Combinatorial Recombination of Gene Fragments to Construct a Library of Chimeras

Michelle M. Meyer¹ Kaori Hiraga²,³ and Frances H. Arnold¹,²


¹Biochemistry Option
²Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, CA 91125

³current address:
Wadsworth Center
New York State Department of Health
Albany, NY 12201

Frances H. Arnold
Phone: (626) 395-4162
FAX: (626) 568-8753
e-mail: frances@cheme.caltech.edu

Key Terms
chimera, combinatorial library, directed evolution, library creation, protein design, recombination
Mini-Abstract:

Recombination of distantly-related and non-related genes is difficult using traditional PCR-based techniques, and truncation-based methods result in a large proportion of inviable 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 together 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 3-5 base-pairs of identity at the desired recombination sites. Simple adaptations of the method allow incorporation of specific gene fragments.
This protocol contains protocols that allow creation of libraries of chimeras through combinatorial assembly of gene fragments (Figure 26.1.1). 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 is achieved by assembling blocks of sequence with specific overhangs, or sticky ends. 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 (see Figure 26.1.1). The overhangs permit even unrelated fragments to be ligated together 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 limits the size of the blocks and the number of recombination sites. It is also intended to incorporate fragments from each parental gene at all blocks. Alternative protocols have been provided to circumvent these limitations. Choosing the recombination sites effectively, however, is a key part of a successful experiment and is discussed in the Commentary.
Strategic Planning

Block Size

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 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 DNA clean-up steps that remove the tag sequences (~30 bp). However, Alternate Protocol 1 provides options for using blocks smaller than 40 bp.

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 another 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. Alternative Protocol 2 details a protocol that incorporates more than eight blocks for using more recombination sites.
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 substitutes with any restriction enzyme of the researcher’s choice.

Construction of a Combinatorial Gene Library

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 26.1.2. 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 x 2 = 6 plasmids (Figure 26.1.2). 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 \( \frac{1}{2} \) of the full-length gene. 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**

**Solutions and equipment**

- Custom-synthesized oligonucleotides for PCR primers
- Gene sequences for proteins to be recombined
- High Fidelity DNA polymerase for PCR, e.g. *Pfu*
- DNA gel extraction kit (Zymogen Zymoclean Gel DNA Recovery Kit *or* Qiagen: Qiaquick Gel Extraction Kit)
- *PstI* restriction endonuclease
- *SalI* restriction endonuclease
- DNA clean-up kit (Zymogen: DNA Clean and Concentrator *or* Qiagen: Qiaquick PCR Purification Kit)
- Plasmid for construction (e.g. pBC KS+; Stratagene)
Alkaline phosphatase (Tabor, 1987a)

*BsaXI* restriction endonuclease

T4 ligase (Tabor, 1987b)

*Ndel* restriction endonuclease*

150 x 15 mm LB agar plates (Appendix 4A) with antibiotic for transformation

*SapI* restriction endonuclease

*HindIII* restriction endonuclease*

*BamHI* restriction endonuclease*

Plasmid for final library cloning and expression (e.g. pet28)

16 °C water bath

Additional reagents and equipment for the polymerase chain reaction (Appendix 4H), 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., 2006, Chapter 7), quantification 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. *SapI* and *BsaXI* cleave outside of 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 both _BsaX I_ (leaves 3bp 5’ overhang) and _BaeI_ (leaves 5bp 5’ overhang) for construction and found that _BsaX I_ works better. _BsaXI_ is used in this example. Because _BsaXI_ leaves 3bp overhangs, the recombination sites require 3bp 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 26.1.3A. 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. 26.1.3B).

