PCR-Mediated Recombination and Mutagenesis

SOEing Together Tailor-Made Genes

Robert M. Horton

Abstract

Gene Splicing by Overlap Extension (gene SOEing) is a sequence-independent method for site-directed mutagenesis and/or recombination of DNA molecules. It is based on the idea that a PCR product can be engineered by adding or changing sequences at its ends so that the product can itself be used to prime DNA synthesis in a subsequent overlap-extension reaction to create mutant or recombinant molecules. As the engineered genes are created in vitro without reliance on host organisms or restriction sites, gene SOEing provides a powerful and versatile tool for genetic investigation and engineering.

Index Entries: Polymerase chain reaction; overlap extension; recombinant PCR; gene SOEing; site-directed mutagenesis.

1. Introduction

Gene Splicing by Overlap Extension (gene SOEing) provides a powerful method of recombining sequences without depending on restriction sites or ligase, and a simple, generally applicable way of using polymerase chain reaction (PCR) to perform site-directed mutagenesis in vitro. This technique allows even those with minimal molecular biology expertise to generate quickly genetic constructs that might otherwise have been impractical using only the older (restriction enzyme-based) technology.

This method is made possible by the fact that the ends of a PCR-generated DNA fragment can incorporate new, custom-designed sequences that were not present in the original template. The primers must match their template sequence well enough to prime, but they do not have to match exactly, especially toward the 5' end. Any mismatches will be incorporated into the product, and will represent changes in the original sequence. This idea, originally called "mispriming" by Mullis et al. *(1,2)*, provides a simple way to perform site-directed mutagenesis *(3,4)*. However, simple mispriming is a limited way of creating site-directed mutants because the changes can only be made at the ends of a PCR product.

Overlap extension was originally devised as a way of introducing mutations in the center of PCR-generated sequence segments in order to make PCR mutagenesis more universally applicable *(5,6)*. The concept is illustrated in Fig. 1. First, two separate fragments are amplified from the target gene. The first is amplified with primers a and b (product AB): Primer b introduces a sequence change at the right end of product AB. The second fragment (product CD) is amplified with primers c and d, with primer c introducing the same mutation, but into the left end of product CD. These two products now share a segment of identical sequence called the overlap region. When these intermediate products are mixed together, melted, and reannealed, the top strand

Address to which all correspondence and reprint requests should be sent: Department of Biochemistry, College of Biological Sciences, 1479 Gortner Ave., St. Paul, MN 55108.

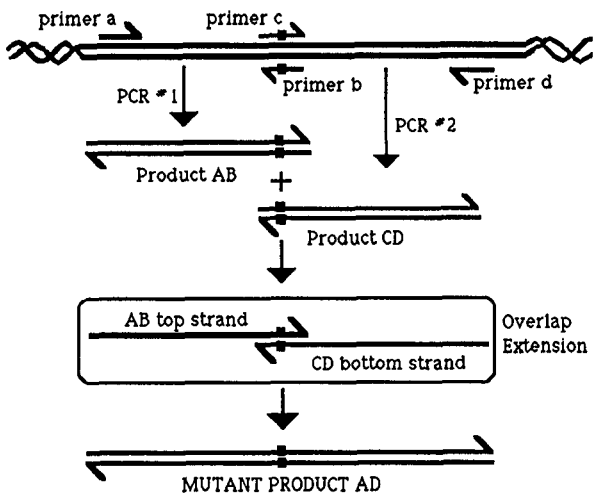

Fig. 1. Mutagenesis by overlap extension. The segment of the "wild type" gene to be mutated is shown as straight lines with the flanking regions coiling away on both sides *ad infinitum*. Synthetic oligonucleotide primers are named with lower case letters. The PCR product made with two primers is named by capital letters of the primers used to make it; for example, primers a and b make product AB. PCR products are shown as two paired strands, with half arrowheads indicating the direction in which each strand can act as a primer for DNA polymerase (the 5'–3' direction). The mutagenic primers, b and c, are complementary to one another, with the mutation shown as a black rectangle. When these intermediate PCR products are mixed, denatured, and reannealed, the top strand of AB and the bottom strand of CD can pair with this overlap at their 3' ends, and act as primers on one another to make the mutant product. The other strands, with the overlap at their 5' ends, cannot prime, and are not shown.

