Preface

Biological systems are very special substrates for engineering—uniquely the products of evolution, they are easily redesigned by similar approaches. A simple algorithm of iterative cycles of diversification and selection, evolution works at all scales, from single molecules to whole ecosystems. In the little more than a decade since the first reported applications of evolutionary design to enzyme engineering, directed evolution has matured to the point where it now represents the centerpiece of industrial biocatalyst development and is being practiced by thousands of academic and industrial scientists in companies and universities around the world. The appeal of directed evolution is easy to understand: it is conceptually straightforward, it can be practiced without any special instrumentation and, most important, it frequently yields useful solutions, many of which are totally unanticipated. Directed evolution has rendered protein engineering readily accessible to a broad audience of scientists and engineers who wish to tailor a myriad of protein properties, including thermal and solvent stability, enzyme selectivity, specific activity, protease susceptibility, allosteric control of protein function, ligand binding, transcriptional activation, and solubility. Furthermore, the range of applications has expanded to the engineering of more complex functions such as those performed by multiple proteins acting in concert (in biosynthetic pathways) or as part of macromolecular complexes and biological networks.

Not surprisingly, the growth in the ranks of practitioners of directed evolution, and also in the range of new applications, has led to a proliferation of experimental methods aimed at simplifying the process and increasing its efficiency. The purpose of this and the accompanying volume in this series is to provide a compendium of experimental protocols accessible to scientists and engineers with minimal background in molecular biology.

Directed Evolution Library Creation focuses on methods for the generation of molecular diversity. Protocols for random mutagenesis of entire genes or segments of genes, for homologous and nonhomologous recombination, and for constructing libraries in vivo in bacteria and yeast are presented. Every one of these methods has been applied for directed evolution purposes. The optimal choice depends on the many factors that characterize each evolution problem, and we have often found that any of several different methods will work. Though there may be multiple molecular solutions to any given functional problem, the library made for directed evolution must nonetheless contain at least one of those solutions. And, the higher the frequency of potential solu-

tions, the easier it is to find them. Thus, the choice of method for creating molecular diversity and its particular implementation are important. In addition to the various protocols for creating libraries, this volume also includes three chapters that describe ways to analyze libraries, particularly those made by recombination.

No directed evolution experiment is successful without a good screen or selection. *Directed Enzyme Evolution: Screening and Selection Methods* is devoted entirely to selection and screening methods that can be applied to directed evolution of enzymes. Directed evolution is not difficult, and these protocols, prepared by practitioners from many leading laboratories, should make this robust protein engineering approach accessible to anyone with a good problem.

Frances H. Arnold

George Georgiou

2

Preparing Libraries in *Escherichia coli*

Alexander V. Tobias

1. Introduction

The process of preparing libraries of mutagenized or recombined gene sequences for screening or selection in *Escherichia coli* is a special application of cohesive-end subcloning ***(1)***. PCR products are digested with restriction endonucleases, ligated into an expression vector digested with the same enzymes, and the resultant recombinant plasmids are transformed into supercompetent bacteria. One difference between routine subcloning of a single gene and preparing libraries is that in the latter case, there can be little allowance for the presence of transformants containing recircularized vector and no insert (so-called "background" ligation products). Furthermore, ligation and transformation of the recombinant plasmids must be performed using materials and conditions that yield a sufficient number of transformants ($\sim 10^3$–10^5) for identifying variants exhibiting desired properties.

Background transformants are a nuisance in routine subcloning, but do not generally ruin the experiment; one can merely pick a few transformants for growth and test them for the presence of the insert by hybridization, PCR, sequencing, or restriction digest. One does not have this luxury when preparing libraries since the number of clones to be screened vastly exceeds the number that can be tested for the presence of insert. Background transformants waste screening effort and decrease the diversity of the transformant library. There can be zero tolerance for background transformants when one is screening for loss-of-function mutants, as these are sure to be confused with desired clones. When screening a library for gain-of-function mutants, a small fraction (<1%) of background transformants may be acceptable, since these will not be confused with positive clones.

From: *Methods in Molecular Biology, vol. 231: Directed Evolution Library Creation: Methods and Protocols*
Edited by: F. H. Arnold and G. Georgiou © Humana Press Inc., Totowa, NJ

Probably the most difficult and frustrating aspect of preparing libraries in *E. coli* is the requirement for a high number of transformants per plate compared to routine subcloning. For example, to screen 50,000 clones, one can grow 50 plates of 1000 clones each or 500 plates of 100 clones each. Obviously, the latter case is much more labor- and resource-intensive, making the former case vastly preferred. The major obstacle to obtaining a sufficient number of transformants per plate lies in the poor transformation efficiency of ligation products—typically one to three orders of magnitude lower than that of supercoiled plasmids. This can represent a challenge to the researcher who wants to screen 10^4–10^5 clones. Nonetheless, with careful preparation of DNA fragments and some optimization of the ligation reaction, one can obtain transformation efficiencies of 10^7 colony-forming units (cfu) per microgram of vector DNA when transforming supercompetent cells. Generally, a transformation efficiency of ~10^6 cfu/μg DNA provides a sufficient number of transformants to screen for most directed evolution applications.