_The DNA between recombination sites is referred to generally as a block._ A fragment 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 last construction phase (Fig. 26.1.3B).*

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. 26.1.3B).*

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

*They must be unique and non-palindromic. For example CTA and TAG will cross-ligate, and AATT will cross-ligate with itself.*

7. Design template tag sequences

*For each mini-library a set of tag sequences will be inserted into the gene at the recombination points. Some portions of these tag sequences are identical for all recombination sites. To design the tags, first make a template tag as shown in Figure 26.1.3C. The BsaXI recognition site is shown in bold, and a second restriction site (Nde I) site in italics. The S, X, Y and N portions of the tag can be any base in our template, but will be filled in 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 we will use BsaX I as our 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 joining site (see step 5), copy the template tag and replace the S portions with the overhang bases as shown (Fig. 26.2.3D).

9. Next, replace the X, Y, and N portions of each tag with bases that are ~50% GC and will not cross-anneal during PCR or generate a restriction site used during the construction process (Fig. 26.1.3E).

**Design phase 2: Design primers**

10. For each gene, design two primers to insert each tag. 

One primer amplifies the DNA fragment before the tag, and the second amplifies the DNA fragment after the tag (Fig. 26.1.4). The 5’ portions of the primers match the tag sequence and overlap ($T_m \geq 50 \, ^\circ C$) to allow PCR-based construction of the gene with tags inserted (Fig. 26.1.4B). The 3’ portion of the primer is unique for each gene.

11. For each gene, design primers for the N- and C-termini.
These 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. 26.1.4D).

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 the Critical Parameters. However, two sets of restriction sites are required for the procedure.

12. Design primers for junction of mini-libraries. See considers below for cloning final library. The last primers are those that 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. 26.1.4C). Choose this restriction enzyme and design the primer to ensure correct cleavage to release overhangs.

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. We have used SapI in with good results and will use it in our demonstration. 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. 26.1.3C). 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.

14. 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

15. 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 Pfu, is recommended for all PCR.*
16. 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 the gel extraction kit.

17. For each gene, connect the PCR products of adjacent fragments using overlap extension PCR. Use the two fragments amplified above as template and the outer 5’ and 3’ primers from the reactions carried out previously (Fig. 26.1.4A).

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

19. Cut the product with PstI and SalI and remove the enzymes using a DNA clean-up 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 manufacturer’s instructions. Transform the product into E. coli (Appendix 4D).

20. Pick several colonies, isolate the DNA, and perform an analytical digest with SalI and PstI to verify full-length tag-inserted sequences (Appendix 4C). Sequence the DNA (Ausubel et al. 2006, 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 3 genes and 2 mini-libraries 6 plasmids are needed).

Construction phase 2: Generate library pieces

This construction phase is done separately with each mini-library (Fig. 26.1.2).

21. Prepare a large quantity (20-100 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 Pfu polymerase. This may result in a few extra mutations in the final library, but can save some time as steps 23 to 25 are no longer necessary.

22. 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 ug of a 3kb 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 is mixed in the molar ratio desired in the library.

Even if the tag-inserted gene is PCR amplified (see step 21), the resulting DNA must be quantified and mixed appropriately.
23. Digest with *PstI* and *SalI* restriction endonucleases to remove the insert from the construction plasmid.  

*This step is unnecessary if the tag-inserted genes were PCR amplified in step 21.*

24. 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 21.*

25. 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 21.*

26. Cut the insert with *BsaXI* (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.
27. Remove the released tags and *BsaXI* using a DNA clean-up kit according to the manufacturer’s instructions.

*It is important that the tags are removed from the reaction. Not all DNA clean-up kits remove small pieces of DNA. Blocks smaller than the cutoff for the DNA clean-up 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 the Alternate Protocol 1.*

**Construction phase 3: Generate multiple mini-libraries**

This construction phase is done separately for each mini-library (Fig. 26.1.2).

28. Ligate the fragments together using T4 ligase (also see Tabor, 1987b) for 4 hrs at 16°C. Heat-kill the ligase after the reaction according the 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 x 3 pmol of each plasmid, (where N is the number of genes you started with) N x 50 ul is ~0.15 μM.*

*There should be N x 6 Weiss units of ligase added to the reaction, proportional to the amount of DNA. One Weiss unit of T4 DNA Ligase converts 1 nmole of $^{32}$P from*
pyrophosphate into Norit-adsorbable material in 20 minutes at 37 °C (Weiss et al., 1968). One Weiss unit equals approximately 67 Cohesive-End Units.

