Combinatorial Recombination of Gene Fragments to Construct a Library of Chimeras

Mary Farrow[1] and Frances H. Arnold[1]

[1]Caltech, Pasadena, California

ABSTRACT

Recombination of distantly related and nonrelated genes is difficult using traditional PCR-based techniques, and truncation-based methods result in a large proportion of nonviable sequences due to frame shifts, deletions, and insertions. This unit describes a method for creating libraries of chimeras through combinatorial assembly of gene fragments. It allows the experimenter to recombine genes of any identity and to select the sites where recombination takes place. Combinatorial recombination is achieved by generating gene fragments with specific overhangs, or sticky ends. The overhangs permit the fragments to be ligated in the correct order while allowing independent assortment of blocks with identical overhangs. Genes of any identity can be recombined so long as they share 3 to 5 base pairs of identity at the desired recombination sites. Simple adaptations of the method allow incorporation of specific gene fragments. *Curr. Protoc. Protein Sci.* 61:26.2.1-26.2.20. © 2010 by John Wiley & Sons, Inc.

Keywords: chimera • recombination • directed evolution • protein design • library creation • combinatorial library

INTRODUCTION

This unit contains protocols that allow creation of libraries of chimeras through combinatorial assembly of gene fragments (Fig. 26.2.1). The experimenter specifies where recombination is allowed to occur, giving the method several advantages over annealing-based and truncation-based methods for chimeragenesis. Among these advantages are that many genes of any identity can be recombined with up to eight recombination sites, and no frame-shifted chimeras are produced. Additionally, unlike total gene synthesis using oligonucleotides, many different chimeras are made simultaneously, it is easy to create full-length genes with few or no point mutations, and fewer oligonucleotides are required for the procedure, reducing the cost of constructing chimeras. Furthermore, it is easy to modify the protocol to incorporate desirable insertions, deletions, or rearrangements in the genes being recombined.

Combinatorial recombination of parent genes, outlined in Figure 26.2.1, is achieved by assembling blocks of sequence with specific, unique overhangs that are generated using Type IIB restriction enzymes. A "block" is a portion of sequence that has a particular set of overhangs and position in the gene; it may be from any parent. A "fragment" is a block from a specific parent gene. The overhangs permit even unrelated fragments to be ligated in the correct order while allowing independent assortment of fragments with identical overhangs. Genes of any identity can be recombined so long as they share 3 to 5 base pairs of identity at the desired recombination site.

The Basic Protocol focuses on the mechanistic details necessary for creating a combinatorial gene library. The Basic Protocol limits the size of the blocks and the number of recombination sites. It is also intended to incorporate fragments from each parent gene at all blocks. Alternate protocols are provided that circumvent these constraints.
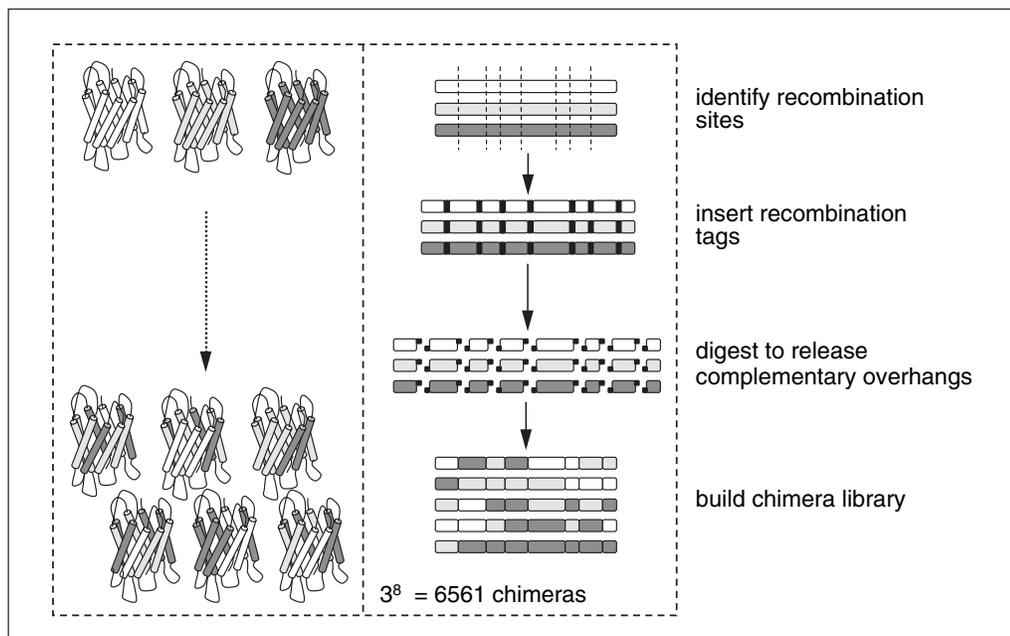
Protein Engineering

Supplement 61

Figure 26.2.1 An overview of the chimeragenesis protocol. By combinatorially recombining three genes with eight independently assorting blocks, $3^8 = 6561$ possible chimeras can be formed in a single experiment.

Choosing the recombination sites effectively is a key part of a successful experiment and is discussed in the Commentary. A simplified protocol is also included as a reference for experienced users.

STRATEGIC PLANNING

Inserting Tag Sequences for Recombination

The Basic Protocol details a method for generating recombination sites by incorporating "tags" (Type IIB restriction enzyme recognition sequences) into parent genes. The incorporation of these tags, construction phase 1 in the Basic Protocol, is achieved through multiple rounds of PCR. Alternate Protocol 1, in which parent genes are synthesized with recombination tags already inserted, allows the researcher to omit the first construction phase and significantly decreases the total time of library development.

Block Choices

Choosing the number and locations of recombination sites is an essential component of the experimental design, and a more detailed discussion of how to choose recombination sites is provided in the Commentary. However, a few details resulting from these choices affect the construction methodology and are discussed here. First, the smallest block should be larger than ~40 bp for the standard procedure to work correctly. Blocks smaller than this are lost during the purification steps that remove the tag sequences (~30 bp). Alternate Protocol 2 provides options for using blocks smaller than ~40 bp. The recombination sites chosen for this protocol must have compatible overhangs for this method to work. Therefore, if experimental constraints require that recombination occur at a site in which parent gene sequences differ, the parent sequences must be mutated so that the sites are identical.

Number of Recombination Sites

The complexity of the procedure increases significantly with the number of recombination sites. This protocol is based on ligation reactions, which are inherently

inefficient. The authors have found it difficult to ligate more than four DNA pieces together in a single reaction and achieve sufficient yield. Therefore, in order to have more than four independent blocks, two separate ligation steps are performed with a cloning step between them. Choosing four or eight blocks takes the best advantage of these two ligation steps. Choosing a different number of blocks results in fewer than four blocks in one or more of the mini-libraries. Using four or fewer recombination sites allows a significantly shorter procedure, where the final construction phase can be omitted. Alternate Protocol 3 details a protocol that incorporates more than eight blocks.

Restriction Sites

The restriction sites used for cloning during construction and for the final cloning of the library should be different. This can be achieved by using a different plasmid for the construction steps, or by using a different set of restriction sites for construction. Using different sites prevents cross-ligation during the final ligation phase of construction. It also allows the mini-libraries to be inversely cloned into plasmids that have leaky or constitutive expression without fear that low-level expression of the mini-libraries will affect the fragment proportions in the final library. Several restriction endonucleases are used as examples in the Basic Protocol but can be substituted with any restriction enzyme of the researcher's choice.

CONSTRUCTION OF A COMBINATORIAL GENE LIBRARY

BASIC PROTOCOL

This protocol is split into six phases—two design phases and four construction phases. The first design phase determines the recombination sites and the gene fragments between them. Additionally, the tag sequences used to generate specific overhangs at the recombination sites are designed. PCR primers to insert the tags into the genes are designed during the second design phase.

The construction phases are summarized in Figure 26.2.2. The first construction phase introduces tag sequences into the genes that will be recombined. The blocks are also divided into groups of four or fewer to facilitate ligation steps during construction phase 3. The product of the first construction phase is a set of plasmids; each group of four blocks has a plasmid for each gene recombined. For example, three genes with eight blocks (two groups) would have a set of $3 \times 2 = 6$ plasmids (Fig. 26.2.2). These plasmids should be sequenced to ensure that no mutations have been introduced during the construction process.

The second construction phase generates sequence blocks that ligate independently through cleavage of the tag-inserted genes with a Type IIB restriction enzyme that leaves specified overhangs. The third construction phase ligates these blocks together to generate mini-libraries that contain one-half of the full-length genes. These mini-libraries are cloned separately and provide an opportunity to check for proper construction. The final construction phase ligates the mini-libraries together to generate full-length genes and a full-size library.