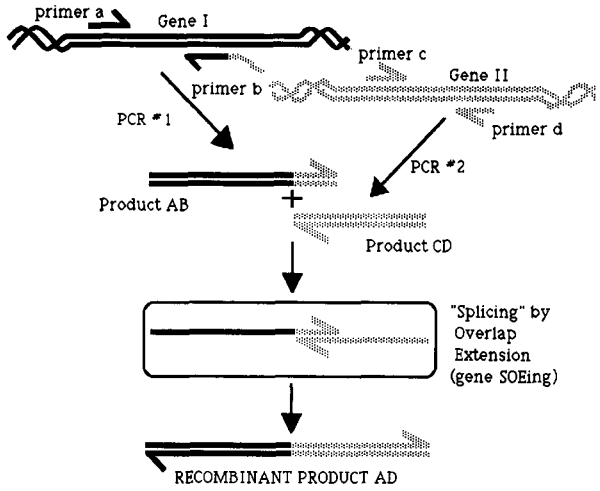

Fig. 2. Gene splicing by overlap extension (SOEing). Here, products AB and CD are derived by two different genes. SOEing primer b has sequences added to its 5' end such that the right end of AB is made complementary to the left end of CD. This allows their sequences to be joined by overlap extension, or "SOEn" together.

of AB can anneal to the bottom strand of CD in such a way that the two strands act as primers on one another. Extension of this overlap by DNA polymerase creates the full-length mutant molecule AD, which has the mutation at an arbitrary distance from either end.

In the final overlap extension step, two separate PCR-generated sequences (AB and CD) are joined together. If AB and CD are made from different genes, then product AD is a recombinant molecule *(7,8)*. Figure 2 illustrates the concept of PCR-mediated recombination, or gene SOEing. Here, extra sequence is added to the 5' end of primer b, which results in a short segment of gene II being added to the right-hand end of the PCR product amplified from gene I (product AB). This causes the two intermediate products to have an overlap region of common sequence, so that they can be joined together by overlap extension. The two strands having the overlap at their 3' ends (the "productive" strands, shown in the boxes in Figs. 1 and 2) each act as both a primer and a template to produce a giant "primer dimer," which is the recombinant molecule.

When should you use gene SOEing? For site-directed mutagenesis, overlap extension is simple, and has advantages over other methods in that

1. Recombination and mutagenesis can be performed simultaneously *(7)*;
2. Essentially all of the product molecules are mutated (i.e., 100% mutation efficiency; *6*); and
3. The product is produced in vitro (without having to grow it up in a plasmid or phage) and may be used directly in experiments *(5,9)*.

As a method of DNA recombination, gene SOEing is tremendously useful in situations where no leeway can be given to use nearby restriction sites. Protein engineering projects provide excellent examples *(7,8,10,11)*. Two major drawbacks of recombination by SOEing are the cost of the primers and the potential for introducing random errors with PCR (*see* Notes 3, 5, and 6). Therefore, gene SOEing is most practical in "complicated" constructions where there are no convenient restriction sites.

2. Materials

1. Thermal cycling ("PCR") machine.
2. *Taq* DNA polymerase.
3. 10X PCR buffer: 500 m*M* KCl, 100 m*M* Tris-HCl, pH 8.3.
4. 10 m*M* $MgCl_2$.
5. dNTPs.
6. Primers (*see* Section 3.1. and Note 1).
7. DNA templates (*see* Note 2).
8. Agarose gel electrophoresis supplies and equipment.
9. GeneClean (Bio101, La Jolla, CA), or your favorite method for purifying DNA from agarose gels.