29. Digest the ligated insert with BsaXI and NdeI according to the manufacturer’s instructions (also see Appendix 4I). Heat-kill the restriction enzymes or remove them using a DNA clean-up kit.

Make sure to use the appropriate buffer conditions. A DNA clean-up may be required after the ligation and prior to the restriction digest.

30. 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.

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

Make sure the pieces are appropriately sized. Bands that are even 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.

32. 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.

33. Cut the mixed reaction with *SalI* and *PstI*, remove the enzymes using a DNA clean-up kit (or heat-kill the enzymes). Using T4 ligase (Tabor, 1987b), ligate the insert into the construction plasmid cut with the same enzymes.

34. 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 greater (i.e. 2 orders of magnitude) 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 using 150x15 mm plates for this transformation to make step 36 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.

35. Pick about 10 to 20 colonies for each mini-library, isolate the DNA, and perform an analytical digest with SalI and PstI to determine insert incorporation rate and sequence to confirm correct construction.

Construction phase 4: Construct full-length genes

36. 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).

37. Digest 3 pmol of the N-terminal mini-library with SalI to linearize the plasmid.

    Digest 3 pmol of the C-terminal mini-library with PstI to linearize the plasmid.

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

38. Dephosphorylate the overhangs with calf alkaline phosphatase (Tabor, 1987a).
This prevents any cross-ligation during step 42.

39. Purify the linearized plasmid with a DNA clean-up kit.

40. Cut the plasmid with SapI to release the library inserts.

41. Purify the inserts using preparative agarose gel electrophoresis and a gel extraction kit.

42. Ligate for 4 hrs at 16°C using T4 ligase. Heat-kill the ligase after the reaction.

The ideal concentration of 5’ ends in the reaction is 0.15 uM. Assuming a 70% yield in the above steps, that is ~40ul. Approximately 6 Weiss units of ligase should be sufficient for this reaction (see step 28 for information about Weiss units).

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

43. Digest the ligated product with BamHI and HindIII according to the manufacturer’s instructions. Remove the enzymes using a DNA clean-up kit or heat-kill the enzymes.
44. Cut the final expression plasmid (e.g. pET28) with the same enzymes, then ligate the cut inserts into the cut plasmid using T4 ligase.

45. Transform into *E. coli* to generate the full library.

**Alternate Protocol 1: Library Construction to Incorporate Small (<40) Blocks**

This is a modification of the Basic Protocol to add small blocks (<40 bp) into the library. Blocks of this size cannot be incorporated 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:**

- T4 polynucleotide kinase
- Kinase buffer with 1mM ATP
- Thermocycler capable of 1 °C/sec ramp

1. Perform steps 1 to 27 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 are short (<50 bp) and are 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 (or purchase them 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/s to 16 °C.

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

5. Add 6 pmol (2 μl) of each annealed oligonucleotide pair to the ligation reaction in step 28 of the Basic Protocol (construction phase 3).
Add twice as much of the phosphorylated oligonucleotides to the reaction compared to the cut pieces, because the phosphorylation reaction is inefficient and many oligonucleotides may not be phosphorylated.

Alternate Protocol 2: 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. We 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. We outline here a few key modifications during the design of additional mini-libraries and the final construction phase.

1. During the design phase, make sure that all of the overhangs generated for the final ligation do not cross-ligate (similar to ensuring that all overhangs within a mini-library do not 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 37 to 41. For the N- and C-terminal mini-libraries use steps 37 to 41 from the basic protocol.

   a. Digest 3 pmol of the mini-library with SapI according to the manufacturer’s instructions.
b. Purify the released insert using preparative gel electrophoresis and a gel extraction kit.

3. Following step 42 (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 30 to 32 of the Basic Protocol (construction phase 3). Use all possible pairs of N- and C-terminal primers. Continue the ligation protocol starting at step 43 (construction phase 4).