Materials

Custom-synthesized oligonucleotides for PCR primers
Gene sequences for proteins to be recombined
High-fidelity DNA polymerase for PCR, e.g., *Phusion*
DNA gel extraction kit (Zymogen Zymoclean Gel DNA Recovery Kit *or* Qiagen Qiaquick Gel Extraction Kit)
*Pst*I restriction endonuclease
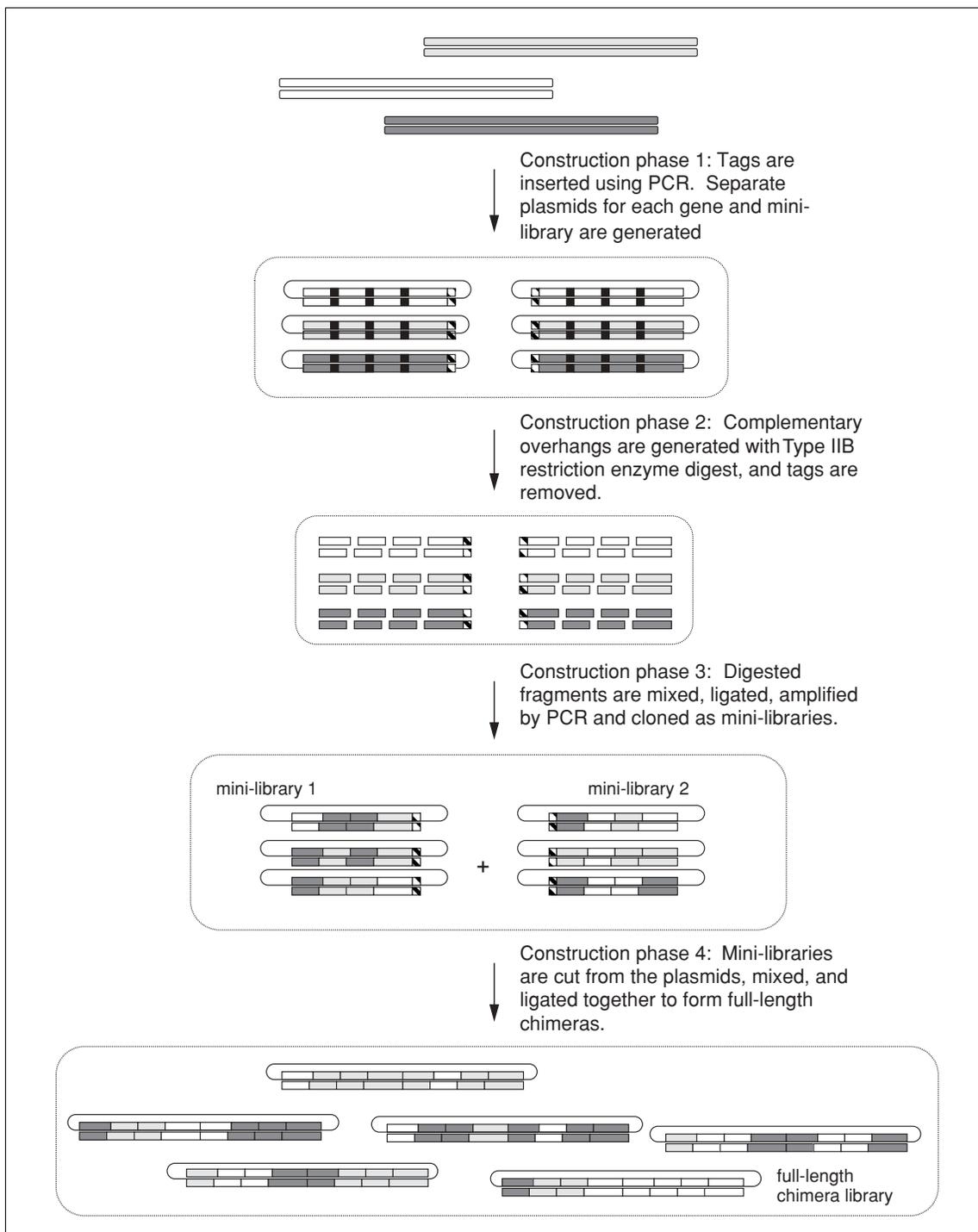*Sal*I restriction endonuclease

26.2.3

Figure 26.2.2 The chimera construction process is broken down into the four phases. This figure demonstrates the recombination of three parental genes, broken into eight blocks (seven recombination sites). First, tag sequences that will allow specific overhangs to be generated are inserted into the genes using PCR. Next, the tag-inserted genes are cut with a Type IIB restriction enzyme to expose the DNA fragments with desired overhangs, and the tag sequences are removed. The DNA fragments are then ligated together to form two mini-libraries, which are cloned individually. Finally, the two mini-libraries are ligated to form full-length genes. Sequences cloned and transformed into *E. coli* are shown with the plasmid backbone.

Combinatorial Recombination of Gene Fragments

26.2.4

DNA purification kit (Zymogen DNA Clean and Concentrator *or* Qiagen Qiaquick PCR Purification Kit)

Plasmid for construction (e.g., pBS KS+; Stratagene)

Calf alkaline phosphatase (Tabor, 1987a)

*Bsa*XI restriction endonuclease

T4 DNA ligase (Tabor, 1987b)

*Nde*I restriction endonuclease

150 × 15–mm LB agar plates (*APPENDIX 4A*) with antibiotic for transformation

*Sap*I restriction endonuclease

*Bam*HI restriction endonuclease

*Hind*III restriction endonuclease

Plasmid for final library cloning and expression (e.g., pET28)

16°C water bath

Additional reagents and equipment for polymerase chain reaction (*APPENDIX 4J*), agarose gel electrophoresis (*APPENDIX 4F*), restriction enzyme digestion (*APPENDIX 4I*), introduction of plasmid DNA into cells (*APPENDIX 4D*), preparation of plasmid DNA (*APPENDIX 4C*), DNA sequencing (Ausubel et al., 2010, Chapter 7), and quantitation of nucleic acids with absorption spectroscopy (*APPENDIX 4K*)

NOTE: All restriction enzymes mentioned in this protocol are examples, and may be replaced with other restriction enzymes compatible with the design of the experiment. *Sap*I and *Bsa*XI cleave outside their recognition sequences and should only be replaced with enzymes that have similar properties.

Design phase 1: Choose recombination sites and design tag sequences

1. Choose the Type IIB restriction site to be used during construction to generate 5′ overhangs.

   Type IIB restriction enzymes cut asymmetrically outside of their recognition sites, allowing any desired overhang to be generated. The authors have used BsaXI successfully. Because BsaXI leaves 3-bp overhangs, the recombination sites require 3 bp of identity in the parental sequences.

2. Choose recombination sites using methods discussed in Background Information.

   The recombination sites chosen using these methods may not be experimentally feasible and may have to be adjusted. A recombination site should have identical base pairs in all the parental genes to form the overhangs, as shown for 3 bp in Figure 26.2.3A. The number of base pairs required varies depending on the Type IIB enzyme chosen.

3. Make a list of the recombination sites with the overhangs (Fig. 26.2.3B).

   The DNA between recombination sites is referred to generally as a <u>block</u>. A <u>fragment</u> is a block from a specific gene.

4. Starting at the N-terminus, group the blocks so that four or fewer sequential blocks are in each group.

   These groups will make up the mini-libraries that are each treated separately until the final construction phase (Fig. 26.2.3B).

5. Separately list the recombination site overhangs internal to each mini-library.

   There will be one recombination site that is not in either group. This site will be used to join the two mini-libraries in the final construction phase (Fig. 26.2.3B).