3. Methods

3.1. Primer Design (see Note 3)

1. Unlike a regular PCR primer, a SOEing primer (such as b in Fig. 2) has to include two sequence regions capable of hybridizing to a template well enough to act as primers. The first is the priming region at the 3' end of the oligonucleotides which allows it to act as a PCR primer. The second is the overlap region at the 5' end of the oligonucleotide; this allows the strand of the PCR product that is complementary to the oligonucleotide to act as a primer in the overlap extension reaction. As shown in Fig. 2, the priming region of primer b contains a sequence from gene I, whereas the overlap region contains a sequence from gene II which is complementary to primer c.

   For examples, some primers from published overlap extension projects *(8)* are shown in the following (*see* Note 4):

```
        priming
   | - region - ||-overlap region-|
3'-gggtccgggtgtgcttctgctgtaactccg-5'  (primer b)
              5'-gaagacgacattgaggc-3'  (primer c)
                 |-priming region-|
```

2. In mutagenesis, there is only one template, so the priming and overlap regions may completely coincide *(6)*. The mismatches are placed in the center of the oligo so that both its 3' end and the 3' end of the product made with it can act as primers.
3. My colleagues and I have adopted the convention of designing both the priming and overlap regions of SOEing primers to have an estimated T_m of around 50°C using the following formula *(12)*:

$$T_m=[(G + C) \times 4] + [(A + T) \times 2] \quad (\text{in } °C)$$

   Using SOEing primer b as an example, the T_m for its overlap region is 52°C $\{[(3 + 6) \times 4] + [(2 + 6) \times 2]\}$ and for its priming region is 46°C $\{[(8 + 2) \times 4] + [(0 + 3) \times 2]\}$. The mismatches in a mutagenic oligo do not count when estimating T_m. These T_m estimates, though crude, are generally conservative approximations of the annealing temperature needed for an oligo to be used in PCR (e.g., a 50°C annealing step will work with these primers). They lead to priming and overlap regions being somewhere around 13–20 bases long, depending on the GC content. The "~50°C rule" has resulted in quite reliable primers, but it does not represent a systematic effort to determine the absolute minimum lengths that SOEing primers can be (*see* Note 5).
4. The normal considerations in primer design also apply, such as avoiding complementarity between or within primers, and so forth.
5. The flanking primers a and d should also be capable of priming at 50°C, and may include restriction sites at the ends to facilitate cloning of the product.
6. "Megaprimer" reactions. Modifications of the gene-SOEing protocol make it possible to use only one primer at the recombinant joint *(13,14)*. Rather than using the symmetrical approach shown in Fig. 2, in which the recombination event is formation of a giant primer-dimer, the top strand of AB can be used as a

megaprimer in place of primer c directly on template gene II to make the final product AD. For recombination of two sufficiently different genes, only the recombinant product can be amplified using primers a and d, since each of these primers will match one of the template genes but not the other. Therefore, a and d can be included in the reaction to drive the synthesis of the productive strands, and to amplify the product as soon as it forms *(13,15)*. For megaprimer mutagenesis, however, primer a cannot be included in the final reaction containing primer d because a and d would simply amplify the wild-type sequence from the template. This means that product AB must be added in sufficient quantity to supply all of the megaprimer strand necessary to generate AD *(14)*.

3.2. PCR Generation of Intermediate Products AB and CD (Fig. 2)

1. Set up two separate PCRs as follows (*see* Note 6):

Ingredients	PCR 1	PCR 2
10X buffer II	10 µL	10 µL
10 m*M* $MgCl_2$	~15 µL	~15 µL
10X dNTPs	10 µL	10 µL
5' primer (10 µ*M*) a,	10 µL	c, 10 µL
3' primer (10 µ*M*) b,	10 µL	d, 10 µL
template	gene I, ~0.5 µg	gene II, ~0.5 µg
H_2O	to 100 µL	to 100 µL
Taq polymerase	~3 U	~3 U
Product	AB	CD

2. Amplify by PCR using the following cycle profile:

20–25 main cycles	94°C, 1 min (denaturation)
	50°C, 1 min (annealing)
	72°C, 1 min (extension)