**Alternate Protocol 3: Construction Using a Subset of Gene Fragments**

It may be desirable to not allow all genes to occur at all 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.

Commentary:

Background Information

Site-directed recombination was developed as an alternative to PCR and truncation-based methods of recombining distantly related genes (Hiraga & 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 inviable variants due to insertions, deletions and frame-shift issues (Lutz et al., 2001; Ostermeier et al., 1999a; Ostermeier et al., 1999b; 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 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 (Drummond et al., 2005; Meyer et al., 2003). Alternative Protocol 3 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 & Khosla, 2000) as well as with boundaries of exons (Back & 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 in silico based on sequence and structure information (Hernandez & LeMaster, 2005; Saraf et al., 2004; Voigt et al., 2002). 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 Angstroms). 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.,
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 by developing an algorithm which 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, we 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.
Table 26.2.1 Troubleshooting Guide for Creating a Combinatorial Gene Library

<table>
<thead>
<tr>
<th>Problem</th>
<th>Probable Cause</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction phase 1 (step 15-20)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual blocks amplify but the assembly PCRs are not working.</td>
<td>Not enough overlap between the primers in the tag regions.</td>
<td>Lower annealing temperature. Design new primers with more overlap.</td>
</tr>
<tr>
<td>Tags did not insert correctly.</td>
<td>Tag overlaps cross-annealed during PCR.</td>
<td>Raise annealing temperature. Design new primers with more distinct regions in the tags.</td>
</tr>
<tr>
<td><strong>Construction phase 3 (steps 28-35)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR after ligation gives bands that are too small.</td>
<td>One of the blocks is missing because it is too small (&lt;40 bp).</td>
<td>Use Alternate Protocol 1 to add in the piece right before the ligation.</td>
</tr>
<tr>
<td>PCR after ligation gives bands that are too large.</td>
<td>The digests with <em>BsaXI</em> and <em>NdeI</em> to remove tags after ligation are not working correctly.</td>
<td>Optimize these digests using plasmid DNA.</td>
</tr>
<tr>
<td></td>
<td>DNA clean-up is not removing cut tags correctly.</td>
<td>Try a different manufacturer.</td>
</tr>
<tr>
<td></td>
<td>Dephosphorylation in step 24 was not complete.</td>
<td>Optimize dephosphorylation.</td>
</tr>
<tr>
<td>PCR after ligation gives multiple bands.</td>
<td>Combination of things listed above. Also check to make sure that <em>BsaXI</em> is cutting to release the fragments.</td>
<td>Run analytical agarose gel prior to the ligation to ensure that pieces are cut.</td>
</tr>
<tr>
<td>One (or more) PCR reaction after ligation gives no product.</td>
<td>Not enough template present.</td>
<td>Increase amount of template. Check the proportionality of the genes added to the mix.</td>
</tr>
<tr>
<td></td>
<td>PCR conditions are not optimized.</td>
<td>Optimize PCR reactions on plasmids with small amounts of template.</td>
</tr>
<tr>
<td>None of the PCR reactions work after ligation.</td>
<td>Template is not full length.</td>
<td>Make sure there are no <em>BsaXI</em> or <em>NdeI</em> sites within the gene.</td>
</tr>
<tr>
<td></td>
<td>There is no template.</td>
<td>Run an analytical gel to check concentration of PCR template. There should be something visible in the amount you add to the PCR, although it may be a smear.</td>
</tr>
<tr>
<td>Issue</td>
<td>Solution</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>PCR reactions do not all give similar yields.</td>
<td>Template is not proportional. Check the proportionality of the genes added to the mix. Make sure all genes cut with BsaXI individually. Mix the PCR reactions in equal proportions and have excess of some.</td>
<td></td>
</tr>
<tr>
<td>Pieces not in correct order in mini-libraries/extra pieces found in libraries.</td>
<td>Check overhangs for palindromic sequences, or sequences that could possibly cross-anneal. Ligate at a higher temperature (20-25°C).</td>
<td></td>
</tr>
<tr>
<td>Tags still present in mini-libraries.</td>
<td>BsaXI and NdeI are not removing tags correctly. Optimize the restriction digest using plasmid DNA. DNA clean-up not removing tags prior to ligation. Try a different manufacturer.</td>
<td></td>
</tr>
<tr>
<td>Block missing in mini-library.</td>
<td>Block is too small and getting lost in DNA clean-up used to remove tags. Use Alternative Protocol 1.</td>
<td></td>
</tr>
<tr>
<td>Block from one gene is missing in mini-library.</td>
<td>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 BsaXI site.</td>
<td></td>
</tr>
<tr>
<td>Mini-libraries not assembling in correct order.</td>
<td>Mini-libraries not assembling, or missing a component. Not enough insert to ligate into final expression plasmid. Cut more DNA. PCR amplify the full-length product prior to cutting with HindIII/BamHI to generate more product.</td>
<td></td>
</tr>
<tr>
<td>Alternate Protocol: Adding small blocks</td>
<td>Blocks do not incorporate in mini-library Phosphorylation inefficient. Optimize phosphorylation reaction. ATP may be missing from buffer.</td>
<td></td>
</tr>
</tbody>
</table>
Oligonucleotides designed incorrectly.