6. Ensure that the overhangs will not cross-ligate within each list of recombination sites.

   They must be unique and nonpalindromic. For example, CTA and TAG will cross-ligate, and AATT will cross-ligate with itself.
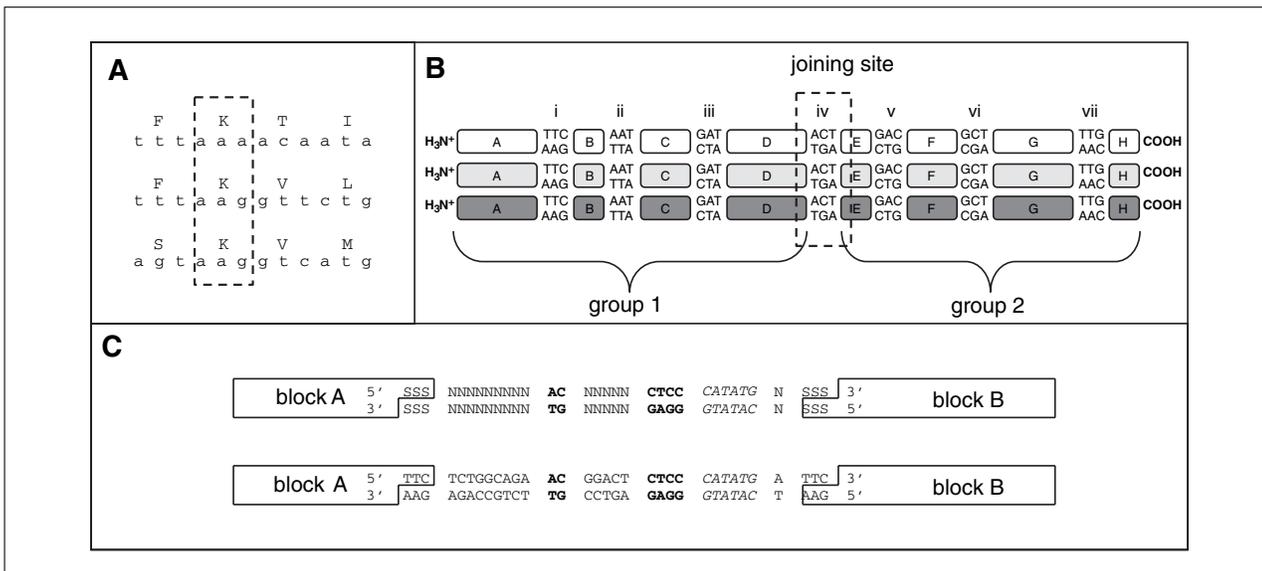
Figure 26.2.3 (**A**) A recombination site must have at least 3 bp that are, or can be made, identical for all the genes to be recombined. For the recombination site shown, the last base pair of the codon can be changed to accommodate the recombination site, without changing the amino acid sequence of the proteins. (**B**) Recombination sites are shown with the unique overhangs and then grouped for the mini-libraries. Within each group, these overhangs must be different. However, the same overhang can be repeated in different groups (e.g., recombination sites II and V could have the same overhang, because they are in different mini-libraries). (**C**) The template for the tag sequence contains the *Bsa*XI (bold) and *Nde*I (italics) restriction sites. A tag for each recombination site is formed by first adding the overhang bases, as listed in panel B. The remaining bases in the tag are filled in using bases that will not cross-anneal during PCR.

7. Design template tag sequences.

   For each mini-library, a set of tag sequences will be inserted into the gene at the recombination sites. Some portions of these tag sequences are identical for all recombination sites. To design the tags, first make a template tag as shown in Fig. 26.2.3C. The BsaXI recognition site is shown in bold and a second restriction site (NdeI) is shown in italics. S represents the desired overhang in each tag. N represents bases in the tag that can be any base in this template, but will be specified as the tags are designed.

   This template will be different for different Type IIB restriction enzymes, depending on their recognition and cleavage sites. The second restriction enzyme will be used to remove unwanted tags during the procedure and should be robust and ideally cut in the same buffer as the Type IIB enzyme. For demonstration purposes here, BsaX I is used as the IIB restriction site and NdeI as the secondary site. Neither enzyme should cut within any of the genes.

8. For each recombination site, except the final joining site (see step 5), copy the template tag and replace the S portions with the overhang bases as shown (Fig. 26.2.3C).

9. Next, replace the N portions of each tag with bases that are ~50% GC and that will not cross-anneal during PCR or generate a restriction site used during the construction process (Fig. 26.2.3C). The same tag sequence, except for the overhangs, can be used for multiple recombination sites, as long as they do not generate unwanted restriction sites. However, it may be necessary to design several unique tag sequences. Once tags are designed, examine the sequences of the tag-inserted parent genes to ensure that only the desired restriction sites are present.

Design phase 2: Design primers

Each primer pair amplifies a DNA fragment with portions of a unique tag at the 5′ and 3′ ends, except for the fragments that correspond to the N- and C-terminal portions of the parent protein. These fragments have portions of a tag only at the 3′ or 5′ end, respectively (Fig. 26.2.4A). Once all fragments have been amplified, their overlapping

regions are used in PCR-based assembly of the full-length construct (Fig. 26.2.4B). Most of this phase may be omitted if genes were synthesized with tags already inserted, as described in Alternate Protocol 1, and the experimenter may go to construction phase 2. See Alternate Protocol 1 for more details.

10. For each gene, design primers to amplify fragments with portions of each tag.

> *Template primers are shown in Figure 26.2.5. The 5' portions of the primers match the tag sequence and overlap ($T_m \geq \sim 50°C$) to allow PCR-based construction of the gene with tags inserted (Fig. 26.2.5A). The 3' portion of the primer is unique for each gene.*
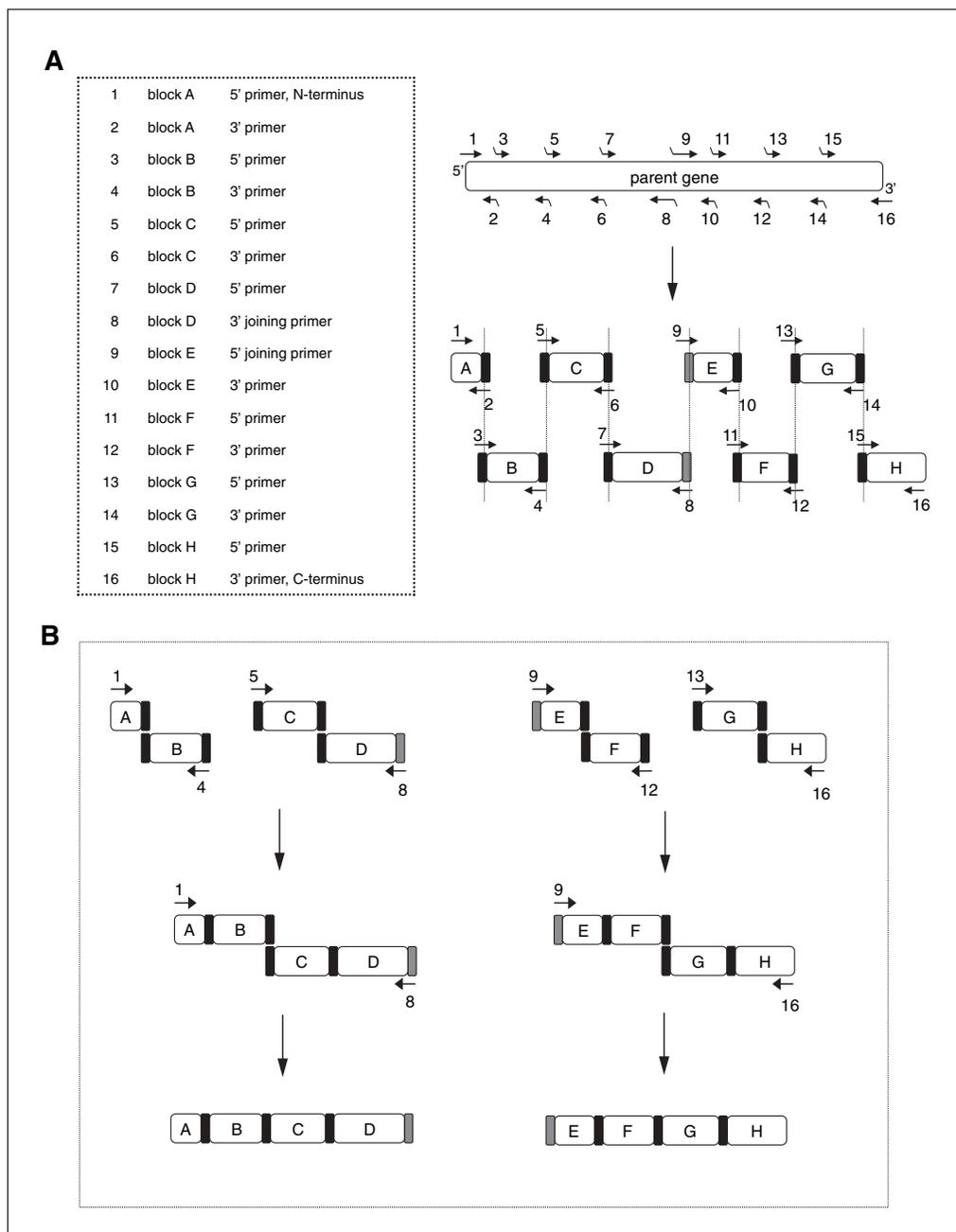

Figure 26.2.4 (**A**) An overview of the PCR protocol necessary to form the tag-inserted sequences. Primers to insert the tag sequences into each gene are used to amplify the fragments before and after the recombination site, and to form part of the tag. (**B**) Fragments overlap in order to allow PCR-based assembly of the full-length construct with the tags inserted.