3.3. Purification of Intermediates (see Note 7)

1. Run the PCR products on an agarose gel to size-purify them (*see* Note 8).
2. The method you use to extract the DNA from the agarose also depends on the size of the piece of DNA with which you are working. Purify fragments larger that ~300 bp using GeneClean. Be aware that GeneClean requires the use of TAE (not TBE) electrophoresis buffer.
3. Smaller fragments are not recovered efficiently by this method, so you should use another procedure, such as electroelution. Cut a well in front of the band, run the band into the well, and recover the DNA in the well with a pipet (this procedure is described in ref. *16*).
4. Recover the DNA from the buffer by ethanol precipitation. Add 0.1 vol of 3*M* sodium acetate, pH 5.2, 1 µL of yeast tRNA (10 mg/mL) as a carrier, and 2.5 vol of cold 95% ethanol. Incubate on dry ice for 15 min, and then spin at approx 10,000*g* for 15 min.

3.4. Generation of the Recombinant Product AD

1. Set up the overlap extension (SOE) reaction as follows (*see* Note 9):

Ingredients	
10X buffer II	10 µL
10 m*M* $MgCl_2$	~15 µL
10X dNTPs	10 µL
5' primer (10 µ*M*) a,	10 µL
3' primer (10 µ*M*) d,	10 µL
Intermediate 1	AB
Intermediate 2	CD
H_2O	to 100 µL
polymerase	~3 U

2. Amplify by PCR using the following cycle profile (*see* Note 10):

20–25 main cycles	94°C, 1 min (denaturation)
	50°C, 1 min (annealing)
	72°C, 1 min (extension)

4. Notes

1. Primer purification. The final step in oligonucleotide synthesis by the phosphoramidite method involves treatment in NH_4OH. The ammonium hydroxide should be evaporated off using a SpeedVac (Savant, Farmingdale, NY). (Please note that ammonia vapors are extremely hard on vacuum pumps! Either install an appropriate chemical trap in your vacuum line or use a water-powered aspirator instead.) Resuspend the dried primer in water and desalt it over a Sephadex G25 column. Prepacked columns are available for this purpose (NAP-10 columns, Pharmacia LKB, Uppsala, Sweden). No special buffer is needed: Distilled water works fine. Extensive (and

expensive!) primer purification schemes, such as acrylamide gel electrophoresis or HPLC, are not necessary. If you plan to purchase primers commercially, they are available already desalted for ~$2.25/base (Integrated DNA Technologies, Coralville, IA). There is no need to pay more than this.

