensure random annealing. While longer primers can also be used, annealing may not be random for short genes. Unlike DNase I fragmentation, this method does not require double-stranded DNA and can be used on both double-stranded and single-stranded DNA templates. Moreover, random primers can also be used with RNA templates.

Protocol 3

Random priming recombination

Equipment and reagents

- Klenow fragment (3'→5' exo)
  (New England Biolabs)
- Klenow fragment buffer (3'→5' exo)
  (New England Biolabs)
- Template DNA samples (1-50 ng each)
- Random hexanucleotides
  (Amersham Biosciences)
- 5 mM of each dNTP
  (Amersham Biosciences)
- QIAquick spin column (Qiagen) or QIAEX II gel extraction kit (Qiagen)

Method

1. Combine 0.2–0.5 pmol of each template DNA and 7 nmol of random hexanucleotide primers in a total volume of 78 μl.
2. Incubate 5 min at 100°C and immediately transfer to an ice water bath.
3. Add 10 μl of 10 x exo- Klenow fragment buffer, 10 μl of dNTP mix, and 10 U Klenow fragment to bring to 100 μl.

4. Incubate 3-6 h at 22°C.

5. Stop the reaction by placing the tube on ice.

6. Separate the DNA fragments by electrophoresis in 2% (w/v) agarose gel containing ethidium bromide (0.5 μg/ml). Visualize the DNA with a preparative UV illuminator and excise the DNA of the desired size.c

7. Purify the fragments (e.g. using a QIAspin column or QIAEX II kit) and elute in 30 μl water. Use 3-10 μl of the products to continue with the assembly reaction (section B of Protocol 1).

Modified from (25).

bTemplates can be plasmids carrying target sequences, sequences excised by restriction endonucleases, or amplified by PCR. However, because the products are separated from template by size, it is not recommended to use very short DNA templates.

cTracking dyes in standard loading buffer (10 x gel loading buffer: 50% (v/v) glycerol, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol) can mask fluorescence from DNA fragments. For optimal gel transparency, the dyes can be diluted with glycerol until they are slightly visible. Alternatively, use dye-free loading buffer for the samples and load dyes in empty lanes to monitor the migration.

Staggered extension process (StEP)

StEP recombination is based on template switching during polymerase-catalyzed primer extension (28). The denaturation and short annealing–extension cycles limit the primer extension in a single cycle. Extension interrupted by heat denaturation resumes during the next annealing–extension step, where the partially extended primers can anneal to different parent sequences. Multiple cycles of partial extension then create a library of chimeric sequences. As it involves no template digestion or fragment assembly, this protocol is very simple.

The faster the full-length product appears in the extension reaction, the fewer the template switches that have occurred and the lower the crossover frequency. Everything possible should be done to minimize time spent in each cycle: selecting a faster thermocycler, using smaller test tubes with thinner walls, and, if necessary, reducing the reaction volume. DNA polymerases currently used in DNA amplification are very fast. Even very brief cycles of denaturation and annealing provide time for these enzymes to extend primers for hundreds of nucleotides: 70 °C, >60 nt/s; 55 °C, ~24 nt/s; 37 °C, ~1.5 nt/s; 22 °C, ~0.25 nt/s (29). Therefore, it is not unusual for the full-length product to appear after only 10–15 cycles. Polymerases are not all equally fast, however. The proofreading activity of Pfu and Vent polymerases slows them down, offering another way to increase recombination frequency. Processivity of the polymerase may also
affect the crossover frequency. Because of the short extension time, the selection of thermal cycler and cycling program can significantly affect the crossover efficiency. It is thus highly recommended to maintain a fixed time between the set temperatures.

**Protocol 4**

**StEP DNA recombination**

**Reagents**
- Taq DNA polymerase (Boehringer, Promega, or Stratagene)
- 10 x Taq buffer (Boehringer, Promega, or Stratagene)
- dNTP mix: 2 mM each dNTP (Amersham Biosciences)
- forward and reverse PCR primers (7.5 pmol/each)
- template DNA samples (1-20 ng/each)

**Method**
1. Combine 5 μl of 10 x Taq buffer, 5 μl of dNTP mix, 1-20 ng of template DNA, 0.15 μM (final) forward and reverse primers, and 2.5 U Taq polymerase in a total volume of 50 μl.
2. Run 80-100 polymerase extension cycles: 94°C for 30 s and 55°C for 5-15 s.
3. Check a small amount of the reaction on an agarose gel.
4. (optional) To remove parent DNA prepared from a dam+ E. coli strain (methylation positive strain, e.g. DH5α, XL1-Blue) or ung- strain (e.g. CJ236), see Section 3.9.
5. (optional) Amplify the target sequence in a standard PCR reaction.
6. Check a small amount of the reaction on an agarose gel.