11. Design N- and C-termini primers for each gene.

   The N- and C-termini primers need to contain two restriction sites, each at the 5′ end for cloning into plasmids. The restriction sites for the final expression plasmid should be closest to the gene DNA, and the restriction sites for the construction plasmid should be placed at the 5′ end of the primer (Fig. 26.2.5B). SalI and PstI are used here as example restriction sites for the N- and C-termini during construction and BamHI and HindIII restriction sites for the final cloning. The reason for using two different plasmids is discussed in Critical Parameters. However, two sets of restriction sites are required for the procedure.

12. Design primers for the junction of mini-libraries.

   These primers allow the mini-libraries to be cloned independently during the first three stages of construction and then to be joined during the final stage. This also involves two restriction sites in each primer. Each primer must contain a restriction site that allows cleavage outside of the recognition site to expose the desired 5′ overhang (Fig. 26.2.5C). Choose this restriction enzyme and design the primer to ensure correct cleavage to release overhangs.
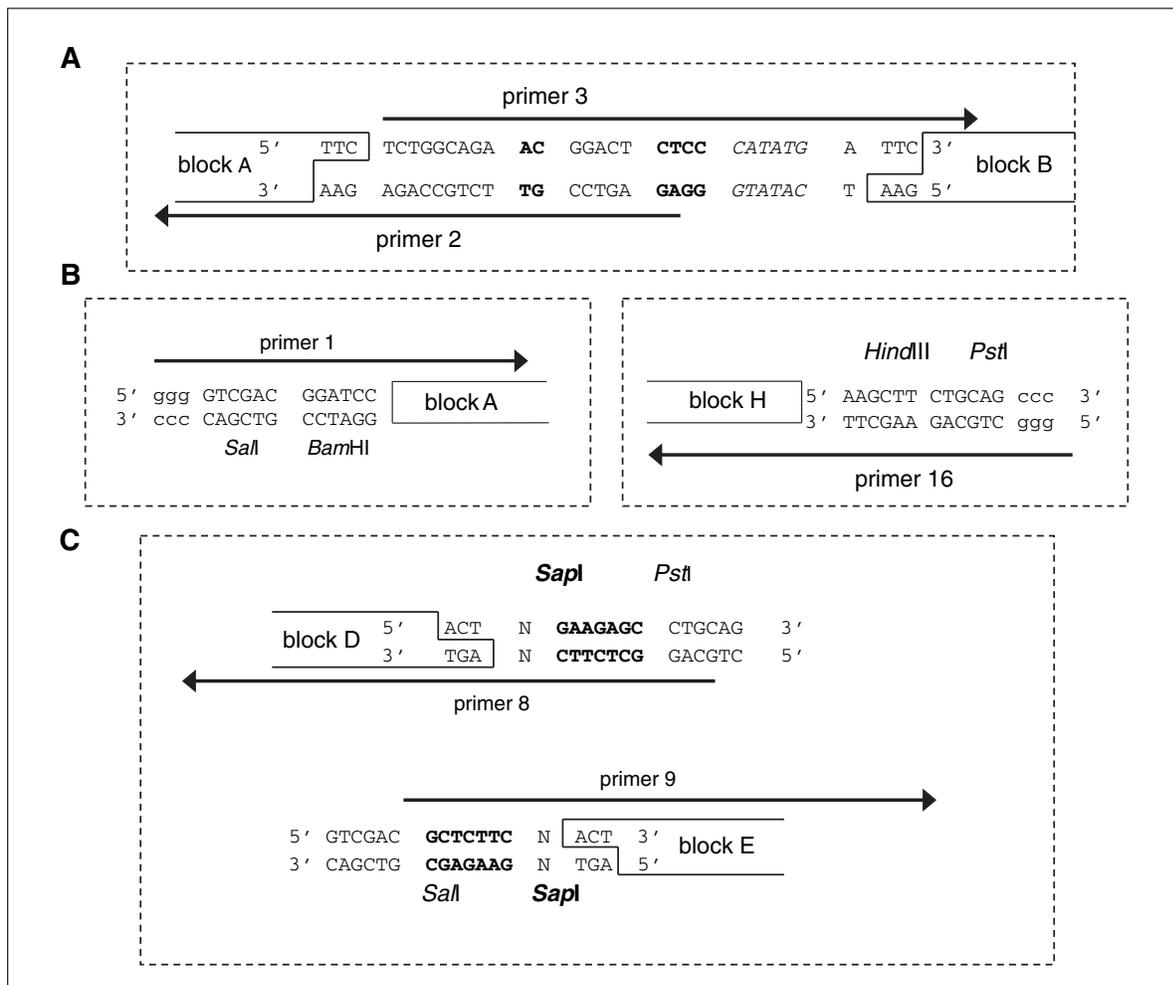

Figure 26.2.5 (**A**) Primers that contain overlapping portions of a *Bsa*XI-based tag for insertion of a recombination site between two blocks. (**B**) The N- and C-termini primers each contain two restriction sites, one for cloning during construction (*Sal*I, or *Pst*I) and a second for cloning into the final expression plasmid (*Bam*HI, or *Hind*III). Lowercase letters are bases added to the 5′ end of the primer to ensure efficient restriction enzyme cleavage. (**C**) There are joining primers at the ends of the mini-libraries. These primers contain *Sap*I restriction sites (bold) to generate the overhangs and secondary sites (*Pst*I, or *Sal*I) for cloning.

For this step, it is better to use a type II restriction site with a smaller number of base-pairs between the recognition and cleavage sites than BsaXI. SapI and EarI generally produce good results, and SapI is used here as an example. It is important to note that many of these enzymes are asymmetric cutters, and the site must be designed in the correct orientation. In addition, restriction sites for cloning during construction are required at the 5′ ends of each joining primer (Fig. 26.2.5C). These restriction sites should be the same as those used for cloning the N- and C-termini into the construction plasmid.

The primer that forms the C-terminal portion of the first mini-library (amplifying block D) has a PstI site, and the primer that forms the N-terminal portion of the second mini-library (amplifying block E) has a SalI site, so that both mini-libraries can be cloned into the same plasmid.

At this point one should have a list of primers that includes a forward and reverse primer for each fragment.

13. Examine primers and starting gene sequences to ensure that none of the restriction sites used during construction are present except where desired.

Construction phase 1: Generate tag-inserted plasmids

This phase may be omitted if genes were synthesized with tags already inserted, as described in Alternate Protocol 1. The experimenter may skip these steps and go to construction phase 2. See Alternate Protocol 1 for more details.

14. Use PCR (*APPENDIX 4J*) to amplify each fragment with the appropriate primers designed above. Use the full-length gene as template.

    A high-fidelity polymerase, such as Phusion, is recommended for all PCR.

15. Use preparatory agarose gel electrophoresis (*APPENDIX 4F*) to separate the individual PCR products, excise the band of the correct size, and purify it from the agarose using a gel extraction kit.

16. For each gene, connect the PCR products of adjacent fragments using PCR. Use the two fragments amplified above as template and the outer 5′ and 3′ primers from the reactions carried out previously (Fig. 26.2.4B) as primers.

17. Repeat steps 15 and 16 on the PCR products until the full-length (4 or fewer) group of fragments is generated.

18. Digest the product with *Pst*I and *Sal*I (*APPENDIX 4I*) and remove the enzymes using a DNA purification kit, or heat-kill the enzymes according to the manufacturer's instructions. Ligate the product into the construction plasmid cut with the same enzymes using T4 ligase according to the ligase manufacturer's instructions. Transform the product into *E. coli* (*APPENDIX 4D*).

19. Pick several colonies, isolate the DNA, and perform an analytical digest with *Sal*I and *Pst*I to verify full-length tag-inserted sequences (*APPENDIX 4C*). Sequence the DNA (Ausubel et al., 2010, Chapter 7) to confirm that no mutations were introduced during the PCR.

    There should be a separate plasmid for each gene and each mini-library (i.e., for three genes and two mini-libraries, six plasmids are needed).

Construction phase 2: Generate library pieces

This construction phase is performed separately with each mini-library (Fig. 26.2.2).

20. Prepare a large quantity (20 to 100 μg scale or larger; *APPENDIX 4C*) of each tag-inserted plasmid.

Instead of using plasmid DNA, the tag-inserted gene can be PCR amplified using Phusion polymerase. This may result in a few extra mutations in the final library, but can save some time, as steps 21 to 24 are no longer necessary.

21. Quantify the concentration of DNA using absorption spectroscopy (*APPENDIX 4K*) and mix together 3 pmol of each tag-inserted plasmid.