2. The starting templates contain the gene sequences that you want to recombine into a tailor-made molecule. Any template suitable for PCR, such as reverse-transcribed RNA, can be used for gene SOEing (e.g., *see* ref. *10*). However, because high template concentrations minimize the probability of the polymerase introducing errors into the sequences (*see* Notes section), your starting templates will probably be cloned genes in plasmids. My coworkers and I have used both cesium chloride and alkyline lysis-purified plasmids with equal success (*see* ref. *16* for plasmid purification protocols).
3. The single simplest and most complete way to mess up a SOEing reaction is to have mistakes in the primer sequences (R.H., personal experience!). As PCR reactions go, amplifying an insert from a plasmid that is present in microgram amounts is like falling off the proverbial log. If this reaction cannot be made to work after the normal titrations of Mg^{2+}, template, and so forth, then something is drastically wrong, and you should recheck the design of your primers. Similarly, if products AB and CD, when mixed together in near-microgram amounts in a SOE reaction, completely fail to produce a recombinant product, the sequences of the primers in the overlap region should be immediately suspect. Writing out the sequence of the desired product and making sure that the SOEing primers each match one strand of the desired product at the recombination point is helpful.
4. The overlap region does not all have to be added to one primer. If, for example, instead of adding a 17-bp overlap region to primer b we had added 9 bases complementary to primer c to the 5' end of b and 8 bases complementary to b to the 5' end of c, the overlap between AB and CD would still have been 17 bp (*see* ref. *7*). This approach avoids using very long primers. However, since making a 40- or 50-mer is now routine, there is usually no need to split the overlap region.
5. The optimal length to design the overlap region has not been settled. Regions as long as 164 bp *(15)* or as short as 12 bp *(17)* have been reported. Although the "50°C rule" reliably produces overlaps that are long enough to work, it does not indicate the minimum workable length. For example, the 12-bp-long overlap region reported in ref. *17* has an estimated T_m of 34°C, although it was used at an annealing temperature of 50°C. On theoretical grounds, however, merely producing the recombinant molecule is not the only consideration to bear in mind. A very short overlap region might lead to an inefficient SOEing reaction, requiring the final product to be amplified through more rounds of PCR. This in turn may increase the error frequency.
6. Errors introduced by polymerase. Studies have demonstrated that *Taq* polymerase is capable of high fidelity DNA synthesis under PCR conditions *(18)*. Clones produced by a single round of overlap extension have error frequencies of around 1 in 4000 bases *(6)*, whereas more complicated constructs involving random SOEing reactions lead to slightly higher frequencies (~1 in 1800; ref. *7*). Because of the possibility that random mutations will be introduced by polymerase errors, several precautions are in order. First, the highest concentration of template plasmid consistent with amplification should be employed (this will probably be around 500 ng of plasmid in 100 µL). This minimizes the number of rounds of replication required to produce enough product with which to work, and gives the polymerase fewer opportunities to make errors. Since the reaction plateaus after producing a certain amount of product, it is probably not necessary to minimize the number of heating/cooling cycles to which the samples are subjected. Second, the lowest concentration of magnesium compatible with amplification should be used, as error rates increase with increasing $[Mg^{2+}]$ *(18)*. Finally, for many applications, it is advisable to sequence the final product to ensure that it is free of errors. For this reason, a "cassette"

approach, in which PCR manipulations are performed on a small recombinant segment, which is then ligated into a vector containing the remaining portions of the construct, may in some cases be preferable to SOEing directly into a vector. Around 300–500 bp is a convenient size for a cassette because it is large enough to handle conveniently yet small enough to sequence quickly. It should be pointed out that other thermostable DNA polymerases (i.e., Vent polymerase, New England Biolabs, Beverly, MA) are reportedly capable of significantly higher fidelity synthesis than *Taq*, and can be used for overlap extension *(19)*.

7. Ho et al. *(6)* found that gel purification of the intermediate products AB and CD led to a cleaner reaction and increased product yield. Purification of intermediates may be most important when the initial template concentration is high. Gel purification removes not only the template plasmids, but also open-ended primer extension products, which, coming from the template, may be longer than the PCR product. These open-ended products may not be obvious on an ethidium-stained gel because they are of indeterminate length and possibly single-stranded, but they have the potential to generate unwanted side products. Nevertheless, other workers have successfully used less extensive purification schemes *(5)* or none at all *(10,13)*.
8. The percent agarose you use depends on the size of the products you are isolating; smaller fragments need a higher percentage of agarose. For up to 1% agarose, "regular" agarose is fine, but for higher percentages, NuSieve (FMC BioProducts, Rockland, ME) gives better resolution. Please note that NuSieve is used to supplement regular agarose: The first 1% agarose is the regular variety, and only the additional percentage is NuSieve (up to a total of 4%).
9. Using large quantities of the intermediates should minimize polymerase errors (*see* Note 6). About 25% of what you recover from the gel should be plenty if the PCRs worked well, and this will leave you some extra in case you have to repeat it. The recombinant may now be cloned, or otherwise used, like a normal PCR product.
10. Related applications. Although the concept is simple, overlap extension is a tremendously powerful technology, and the reader is encouraged to spend some time contemplating modifications and applications. A general theoretical review of the subject is given in ref. *15*. Some of the more important and thought-provoking technical developments related to synthetic uses of PCR are given in refs. *14,20–22; see also* Chapters 24, 26, 27, and 28 in reference *23*.