---

*a* Modified from the method given in (28).

*b* Templates may be plasmids carrying target sequences, sequences excised by restriction endonuclease or amplified by PCR. When the size of parent DNA is close to those of products, enzymatic degradation of the parent is necessary. See Section 3.6 for details.

*c* Possible reaction products are full-length amplified sequence, a smear, or a combination of both. If a discrete band is obtained after the StEP reaction, the products may be cloned into a vector without amplification.

*d* If the amplification reaction is not successful and a smear rather than a discrete band occurs, repeat the amplification with a lower concentration of the StEP reaction mixture. Run the reaction with several dilutions of the StEP reaction: 1/10, 1/20, and 1/50, and select the most successful one for cloning. Another solution to the smear problem is to use nested primers (see Section 3.7.3).
In vitro heteroduplex formation and in vivo repair (heteroduplex recombination)

This method comprises two simple processes: in vitro heteroduplex formation and subsequent mismatch repair in vivo (30). In contrast to other methods, recombination occurs in non-homologous (mismatched) sequences rather than in regions of homology. Either whole plasmids or target fragments can be used. For creating the plasmid heteroduplex, the plasmids have to be linearized with restriction enzymes (one plasmid with restriction enzyme A and the other with B), whose restriction sites are located outside the target region. This enables one to differentiate a heteroduplex from the homoduplex in a bacterial cell. The linear, homoduplex form is immediately digested by the host and will not contribute to transformation. On the other hand, the circular, heteroduplex form is modified by the mismatch repair system. Sealing nicks by DNA ligase treatment increases the efficiency of recombination several-fold (Table 1, (30)).

If short target sequences are used for parents, the discrimination system in vivo cannot be used. Further, if double-stranded fragments are used, the probability of forming heteroduplex is at most 50%, and it is hard to isolate it from homoduplex by any physical method (e.g. agarose gel, membrane). To maximize the formation of heteroduplex, asymmetric PCR synthesis of one of the two strands of the target genes is advantageous. The resulting single-stranded sequences are used for annealing. The heteroduplex products are cloned into a vector, and the plasmids are used to transform bacteria.

Protocol 5

Heteroduplex recombination: Plasmid heteroduplex

Reagents

- 20 × SSPE: 0.2 M NaH₂PO₄, pH 7.4, 3.6 M NaCl, 20 mM EDTA
- Two plasmids for recombination (2–4 μg/each)
- DNA ligase (Stratagene)

Method

1. Combine equal amounts (~2–4 μg) of two plasmids, 5 μl of 20 × SSPE to bring to 100 μl.
2. Incubate the mixture at 100 °C for 10 min, immediately transfer to an ice water bath, and continue incubation at 68 °C for 2 h.
3. Repair the nicks with DNA ligase.
4. Transform bacteria with the products.

aMethod taken from (30).
Protocol 6

**Heteroduplex recombination: insert heteroduplex**

**Equipment and reagents**
- 20× SSPE: 0.2 M NaH₂PO₄, pH 7.4, 3.6 M NaCl, 20 mM EDTA
- DNA ligase (Stratagene)
- QIAspin PCR purification kit (Qiagen)
- DNA template samples (5 ng/each)
- Forward and reverse PCR primers (15–25 pmol/each)

**Method**

1. Combine 5 μl of 10× Taq buffer, 5 μl of the dNTP mix, 0.3–0.5 μM of one primer, 5 ng of one DNA template and 5 U Taq polymerase in a total volume of 50 μl. One DNA template is combined with the forward primer, another with the reverse primer.

2. Synthesize the ssDNA templates in a PCR-like reaction for 100 cycles: 94 °C for 30 s; 56 °C for 30 s; and 72 °C for 1 min.

3. Combine the annealing products together and purify them using a QIAspin PCR purification kit or an equivalent kit. Elute the DNA in water.

4. Combine 47.5 μl of purified products and 2.5 μl of 20× SSPE.

5. Incubate the mixture at 100 °C for 10 min, immediately transfer to an ice water bath, and continue the incubation at 68 °C for 2 h.

6. Purify the annealed products, digest them with the appropriate restriction enzymes and ligate them into a cloning vector.

7. Transform bacteria with the products.

*a Method taken from (30).