    3 pmol is roughly 1 µg/kb of plasmid (i.e., 3 pmol is 9 µg of a 3-kb plasmid). This amount does not need to be exact, but is a guide to the appropriate scale necessary for the procedure. It is important that the DNA be mixed in the molar ratio desired in the library.

    Even if the tag-inserted gene is PCR amplified (see step 20), the resulting DNA must be quantified and mixed appropriately.

22. Digest with *Pst*I and *Sal*I restriction endonucleases (*APPENDIX 4I*) to remove the insert from the construction plasmid.

    This step is unnecessary if the tag-inserted genes were PCR amplified in step 20.

23. Dephosphorylate with calf alkaline phosphatase according to the manufacturer's instructions (also see Tabor, 1987a).

    This will ensure that no ligation occurs at the termini of the insert, which would interfere with the rest of the reaction during construction phase 3.

    This step is unnecessary if the tag-inserted genes were PCR amplified in step 20.

24. Use preparatory agarose gel electrophoresis (*APPENDIX 4F*) to isolate the dephosphorylated inserts and remove the plasmid backbone. Purify the inserts using a gel extraction kit.

    This removes the unnecessary plasmid DNA from subsequent reactions where it may hinder the process.

    This step is unnecessary if the tag-inserted genes were PCR amplified in step 20.

25. Cut the insert with *Bsa*XI (or other type IIB restriction site, as discussed during the design phase) according to the manufacturer's instructions to expose the 5′ overhangs on the blocks.

26. Remove the released tags and *Bsa*XI using a DNA purification kit according to the manufacturer's instructions.

    It is important that the tags be removed from the reaction. Not all DNA purification kits remove small pieces of DNA. Blocks smaller than the cutoff for the DNA purification kit are also removed during this step. Running an analytical agarose electrophoresis gel to verify that smaller fragments are not lost is often worthwhile here. If using small blocks, see Alternate Protocol 2.

Construction phase 3: Generate multiple mini-libraries

This construction phase is done separately for each mini-library (Fig. 26.2.2).

27. Ligate the fragments together using T4 ligase (also see Tabor, 1987b) for 4 hr at 16°C. Heat-kill the ligase after the reaction according to the ligase manufacturer's instructions.

    The ideal concentration of 5′ ends in the reaction is 0.15 µM. Assuming a 70% yield on the above steps starting with N × 3 pmol of each plasmid (where N is the number of genes with which the procedure started), N × 50 µl is ~0.15 µM.

    There should be N × 6 Weiss units of ligase added to the reaction, proportional to the amount of DNA. One Weiss unit of T4 DNA ligase converts 1 nmol of ^{32}P from pyrophosphate into Norit-adsorbable material in 20 min at 37°C (Weiss et al., 1968). One Weiss unit equals ~67 cohesive-end units.

28. Digest the ligated insert with *Bsa*XI and *Nde*I according to the manufacturer's instructions (also see *APPENDIX 4I*). Heat-kill the restriction enzymes or remove them using a DNA purification kit.

    Make sure to use the appropriate buffer conditions. Purification of the DNA may be required after the ligation and prior to the restriction digest.

29. Amplify the ligated product by PCR (*APPENDIX 4J*) using the appropriate joining primers and N- or C-termini primers. For each mini-library, use every possible primer pair, resulting in four reactions for two genes, nine for three genes, etc.

    Start with small amounts of template; it is important that all the PCR reactions work on the same batch of ligated template.

30. Purify the PCR products using preparatory agarose gel electrophoresis and a gel extraction kit.

    Make sure the pieces are appropriately sized. Bands slightly larger than expected may indicate the tags were not completely removed and bands smaller than expected may indicate that one or more of the fragments was not correctly incorporated. It may be helpful to use PCR products from unaltered parent genes as controls to ensure the band sizes are correct.

31. Quantify the gel-purified product using absorption spectroscopy (*APPENDIX 4K*) and mix the reactions in the molar ratio desired in the final library.

    This step is only necessary if the PCR reactions did not have roughly equivalent yields or if there are significant variations between samples during the purification of the PCR products. However, better quantification will only decrease undesirable biases in the final library.

32. Digest the mixed reaction with *Sal*I and *Pst*I, then remove the enzymes using a DNA purification kit (or heat-kill the enzymes). Using T4 ligase (Tabor, 1987b), ligate the insert into the construction plasmid cut with the same enzymes.

33. Transform the product into *E. coli* (*APPENDIX 4D*) and plate the bacteria on 150 × 15–mm plates.

    The number of colonies containing a full-length insert should be significantly (i.e., 2 orders of magnitude) greater than the maximum possible number of combinations in the mini-library (e.g., 81 for 3 genes and 4 blocks). This is to ensure that all combinations are represented in the final library.

    The authors recommend the use of 150 × 15–mm plates for this transformation to make step 34 easier. It is possible to estimate the complexity of the library by plating a small aliquot of the transformation onto a separate plate in order to count the colonies.

34. Pick 10 to 20 colonies for each mini-library, isolate the DNA, and perform an analytical digest with *Sal*I and *Pst*I to determine insert incorporation rate and sequence to confirm correct construction.

Construction phase 4: Construct full-length genes

35. Pool the colonies from each mini-library transformation and isolate the DNA (*APPENDIX 4C*). If there is insufficient cell mass for an effective DNA preparation, then grow the cells for a few hours to increase the cell mass. Quantitate the DNA using absorption spectroscopy (*APPENDIX 4K*).

36. Digest 3 pmol of the N-terminal mini-library with *Sal*I to linearize the plasmid (*APPENDIX 4I*). Digest 3 pmol of the C-terminal mini-library with *Pst*I to linearize the plasmid.

    *This should be the restriction enzyme that is **not** next to SapI in the mini-library.*

37. Dephosphorylate the overhangs with calf alkaline phosphatase (Tabor, 1987a).

   This prevents any cross-ligation during step 41.

38. Purify the linearized plasmid with a DNA purification kit.

39. Digest the plasmid with *Sap*I (*APPENDIX 4I*) to release the library inserts.

40. Purify the inserts using preparative agarose gel electrophoresis (*APPENDIX 4F*) and a gel extraction kit.

41. Ligate the inserts together for 4 hr at 16°C using T4 ligase (Tabor, 1987b). Heat-kill the ligase after the reaction.

   The ideal concentration of 5' ends in the reaction is 0.15 μM. Assuming a 70% yield in the above steps, that is ∼40 μl. Approximately 6 Weiss units of ligase should be sufficient for this reaction (see step 27 for information about Weiss units).

   It may be beneficial to purify the correct-size insert using preparative agarose gel electrophoresis and a gel extraction kit, but this is optional.

42. Digest the ligated product with *Bam*HI and *Hind*III according to the manufacturer's instructions (also see *APPENDIX 4I*). Remove the enzymes using a DNA purification kit or heat-kill the enzymes.

43. Cut the final expression plasmid (e.g., pET28) with the same enzymes, and then ligate the cut inserts into the cut plasmid using T4 ligase (Tabor, 1987b).

44. Transform into *E. coli* (*APPENDIX 4D*) to generate the full library.

LIBRARY CONSTRUCTION USING SYNTHETIC GENES

This is a modification of the Basic Protocol that omits the PCR-based insertion of recombination tags in parental genes. This protocol significantly reduces the time and effort required to construct a chimera library.

Additional Materials (also see Basic Protocol)

   Tag-inserted genes made by a gene synthesis company, (e.g., DNA 2.0, *https://www.dna20.com/*)

1. Perform steps 1 to 9 of the Basic Protocol (design phase 1), in order to design tag sequences for the desired recombination sites.

2. Incorporate tags into parent sequences in silico, and have full-length tag-inserted genes synthesized by a company that provides synthetic genes.

   Most gene synthesis companies charge a flat fee per synthetic reaction. Therefore, it is more cost effective to have full-length genes (with all tags incorporated) synthesized, than to have mini-libraries or tagged fragments made individually.

3. For each gene, design two oligonucleotides that amplify the tag-inserted parent sequence with the appropriate N- and C-terminal restriction enzyme cloning sites, and the appropriate "joining" sites for combining mini-libraries, if applicable.

   The N- and C-termini primers, as well as the "joining" primers (Fig. 26.2.5), are designed exactly as described in the Basic Protocol steps 11 and 12 (design phase 2).

4. Use PCR (*APPENDIX 4J*) to amplify the tag-inserted sequences, and clone into the construction plasmid.

5. Pick several colonies, isolate the DNA, and perform an analytical digest with *Sap*I and *Pst*I (*APPENDIX 4D*) to verify the full-length sequences. Sequence the DNA (Ausubel et al., 2010, Chapter 7) to ensure no mutations were introduced during PCR.