Acknowledgments

I am deeply grateful to my colleagues Steffan Ho, Jeff Pullen, Henry Hunt, Zeling Cai, and Larry Pease, for making it possible for me to participate in the development of this technology, and to Bianca Conti-Tronconi for the use of her computer.

References

1. Mullis, K. B. and Faloona, F. A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Methods Enzymol.* **155,** 335–350.
2. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.* **51,** 263–273.
3. Kadowaki, H., Kadowaki, T., Wondisford, F. E., and Taylor, S. I. (1989) Use of polymerase chain reaction catalysed by Taq DNA polymerase for site-specific mutagenesis. *Gene* **76,** 161–166.
4. Vallette, F., Mege, E., Reiss, A., and Milton, A. (1989) Constuction of mutant and chimeric genes using the polymerase chain reaction. *Nucleic Acids Res.* **17,** 723–733.
5. Higuchi, R., Krummel, B., and Saiki, R.K. (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16,** 7351–7367.
6. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77,** 51–59.
7. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77,** 61–68.
8. Horton, R. M., Cai, Z., Ho, S. N., and Pease, L. R. (1990) Gene splicing by overlap extension: tailor-

made genes using the polymerase chain reaction. *BioTechniques* **8,** 528–535.
9. Kain, K. C., Orlandi, P. A., and Lanar, D. E. (1991) Universal promoter for gene expression without cloning: expression-PCR. *BioTechniques* **10,** 366.
10. Davis, G. T., Bedzyk, W. D., Voss, E. W., and Jacobs, T. W. (1991) Single Chain Antibody (SCA) encoding genes: one-step construction and expression in eukaryotic cells. *Biotechnology* **9,** 165–169.
11. Daughtery, B. L., DeMartino, J. A., Law, M.-F., Kawka, D. W., Singer, I. I., and Mark, G. E. (1991) Polymerase chain reaction facilitates the cloning, CDR-grafting, and rapid expression of a murine monoclonal antibody directed against the CD18 component of leukocyte integrins. *Nucleic Acids Res.* **19,** 2471–2476.
12. Suggs, S.V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J., Itakura, K., and Wallace, R. B. (1981) Use of synthetic oligo-deoxyribonucleotides for the isolation of cloned DNA sequences, in *Developmental Biology Using Purified Genes* (Brown, D. D. and Fow, C. F., eds.), Academic, New York, pp. 683–693.
13. Yon, J. and Fried, M. (1989) Precise gene fusion by PCR. *Nucleic Acids Res.* **17,** 4895.
14. Sarkar, G. and Sommer, S. S. (1990) The "megaprimer" method of site-directed mutagenesis. *BioTechniques* **8,** 404–407.
15. Horton, R. M. and Pease, L. R. (1991) Recombination and mutagenesis of DNA sequences using PCR, in *Directed Mutagenesis: A Practical Approach.* (McPherson, M. J., ed.) IRL, Oxford, pp. 217–247.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed.) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
17. Yolov, A. A. and Shaborova, Z. A. (1990) Constructing DNA by polymerase recombination. *Nucleic Acids Res.* **18,** 3983–3986.
18. Eckert, K. A. and Kunkel, T. A. (1990) High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res.* **18,** 3739–3744.
19. Hanes, S. D. and Brent, R. (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. *Science* **251,** 426–430.
20. Rudert, R. A. and Trucco, M. (1990) DNA polymers of protein binding sequences generated by PCR. *Nucleic Acids Res.* **18,** 6460.
21. Shuldiner, A. R., Scott, L. A., and Roth, J. (1990) PCR-induced (ligase-free) subcloning: a rapid reliable method to subclone polymerase chain reaction (PCR) products. *Nucleic Acids Res.* **18,** 1920.
22. Jones, D. H. and Howard, B. H. (1990) A rapid method for site-specific mutagenesis and directional subcloning by using the polymerase chain reaction to generate recombinant circles. *BioTechniques* **8,** 178–183.
23. White, B. A. (ed.) (1993) *PCR Protocols: Current Methods and Applications,* Humana, Totowa, NJ.