**Choice of recombination method**

The recombination methods (DNA shuffling, random priming recombination, StEP, and heteroduplex recombination) each have their own advantages and disadvantages. Experimenters choose a method taking several factors into account, including type of nucleic acid (DNA or RNA), number of parents, length of genes, operational simplicity, secondary structure, and sequence similarity. Relative performance will probably differ from sequence to sequence and optimal conditions have to be determined experimentally.

3.6.1 Type of nucleic acids

Double-stranded DNA is usable in all DNA recombination methods. DNA shuffling requires dsDNA, which can be prepared either in vivo (e.g. from plasmid) or in vitro (e.g. by PCR). Random priming recombination and StEP can use ssDNA and RNA. Heteroduplex formation becomes highly efficient if ssDNA is used (30).
3.6.2 Number of parents

None of polymerase-based recombination methods (DNA shuffling, random priming recombination, StEP) are restricted as to the number of parents. By contrast, heteroduplex recombination is limited to recombining two sequences per reaction.

3.6.3 Gene length

In general, in polymerase-based reactions such as PCR, shorter sequences serve as a better template for the reproducibility of the reaction. However, in recombination experiments, contamination of full-length parent(s) in the shorter fragments has to be avoided. If one would like to recombine short sequences, the full-length contaminant has to be removed other than by a size separation. See Section 3.6 for details.

The length of the parent DNA can affect recombination efficiency. The StEP method requires repetitive short extensions. Recombination frequency depends on extension length per cycle. In random priming DNA synthesis, it is independent of the length of the DNA template. DNA fragments as small as 200 bases can be primed equally well as large molecules such as λ phage DNA (24). The average size of synthesized DNA is an inverse function of primer concentration to template (31). Based on this guideline, proper conditions can be set for a given gene. However, the optimal conditions may vary from gene to gene.

Although PCR (and related polymerase reactions) in theory can be used to generate very long sequences, in practice the efficiency of amplification decreases significantly for very long sequences. Another drawback of PCR amplification is the introduction of unwanted mutations, which is particularly problematic for long sequences. By contrast, heteroduplex recombination neither suffers from the limitations of PCR-based approaches nor requires transformation with multiple gene fragments. In heteroduplex recombination, a whole plasmid is typically used for operational simplicity. However, because of the kinetics of annealing, long sequences adversely affect the recombination efficiency (30). In this case, short insert fragments can be used. However, because unwanted homoduplex and desired heteroduplex are non-separable, a pair of single strand sequences are used.

3.6.4 Operational simplicity

Stemmer's DNA shuffling method is the most widely used, but it requires multiple steps: fragmentation, assembly, and PCR. Each reaction step has to be carefully monitored, and the method is accordingly labor-intensive. This is also the case for the random priming method. Random digestion of uracil-containing DNA with E. coli endonuclease V (23) is less laborious because fragments can be prepared in complete digestion. StEP requires fewer steps, and once optimal conditions are found for a particular gene, one can use identical conditions without major modifications. Heteroduplex recombination is quite simple and quick. However, for maximal recombination efficiency, monitoring the formation of heteroduplex may be necessary.
3.6.5 Secondary structure
Secondary structure formation in ssDNA may adversely affect the performance of all methods, especially at low temperatures. In DNA shuffling, the randomness of gene fragmentation with DNase I may be reduced. Random priming recombination can also be affected at the most crucial step of the reaction: annealing and extension of the random primers. Small primers will be more sensitive to secondary structure than long primers at this annealing step. The low temperature used in the extension of the random primers also stabilizes secondary structure. Extension at elevated temperature requires longer primers, which would probably lead to less efficient recombination, at least for short genes. Termination of elongation of both short and long primers is sensitive to secondary structure. Formation of stable stem-and-loop structures may be the most important source of nonrandom distribution of extended fragments in random priming recombination. StEP recombination, with its very short annealing period, may be sensitive to secondary structure. Randomness created by in vitro heteroduplex formation can also be limited especially at low temperature.

3.6.6 Sequence similarity
Although all the techniques are in theory applicable for recombining genes with many mutations (sequence identity above ~50%) genes, high sequence identity seems to be critical for recombination success. Family shuffling has only been reported for the Stemmer method. The original Stemmer protocol has been devised to increase the recombination efficiency and to avoid contamination of parent sequences; the modified protocol uses various frequent-cutter restriction enzymes for fragmentation instead of DNase I (23).