6. Continue the Basic Protocol at step 20 (construction phase 2).

LIBRARY CONSTRUCTION TO INCORPORATE SMALL (<40 bp) BLOCKS

This is a modification of the Basic Protocol to add small blocks (<40 bp) into the library. Blocks of this size cannot be used in the Basic Protocol because they are lost during the purification steps. Incorporating many blocks in this manner is not recommended because they cannot be sequenced to ensure integrity before they are incorporated into the library.

Additional Materials (also see Basic Protocol)

 T4 polynucleotide kinase (Tabor, 1987a)
 Kinase buffer with 1 mM ATP
 Thermal cycler capable of 1°C/sec ramp

1. Perform steps 1 to 26 of the Basic Protocol, performing steps 2 to 4 below in parallel with the Basic Protocol steps, then proceed to step 5 of this protocol.

2. For each gene, design two oligonucleotides that, when annealed, form the fragment with the appropriate 5′ overhangs.

    It is very important that these oligonucleotides be short (<50 bp) and of the highest purity available (PAGE purified). The small fraction of oligonucleotides containing single-base-pair deletions will result in frame shifts in the library, and all mutations that occur in the oligonucleotides will be transferred into the library.

3. Phosphorylate each oligonucleotide using T4 polynucleotide kinase (see Tabor, 1987a) or purchase phosphorylated from the manufacturer.

    The recommended concentration of oligonucleotide in the reaction is ~3 μM, but follow the manufacturer's instructions. Most kinase buffers do not contain ATP, which therefore must be supplemented to 1 mM.

4. Anneal oligonucleotide pairs by combining equimolar amounts, heating the mixture to 100°C and cooling at 1°C/sec to 16°C.

    Alternatively, cool slowly to room temperature by placing tubes on the bench after heating.

5. Add 6 pmol (2 μl) of each annealed oligonucleotide pair to the ligation reaction in step 27 of the Basic Protocol (construction phase 3).

    Add twice as much of the phosphorylated oligonucleotides to the reaction as compared to the cut pieces that would otherwise be used in Basic Protocol 1, because the phosphorylation reaction is inefficient and many oligonucleotides may not be phosphorylated.

LIBRARY CONSTRUCTION TO RECOMBINE MORE THAN 8 BLOCKS

This is a modification of the site-directed recombination Basic Protocol to allow the recombination of more than eight blocks. The authors do not recommend increasing the number of blocks in the mini-libraries to more than four. However, additional mini-libraries can be constructed and added during the final ligation step in construction phase 4. This protocol outlines a few key modifications of the design of additional mini-libraries and the final construction phase.

1. During the design phase, make sure that none of the overhangs generated for the final ligation cross-ligate (similar to ensuring that no overhangs within a mini-library cross-ligate, as in step 6 of the Basic Protocol).

2. During the final construction phase, use the following procedure for mini-libraries not at the N- and C-termini instead of steps 36 to 40. For the N- and C-terminal mini-libraries use steps 36 to 40 from the Basic Protocol.

    a. Digest 3 pmol of the mini-library with *Sap*I according to the manufacturer's instructions (also see *APPENDIX 4I*).

b. Purify the released insert using preparative gel electrophoresis (*APPENDIX 4F*) and a gel extraction kit.

3. Following step 41 (construction phase 4), perform the optional preparative agarose gel purification of the correct-size ligation product.

   This step is not optional if there are more than two mini-libraries.

4. If there is not sufficient ligation product for cloning, PCR-amplify the ligation product essentially as in steps 29 to 31 (construction phase 3) of the Basic Protocol. Use all possible pairs of N- and C-termini primers. Continue the ligation protocol starting at step 42 (construction phase 4).

LIBRARY CONSTRUCTION USING A SUBSET OF GENE FRAGMENTS

It may be desirable to allow only certain genes to occur at some positions in the combinatorial library. This protocol can be easily modified to remove one or more of the gene fragments from the synthesis reaction. However, this must be planned accordingly during the design phase. The essential component of the method is creation of sequence blocks with specific overhangs that allow correct ligation. To remove a block, simply leave it out and create a tag sequence appropriate for the flanking blocks. To do this, design a tag specific to the gene with the two different overhangs required by the flanking blocks rather than the same overhang on both ends. It is important to make sure that all fragments are present in desired molar proportions and that each block has the appropriate overhangs when cut with the Type IIB restriction enzyme.

CHIMERA LIBRARY CONSTRUCTION

This is a simplified version of the chimeragenesis protocol, intended as a reference guide for experienced users. For materials, see the Basic Protocol.

Design phase: Choose recombination sites and design tag sequences

1. Identify parent genes of interest. Choose each unique recombination site with 3 to 5 base pairs that are identical in all parent genes. Ensure that recombination sites are nonpalindromic, and do not cross-ligate.

2. Design tag sequences that contain the unique overhangs of each recombination site, a Type IIB restriction enzyme site such as *Bsa*XI, and a type I restriction enzyme site such as *Nde*I. If constructing mini-libraries, design a "joining" tag sequence (Fig. 26.2.5) using the recognition sites of another Type IIB enzyme, such as *Sap*I, and a type I enzyme such as *Pst*I or *Sal*I.

Construction phase 1: Generate tag-inserted plasmids

3. Insert tag sequences into parent genes, either by PCR-based overlap extension (Basic Protocol) or by synthesizing the parent genes with tags already inserted (Alternate Protocol 1).

   a. If using the Basic Protocol, insert tags with PCR (*APPENDIX 4J*), using primers that will amplify each fragment with overlapping parts of the appropriate tag(s). Gel-purify individual PCR products from these reactions. Assemble purified fragments with overlapping tags using PCR as depicted in Figure 26.2.4, and purify using a DNA purification kit. Repeat until the full-length group of fragments (4 or fewer) is generated.

   b. If using Alternate Protocol 1, design primers to amplify the full-length (4 or fewer tagged blocks) tagged sequences with the appropriate restriction enzymes for cloning.

4. Digest the tagged, purified parent sequences with *Pst*I and *Sal*I restriction endonucleases (*APPENDIX 4I*). Remove the enzymes with a DNA purification kit, or by heat-killing the enzyme.

5. Ligate the product into the construction plasmid using T4 DNA ligase (see Tabor, 1987b) according to the manufacturer's instructions, and transform the ligation reaction into *E. coli* (*APPENDIX 4D*).

6. Pick several colonies and isolate the DNA. Perform an analytical digest with *Sal*I and *Pst*I, and sequence clones (Ausubel et al., 2010, Chapter 7), to confirm no mutations were introduced.

Construction phase 2: Generate library pieces

7. Prepare a large quantity (20 to 100 µg scale or larger; *APPENDIX 4C*) of each tag-inserted plasmid. Quantify DNA concentration using absorption spectroscopy (*APPENDIX 4I*), and mix together 3 pmol of each. Alternatively, PCR amplify the tag-inserted genes. If the tagged genes are amplified by PCR, skip to step 9.

8. Digest with *Pst*I and *Sal*I (*APPENDIX 4I*) to remove insert from the construction plasmid, and dephosphorylate with alkaline phosphatase (Tabor, 1987a). Isolate the dephosphorylated inserts using agarose gel electrophoresis (*APPENDIX 4F*), and purify the inserts using a gel extraction kit.

9. Digest insert with *Bsa*XI (*APPENDIX 4I*) to expose the 5′ overhangs in the fragments. Remove the tags and enzyme with a DNA purification kit.

Construction phase 3: Generate multiple mini-libraries

This construction phase is done separately for each mini library.

10. Ligate the fragments together using T4 DNA ligase (see Tabor, 1987b) according to the manufacturer's instructions, and heat-kill the ligase.

11. Digest the ligated insert with *Bsa*XI and *Nde*I according to the manufacturer's instructions (also see *APPENDIX 4I*), and heat-kill the enzymes or remove them with a DNA purification kit.

12. Amplify the ligated product by PCR (*APPENDIX 4J*) using the appropriate joining primers and N- or C-termini primers. Use every possible primer pair for each mini-library.

13. Purify the PCR products using agarose gel electrophoresis (*APPENDIX 4F*) and a gel extraction kit, and mix the reactions in the molar ratio desired in the final library.

14. Digest the mixed reaction with *Sal*I and *Pst*I, and remove the enzymes with a DNA purification kit, or heat-killing. Using T4 ligase (Tabor, 1987b), ligate the insert into the construction plasmid. Transform the product into *E. coli* (*APPENDIX 4D*) and plate the bacteria on 150 × 15–mm plates.

15. Pick 10 to 20 colonies for each mini-library and isolate the DNA. Perform an analytical digest with *Sal*I and *Pst*I, and sequence clones, to determine insert incorporation and confirm correct construction.