Make sure that the oligonucleotides when annealed leave the necessary overhangs.

**Anticipated results:**

The library produced by this method is likely to be biased, with some fragments occurring disproportionately compared to the desired molar ratio. This can be minimized through careful quantitation of the plasmids during steps 22 and 32 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 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 distributions of fragments is such that one fragment is either no 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 sequences 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 of work and several days afterward to ensure that the tag-inserted plasmids are correct and confirmed by DNA sequencing. 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, 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 may 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 involved that may require optimization, it is more likely to take even the 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 for producing DNA.

**Literature Cited.**


**Internet Resources**

[http://www.che.caltech.edu/groups/fha/](http://www.che.caltech.edu/groups/fha/)

*Web site that supplies RASPP code for determining optimal recombination points.*
Figure 26.1.1

By combinatorially recombining three genes with eight independently assorting blocks $3^8 = 6561$ possible chimeras can be formed in a single experiment. Each positional block of sequence is numbered 1 to 8, and at each block there are three possible fragments labeled 1 to 3. The designation 1.3 refers to the first block with the fragment from parent 3.
Construction phase 1:
Tags are inserted through PCR. Separate plasmids for each gene and mini-library are generated.

Construction phase 2:
Type IIB restriction enzyme digest generates fragments with overhangs, and tags are removed.

Construction phase 3:
Fragments are mixed, ligated, PCR amplified and cloned as mini-libraries.

Construction phase 4:
Mini-libraries are cut from the plasmids, mixed, and ligated together to form full-length genes.

Figure 26.1.2
Overview of the construction process broken down into the four construction 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.
(A) A recombination site must have at least 3 bp which 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) The recombination sites are listed 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 have the same overhang, but they are in different mini-libraries). (C) The template for the tag sequence contains the \textbf{BsaXI} (bold) and \textit{Ndel} (italics) restriction sites. (D) A tag for each recombination site is formed by first adding the overhang bases, as listed in panel B. (E) The remaining bases in the tag are filled in using bases that will not cross-anneal during PCR.
(A) An overview of the PCR protocol necessary to form the tag-inserted sequences. (B) Primers to insert the tag sequences into each gene are used to amplify the fragments before and after the restriction site, and to form part of the tag. They overlap in order to allow PCR-based assembly of the full-length construct with the tags inserted. (C) There are joining primers at the ends of the mini-libraries. These primers contain SapI (bold) restriction sites to generate the overhangs and secondary restriction site (italics) for cloning. Lowercase letters are bases added to the 5’ end of the primer to ensure efficient restriction enzyme cleavage. (D) The N- and C-terminal primers each contain two restriction sites, one for cloning during construction (italics) and a second for cloning into the final expression plasmid (bold).