3.6.7 Comparison of recombination methods
DNA recombination methods so far reported were compared using the green fluorescent protein system developed by Volkov et al. (30). The green fluorescent protein variants containing a stop codon produce nonfluorescent proteins. Generation of fluorescence requires recombination between the sites to restore the wild type sequence. The percentage of fluorescent host E. coli colonies obtained by recombining two GFP templates via various recombination methods are thereby compared (30, 32). The results are summarized in Table 1. In general, two mutations separated by short nucleotides are less efficiently recombined than are distant mutations. The insert heteroduplex recombination, where ssDNA is used for parent, gives the highest recombination efficiency for mutations separated by >99 base pairs (bp). The Stemmer method and StEP are equally efficient. With DNase I fragmentation, using smaller fragments (<100 bp) yield a slightly higher efficiency than larger fragments (100–200 bp). DNA fragments prepared by endonuclease V digestion of uracil-containing DNA gave nearly the same recombination efficiency (10% when two stop codons were separated by 180 bp; data not included in Table 1).
Table 1. Comparison of recombination methods

<table>
<thead>
<tr>
<th>Recombination methods</th>
<th>Distance between mutations (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>423</td>
</tr>
<tr>
<td>DNA shuffling (&lt;100 bp fragments)b</td>
<td>20.5</td>
</tr>
<tr>
<td>DNA shuffling (100–200 bp fragments)b</td>
<td>19.2</td>
</tr>
<tr>
<td>StEPb</td>
<td>18.5</td>
</tr>
<tr>
<td>Random priming recombinationb</td>
<td>5.0</td>
</tr>
<tr>
<td>Heteroduplex recombination (plasmid, nick)c</td>
<td>3</td>
</tr>
<tr>
<td>Heteroduplex recombination (plasmid, ligase)c</td>
<td>10</td>
</tr>
<tr>
<td>Heteroduplex recombination (insert, dsDNA)c</td>
<td>15–18</td>
</tr>
<tr>
<td>Heteroduplex recombination (insert, ssDNA)c</td>
<td>29</td>
</tr>
</tbody>
</table>

*a* Percentage of fluorescent *E. coli* colonies obtained by recombining two GFP templates containing stop codon mutations.

*b* Data taken from (30).

*c* Data taken from (32).

**Removal of background**

In Stemmer DNA shuffling, undigested parent fragment after DNase I digestion (occurs most frequently when PCR-amplified fragments are used) is the major contaminant. To avoid this, visual monitoring the digestion reaction is effective; gel separation is employed if necessary.

Many polymerase reactions become more reproducible when short DNA is used as a template. Random priming recombination and StEP method are the particular case in recombination experiments. However, if templates are too short and indistinguishable from the products in size, the nonrecombinant full-length genes will be a problem. Because conventional physical separation, such as agarose gel or membrane separation, cannot be employed, chemical and/or enzymatic treatments are applied to reduce the background. *DpnI* and uracil DNA glycosylase break the phosphodiester backbone of parent DNA in such a way that digested parents will not be amplified in the final PCR.

Another problematic contaminant leading to high concentration of inactive clones are the products of mis-recombination. Reaction conditions that promote recombination usually also generate unwanted recombination. Mis-recombination is common to all polymerase-based recombination methods such as Stemmer’s method, random priming recombination, and StEP. Background of this kind can be eliminated by nested PCR.
3.7.1 DpnI digestion
The restriction enzyme DpnI cleaves only when its recognition site (GATC) is fully-methylated (N⁶ methylation of adenine) for both strands, while leaving hemi-methylated DNA intact (33). It thereby specifically removes parent DNA isolated from a dam⁺ E. coli strain such as DH5α or XL-1 Blue. This decontamination method can be used in Step by treating the full-length product.

3.7.2 Uracil DNA glycosylase
Another way to remove parent sequences is to incorporate dUTP in parent DNA but not in product DNA. This can be achieved by PCR in the presence of dUTP instead of dTTP. Taq polymerase recognizes dUTP equally efficiently and gives an equivalent yield of products. Alternatively, one can isolate plasmid from an ung⁻ E. coli strain (e.g. CJ236) and excise the insert fragment by restriction digestion. Uracil N-glycosylase recognizes uracil residues in DNA and removes the base to generate abasic sites (34). Incubating the damaged DNA in alkaline solution at high temperature breaks phosphodiester bonds at the 3'-end of the abasic sites while keeping intact normal DNA (dUTP-free), thereby eliminating parent background. It is important to note that, if the alkaline treatment is insufficient, the parent serves as a highly mutagenic template in subsequent amplification. The resulting library will be full of inactive clones.