Construction phase 4: Construct full-length genes

16. Pool the colonies from each mini-library transformation and isolate DNA (*APPENDIX 4C*). Quantify the DNA concentration using absorption spectroscopy (*APPENDIX 4I*).

17. Digest 3 pmol of the N-terminal library with *Sal*I (*APPENDIX 4I*), and 3 pmol of the C-terminal library with *Pst*I.

Protein Engineering

26.2.15

18. Dephosphorylate the overhangs with calk alkaline phosphatase (Tabor, 1987a) according to manufacturer's instructions, and purify the plasmids with a DNA purification kit.

19. Digest the plasmids with *Sap*I (*APPENDIX 4I*), and gel-purify the inserts.

20. Ligate the inserts together using T4 DNA ligase (see Tabor, 1987b) according to the manufacturer's instructions, and heat-kill the ligase after the reaction.

21. Digest the ligated product with *Bam*HI and *Hind*II according to the manufacturer's instructions (*APPENDIX 4I*). Remove the enzymes with a DNA purification kit, or by heat-killing.

22. Digest the expression plasmid with the same enzymes, and then ligate the cut inserts into the plasmid. Transform into *E. coli* (*APPENDIX 4D*) to generate the full library.

COMMENTARY

Background Information

Site-directed recombination was developed as an alternative to PCR and truncation-based methods of recombining distantly related genes (Hiraga and Arnold, 2003). PCR-based methods are usually limited to genes with more than 70% sequence identity (Joern et al., 2002), and truncation-based methods lead to a large number of nonviable variants due to insertions, deletions, and frame-shift issues (Ostermeier et al., 1999a,b; Lutz et al., 2001; Sieber et al., 2001). Additionally, with these methods there is little or no control over recombination site number or location within the product genes. Site-directed recombination trades the blind approach used in such methods for a more directed tactic that can take advantage of additional information to choose recombination sites that result in a high proportion of folded chimeras. Additionally, the resulting populations have well-defined sequences that allow a much more detailed analysis of the collection of proteins generated in the experiment (Meyer et al., 2003; Drummond et al., 2005; Li et al., 2007; Heinzelman et al., 2009). Alternate Protocol 4 for recombining only specific gene fragments can also be used to build libraries that contain a subset of the possible parental genes at any given position (Saraf et al., 2005).

In order to generate a population with a large percentage of folded and potentially functional chimeras by recombining distantly related proteins, it is necessary to choose recombination sites effectively. The recombination sites chosen dictate the chimeras that will be constructed. Recombination sites between related proteins have been chosen, with varying degrees of success, to correspond with regions of high sequence identity (Burson and Khosla, 2000), as well as with boundaries of exons (Back and Chappell, 1996), secondary structure elements (Jermutus et al., 2001; Koenderink et al., 2001), and clear structural domains (Nicot et al., 2002; Roman et al., 2003). As the sequence identity between the genes to be recombined decreases, the integration of structural information becomes more and more critical to obtaining a highly folded population. In order to meet this need, several computational methods have been developed that rate chimeras based on sequence and structure information (Voigt et al., 2002; Saraf et al., 2004; Hernandez and LeMaster, 2005). Comparison with existing data on chimeric enzymes has shown that they are at least somewhat effective at predicting which chimeras are more likely to retain function. However, most of these methods have not been thoroughly tested.

The authors of this unit have developed a simple computational model for rating chimeras that takes into account both the protein structure and the sequence of the proteins being recombined to predict which chimeras are more likely to retain the parental fold. The metric used, SCHEMA disruption (E), is calculated by first identifying all amino acid pairs that are contacting in the three-dimensional structure (heavy atoms separated by less than 4.5 Å). A chimera's disruption is then determined by counting the number of contacting pairs where the identities of the amino acids have changed ("broken contacts"). A contact is not broken if the new amino acid pairing is found in any of the parental proteins. Using this method, a chimera can be assessed in silico for likelihood of folding before it is constructed in the laboratory (Silberg et al., 2004). The authors have shown that this

metric is better at determining which chimeras will function than simply counting the number of mutations in the chimeras (Meyer et al., 2003).

The authors have applied SCHEMA disruption to library design (choosing the best recombination sites) by developing an algorithm that minimizes the average disruption for all chimeras in a library at many levels of mutation. This algorithm, RASPP (Recombination as a Shortest Path Problem), uses changes in minimum block length to identify libraries with the least disruption at a range of levels of average mutation (Endelman et al., 2004). The output of RASPP is a curve that contains the libraries with the lowest disruption at various levels of average mutation. Disruption can be transformed into a percentage of chimeras folded, if desired. However, the relationship between disruption and percentage folded must be calibrated using previously collected data (Meyer et al., 2003; Otey et al., 2004). At the Web site listed under Internet Resources, the authors provide MATLAB code to perform RASPP and disruption calculations based on a protein structure and an amino acid sequence alignment of the parental proteins.

Critical Parameters

The most critical parameter in building a site-directed recombination library is primer design. It is very easy to make a mistake during the primer design that will detrimentally affect the construction procedure. Be sure to check all of the DNA created during the procedure for all of the restriction sites used to ensure that there are no extra sites. Be aware that sites may be created during the construction process due to the insertion and deletion of the tag sequences. The best way to ensure that there are no extra sites is to build, in silico, all of the intermediate DNA fragments from the designed tags, gene sequences, and plasmid sequences, and check them individually, stepping through each phase of the procedure with the DNA sequences.

Most of the Basic Protocol involves standard molecular biology techniques strung together. However, it is essential that every step work with high efficiency. The PCR reactions, especially those involving the joining and N- or C-termini primers, should be optimized, and the primers for each gene should be specific. The ligations and restriction digests may also be optimized to ensure a higher success rate. There are several places where plasmids are transformed into *E. coli*. These can serve as places to stop and make sure the procedure

is proceeding as planned by sequencing the DNA. It is also essential that the transformed mini-libraries have many more colonies than the expected complexity of the mini-library (at least 1 to 2 orders of magnitude).

Troubleshooting

Table 26.2.1 lists some of the more common problems that may be encountered using the protocols described in this unit, along with explanations of possible causes of the problems and suggested approaches for overcoming these barriers.

Anticipated Results

The library produced by this method is likely to be biased, with some fragments occurring disproportionately in comparison to the desired molar ratio. This can be minimized through careful quantitation of the plasmids during steps 21 and 31 of the Basic Protocol and through proper mixing. However, it is unlikely that the library will be completely unbiased. The bias can be determined by sequencing a set of unselected chimeras. Sequencing of a large number of chimeras is easily achieved in high-throughput format by DNA hybridization (Meinhold et al., 2003).

For a two-parent library, the average percentage of each fragment is 50% ± 11%. For two separate three-parent libraries, the standard deviations are higher, with the average percentage of each fragment 33% ± 13 and 33% ± 19%. While most positions (>60%) typically show even percentages of each fragment, there are always some positions where the distribution of fragments is such that one fragment is either not present in significant quantity (<10%) or is dominant (>70%). However, usually the mutation frequency is very low (0.007%), and most tags are correctly removed (in 40 randomly chosen chimeras, a single uncleaved tag sequence being identified).

Time Considerations

The amount of time the design phase will take varies depending on the complexity of the library and how the recombination sites are chosen. Construction phase 1 will take 1 to 2 days, and several days of DNA sequencing will then be required to ensure that the tag-inserted plasmids are correct. Construction phases 2 and 3 can be completed in a single, very long day, but it is easier to perform them over 2 or even 3 days, allowing the PCR amplification to proceed overnight. As with most molecular biology reactions,

Table 26.2.1 Troubleshooting Guide for Creating a Combinatorial Gene Library

Problem	Probable cause	Solution
Construction phase 1 (steps 14-19)		
Individual blocks amplify but the assembly PCRs are not working	Not enough overlap between the primers in the tag regions	Lower the annealing temperature. Design new primers with more overlap.
Tags did not insert correctly	Tag overlaps cross-annealed during PCR	Raise the annealing temperature. Design new primers with more distinct regions in the tags.
Construction phase 3 (steps 27-34)		
PCR after ligation yields bands that are too small	One of the blocks is missing because it is too small (<40 bp)	Use Alternate Protocol 2 to add in the piece right before the ligation
PCR after ligation yields bands that are too large	The digests with *Bsa*XI and *Nde*I to remove tags after ligation are not working correctly	Optimize these digests using plasmid DNA
	DNA purification is not removing cut tags correctly	Try a DNA purification kit from a different manufacturer
	Dephosphorylation in step 23 was not complete	Optimize dephosphorylation
PCR after ligation yields multiple bands	Combination of issues listed above. Also check to make sure that *Bsa*XI is cutting to release the fragments	Run analytical agarose gel prior to the ligation to ensure that pieces are cut
One (or more) PCR reaction after ligation gives no product	Not enough template present	Increase amount of template. Check the proportionality of the genes added to the mix.
	PCR conditions are not optimized	Optimize PCR reactions on plasmids with small amounts of template
None of the PCR reactions work after ligation	Template is not full length	Make sure there are no *Bsa*XI or *Nde*I sites within the gene
	There is no template	Run an analytical gel to check concentration of PCR template. There should be something visible in the amount added to the PCR, although it may be a smear.
PCR reactions do not all produce similar yields	Template does not contain equal proportions of the different parental fragments	Check the proportionality of the genes added to the mix. Make sure all genes cut with *Bsa*XI individually. Mix the PCR reactions so that an equal quantity of DNA is used from each reaction.
Pieces not in correct order in mini-libraries/extra pieces found in libraries	Check overhangs for palindromic sequences, or sequences that could possibly cross-anneal	Ligate at a higher temperature (20°-25°C)
Tags still present in mini-libraries	*Bsa*XI and *Nde*I are not removing tags correctly	Optimize the restriction digest using plasmid DNA
	DNA purification not removing tags prior to ligation	Try DNA purification kit from a different manufacturer

continued

Table 26.2.1 Troubleshooting Guide for Creating a Combinatorial Gene Library, *continued*