Protocol 7
Decontamination of uracil-containing DNA with uracil DNA glycosylase

Equipment and reagents

- Taq DNA polymerase (Stratagene)
- 10× Taq DNA polymerase (Stratagene)
- dATP, dGTP, dCTP, dUTP (Amersham Biosciences)
- Uracil DNA glycosylase (MBI Fermentas)
- QIAspin column (Qiagen)
- Template DNA (0.2–0.5 pmol)
- Forward and reverse PCR primers (50 pmol/each)

A. Preparation of parent DNA containing dUTP
1. Combine 10 μl of 10× Taq buffer, 0.2 mM each of dATP, dGTP, dCTP, 0.6 mM dUTP,⁸ template DNA, 0.5 μM each primers, and 5 U Taq polymerase in a total volume of 100 μl.
2. Amplify the target sequence under normal PCR conditions.
3. Purify the fragment and dissolve it in water.
B. StEP recombination

1 Run the StEP reaction as in Protocol 3 using the uracil-containing DNA parents as templates.

C. Decontamination of parent DNA with uracil DNA glycosylase

1 Add 5 U of uracil DNA glycosylase to the StEP reaction mixtures. Incubate the reaction mixtures for 2 h at 37°C.

2 Add 50 mM NaOH. Incubate the mixture at 95°C for 30 min.

3 Recover the mixtures using QIAspins column. Use these mixtures for the subsequent process.

*The dUTP concentration is threefold higher than the other dNTP concentrations to ensure optimal PCR efficiency.

3.7.3 Nested PCR

The reaction conditions which promote DNA recombination also promotes mis-annealing. Those unwanted products derived from incorrect recombination have to be eliminated. Nested PCR increases the yield and specificity of a PCR product. An aliquot of the first-round recombination product, for example 1 μl of a 1/10–1/1000 dilution, is subjected to a PCR amplification. The PCR is performed with two new primers which hybridize to sequences internal to the first-round sequences.

Technical tips

Optimizing DNA recombination conditions is labor-intensive and time-consuming. The reaction is frequently controlled kinetically and thus extremely sensitive to temperature, concentration of reagents, and any other seemingly minor conditions. For maximal reproducibility, it is strongly recommended to use the same thermal cycler and thin-wall PCR tubes once good conditions have been identified.

3.8.1 Oil overlay and reaction volume

If a thermal cycler does not heat the lids of the reaction tubes, you need to overlay the reaction mixture with light mineral oil to prevent evaporation. In thermal cyclers where the lid of the reaction tubes is heated continuously >96°C, a PCR can be run without oil overlay. Insufficient oil overlay may result in evaporation, which would lead to higher reagent concentration and to a decrease in temperature. For most applications, 70 μl oil for 100 μl reactions, 40 μl oil for 50 μl reactions, and 30 μl oil for 25 μl reactions is probably sufficient.
Too much oil can interfere with heat transfer. Always use the appropriate amount of oil.

3.8.2 Quality control of reagents
DNA polymerases are physically unstable (e.g. to foaming and high temperature) and always need to be kept at -20°C for storage and on ice when in use. Deoxynucleotide triphosphates are also thermolabile, and the stock solution must be stored in aliquots at -20°C. Avoid repeated freezing and thawing. Manganese chloride can form precipitates after long storage. The stock solution must be stored in aliquots at -20°C. Avoid repeated freezing and thawing.

3.8.3 Agarose gel electrophoresis
Agarose gel electrophoresis is frequently used for the analysis and size fractionation of DNA fragments. Efficient range of separation of linear DNA molecules is 0.1–3 kbp in 2% (w/v), 0.2–4 kbp in 1.5% (w/v), 0.4–6 kbp in 1.2% (w/v), and 0.5–7 kbp in 0.9% (w/v) agarose. Commonly used concentration (0.8–1.2%) may not work well for separation of short DNA fragments (50–300 bp) which are typically used in DNA shuffling.

Short DNA fragments are often hard to see on an agarose gel. This is especially problematic in preparative experiments where long UV transillumination (e.g. 366 nm) is used. Moreover, tracking dyes in loading buffer (10 × loading buffer: 50% (v/v) glycerol, 0.1% bromophenol blue (w/v), 0.1% (w/v) xylene cyanol) can mask the fluorescence from DNA (migration of bromophenol blue is ~300 bp and that of xylene cyanol is ~4000 bp in 0.5–1.4% agarose). For optimal transparency, the loading buffer can be diluted in glycerol until they are slightly visible. Alternatively, use dye-free loading buffer for the samples and load dyes in empty lanes to monitor the migration.

References