Problem	Probable cause	Solution
Block missing in mini-library	Block is too small and getting lost in DNA clean-up used to remove tags	Use Alternate Protocol 2
Block from one gene is missing in mini-library	There is a restriction site in the piece	Check to make sure that the block is not cleaved by one of the enzymes used
	One of the tags is not cleaving completely	Check the tag sequences at both ends for the *Bsa*XI site
Construction phase 4 (steps 35-44)		
Mini-libraries not assembling in correct order	Cross-ligation of overhangs	Check overhangs for palindromes
	Incomplete dephosphorylation	Optimize dephosphorylation
Mini-libraries not assembling, or missing a component	Inefficient *SapI* digestion	Optimize the *SapI* digest using plasmid
Not enough insert to ligate into final expression plasmid		Cut more DNA. PCR amplify the full-length product prior to cutting with *Hind*III/*Bam*HI to generate more product.
Alternate Protocol 2: Adding small blocks		
Blocks do not incorporate in mini-library	Phosphorylation inefficient	Optimize phosphorylation reaction. ATP may be missing from buffer
	Oligonucleotides designed incorrectly	Make sure that the oligonucleotides, when annealed, leave the necessary overhangs

these can be stored overnight at –20°C after almost any step in the procedure. Following construction phase 3, it is also recommended that correct sequences be confirmed, which might take a few days. The final construction phase 4 can be done in a single day. An ideal construction with no problems will take ~2 to 3 weeks to complete, depending on the turnaround time for DNA sequencing. However, due to the number of steps that may require optimization, it is more likely to take even an experienced molecular biologist 4 to 8 weeks to complete the first time. This time is not all dedicated to the experiment, but also includes growing time for the *E. coli* necessary to produce DNA.

Literature Cited

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A, and Struhl, K. (eds.). 2010. Current Protocols in Molecular Biology. John Wiley & Sons, Hoboken, New Jersey.

Back, K. and Chappell, J. 1996. Identifying functional domains within terpene cyclases using a domain-swapping strategy. *Proc. Natl. Acad. Sci. U.S.A.* 93:6841-6845.

Burson, K.K. and Khosla, C. 2000. Dissecting the chain length specificity in bacterial aromatic polyketide synthases using chimeric genes. *Tetrahedron* 56:9401-9408.

Drummond, D.A., Silberg, J.J., Meyer, M.M., Wilke, C.O., and Arnold, F.H. 2005. On the conservative nature of intragenic recombination. *Proc. Natl. Acad. Sci. U.S.A.* 102:5280-5385.

Endelman, J.B., Silberg, J.J., Wang, Z.-G., and Arnold, F.H. 2004. Site-directed protein recombination as a shortest-path problem. *Protein Eng. Des. Sel.* 17:589-594.

Hernandez, G. and LeMaster, D.M. 2005. Hybrid native partitioning of interactions among non-conserved residues in chimeric proteins. *Proteins* 60:723-731.

Heinzelman, P., Snow, C.D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F.H. 2009. A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U.S.A.* 106:5610-5615.

Hiraga, K. and Arnold, F.H. 2003. General method for sequence-independent site-directed chimeragenesis. *J. Mol. Biol.* 330:287-296.

Li, Y., Drummond, D.A., Sawayama, A.M., Snow, C.D., Bloom, J.D., and Arnold, F.H. 2007.

Protein Engineering

26.2.19

A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* 25:1051-1056.

Jermutus, L., Terrier, M., Pasamontes, L., van Loon, A.P.G.M., and Lehmann, M. 2001. Structure-based chimeric enzymes as an alternative to directed enzyme evolution: Phytase as a test case. *J. Biotechnol.* 85:15-24.

Joern, J.M., Meinhold, P., and Arnold, F.H. 2002. Analysis of shuffled gene libraries. *J. Mol. Biol.* 316:643-656.

Koenderink, J.B., Swarts, H.G.P., Stronks, C., Hermsen, H.P.H., Willems, P.H.G.M., and de Pont, J.J.H.H.M., 2001. Chimeras of X+, K+-ATPases. *J. Biol. Chem.* 276:111705-11711.

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. Sci. U.S.A.* 98:11248-11253.

Meinhold, P.M., Joern, J.M., and Silberg, J.J. 2003. Analysis of shuffled libraries by oligonucleotide probe hybridization. *In* Methods in Molecular Biology Vol. 231: Directed Evolution Library Creation (F. H. Arnold and G. Georgiou, eds.) pp. 177-188. Humana Press, Totowa, New Jersey.

Meyer, M.M., Silberg, J.J., Voigt, C.A., Endelman, J.B., Mayo, S.L., Wang, Z.-G., and Arnold, F.H. 2003. Library analysis of SCHEMA-guided recombination. *Protein Sci.* 12:1686-1693.

Nicot, C., Relat, J., Woldegiorgis, G., Haro, D., and Marrerro, P.F. 2002. Pig liver carnitine palmitoyltransferase. Chimera studies show that both the N- and C-terminal regions of the enzyme are important for the unusual high malonyl-CoA sensitivity. *J. Biol. Chem.* 277: 10044-10049.

Ostermeier, M., Nixon, A.E., and Benkovic, S.L. 1999a. Incremental truncation as a strategy in the engineering of novel biocatalysts. *Bioorg. Med. Chem.* 7:2139-2144.

Ostermeier, M., Shim, J.H., and Benkovic, S.J. 1999b. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 17:1205-1209.

Otey, C.R., Silberg, J.J., Endelman, J.B., Bandara, G., and Arnold, F.H. 2004. Functional evolution and structural conservation in chimeric cytochromes P450: Calibrating a structure-guided approach. *Chem. Biol.* 11:309-318.

Roman, L.J., McLain, J., and Masters, B.S.S. 2003. Chimeric enzymes of cytochrome P450 oxidoreductase and neuronal nitric-oxide synthase reductase domain reveal structural and functional differences. *J. Biol. Chem.* 278:25700-26707.

Saraf, M.C., Horswill, A.R., Benkovic, S.J., and Maranas, C.D. 2004. FamClash: A method for ranking the activity of engineered enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 101:4142-4147.

Saraf, M.C., Gupta, A., and Maranas, C.D. 2005. Design of combinatorial protein libraries of optimal size. *Proteins* 60:769-777.

Sieber, V., Martinez, C.A., and Arnold, F.H. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* 19:456-460.

Silberg, J.J., Endelman, J.B., and Arnold, F.H. 2004. SCHEMA-guided protein recombination. *In* Methods in Enzymology Vol. 228 Protein Engineering (D.E. Robinson and J.P. Noel, eds.) pp. 35-41. Elsevier Academic Press, Amsterdam.

Tabor, S. 1987a. Phosphatases and kinases. *Curr. Protoc. Mol. Biol.* 0:3.10.1-3.10.5.

Tabor, S. 1987b. DNA ligases. *Curr. Protoc. Mol. Biol.* 8:3.14.1-3.14.4.

Voigt, C.A., Martinez, C., Wang, Z-G., Mayo, S.L., and Arnold, F.H. 2002. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* 9:552-558.

Weiss, B., Jacquemin-Sablon, A., Live, T.R., Faree, G., and Richardson, C.C. 1968. Enzymatic breakage and joining of deoxyribonucleic acid. *J. Biol. Chem.* 243:4543-4555.

Internet Resources

http://www.che.caltech.edu/groups/fha/
Web site that supplies RASPP code for determining optimal recombination points.

Combinatorial Recombination of Gene Fragments

26.2.20