2. Materials

1. Mutagenic or recombination PCR product.
2. Plasmid vector for cloning.
3. Sterile distilled water (dH_2O).
4. DpnI endonuclease (New England Biolabs, Beverly, MA).
5. Restriction endonucleases and buffers.
6. Zymo-5 DNA Clean and Concentrator Kit (Zymo Research, Orange, CA).
7. Zymo-25 DNA Clean and Concentrator Kit (Zymo Research, Orange, CA).
8. Agarose-dissolving buffer (Zymo Research, Orange, CA).
9. 1% agarose gels (analytical and preparative).
10. Gel electrophoresis equipment.
11. UV transilluminator.
12. Sterile razor blades.
13. Shrimp alkaline phosphatase and 10X buffer (USB, Cleveland, OH).
14. T4 DNA ligase and 10X buffer (Roche, Indianapolis, IN) (*see* **Note 1**).
15. Control circular plasmid.
16. Transformation-competent *E. coli* cells.
17. SOC media (*see* **Note 2**).
18. Luria-Bertani agar plates supplemented with appropriate antibiotic.

3. Methods

3.1. Choice of Restriction Sites, PCR Primers, and Vector

Two different restriction sites must be used for subcloning an insert library in order to ensure proper insert orientation in the recombinant plasmids. Avoid restriction endonucleases known by the manufacturer to have “star” activity (ability to cleave at sequences that are similar but not identical to the primary

recognition sequence). If the two restriction endonucleases have compatible buffers and operating temperatures, it may be possible to save time and eliminate purification steps by digesting with both endonucleases simultaneously. This should be verified experimentally, even if the supplier's instructions indicate that the two enzymes are compatible and may be used simultaneously.

Do not design PCR primers with the restriction site at the 5' terminus. Rather, primers should have at least five "spacer" nucleotides 5' of the restriction site. This permits more efficient digestion of the PCR product by restriction endonucleases. The spacer nucleotides should be complementary to the PCR template. The extra annealing this provides may improve the PCR reaction.

If possible, choose a vector with an ampicillin- or kanamycin-resistance gene. Compared to chloramphenicol, these two antibiotics place less metabolic burden on growing colonies. Use of chloramphenicol can result in growth delays and fewer surviving colonies.

Optimizing ligation and transformation efficiencies may involve trying different cloning vectors and restriction endonuclease recognition sites. The researcher is encouraged to attempt the protocols in this chapter with several different vectors and restriction sites in order to increase the chances of obtaining adequate-sized libraries of transformants for screening.

3.2. Preparation of Insert Library

1. Run a 1 µL aliquot of the completed PCR reaction on an agarose gel to estimate the concentration of PCR product.
2. Digest the PCR template DNA by adding 1 µL (20 U) *Dpn*I directly to the entire completed PCR reaction (*see* **Note 3**).
3. Incubate at 37°C for 1 h.
4. Purify 2 µg of PCR product using the Zymo-5 DNA Clean and Concentrator kit. Follow the instructions supplied with the kit. Elute the DNA from the Zymo-5 column with 20 µL dH_2O (*see* **Note 4**).
5. Digest the purified PCR product with two units of each restriction endonuclease (*see* **Note 5**), following the guidelines supplied with the enzymes. The total volume of the insert digestion reaction(s) should be 100 µL.
6. Purify the digested insert using the Zymo-5 Kit. Elute the DNA from the Zymo-5 column with 20 µL dH_2O.
7. Run a 1-µL aliquot of the insert DNA on an agarose gel to estimate its concentration.

3.3. Preparation of Plasmid Vector

1. Purify 2 µg of vector DNA with the Zymo-5 kit (*see* **Note 6**). Elute the DNA from the Zymo-5 column with 80 µL dH_2O.
2. Digest the vector DNA with the two units of each restriction endonuclease (*see* **Note 5**), following the guidelines supplied with the enzymes. The total volume of the vector digestion reaction(s) should be 100 µL.

3. Run the digested vector DNA on a preparative agarose gel.
4. Viewing the gel with an UV transilluminator under preparative illumination intensity, excise the band corresponding to doubly cut vector DNA with a sterile razor blade.
5. Place the excised band in a 2 mL centrifuge tube and fill to the top with Agarose-dissolving buffer. Incubate at 50°C until the agarose is completely dissolved (~10 min).
6. Pass the entire amount of agarose-dissolving buffer containing the vector DNA through a single Zymo-25 column in order to bind all the DNA to the column (this will require several runs on the centrifuge since the volume of a column is only about 0.6 mL).
7. Wash the vector DNA bound to the column as described in the Zymo-25 DNA Clean and Concentrator Kit instructions.
8. Elute the vector DNA with 80 µL dH_2O.
9. Run a 2-µL aliquot of the doubly cut vector DNA on an agarose gel to estimate its concentration.
10. Set aside a 100 ng aliquot of purified, linearized vector DNA for use in control ligation A (*see* **Subheading 3.4.**).
11. Combine the remaining doubly cut vector DNA, 10 µL 10X phosphatase buffer, 1 µL (1 U) phosphatase, and dH_2O to 100 µL. Incubate the phosphatase reaction at 37°C for 1 h (*see* **Note 7**).
12. Purify the phosphatase-treated vector DNA using the Zymo-5 kit. Elute the DNA with 20 µL dH_2O.
13. Run a 1-µL aliquot of the purified phosphatase-treated vector DNA on an agarose gel to estimate its concentration.

3.4. Ligation of Insert and Vector DNA

Obtaining a sufficient number transformants for screening in directed evolution requires obtaining a near-maximal yield of desired recombinant plasmids from the ligation reaction. To achieve this, some optimization of the reaction is usually necessary. One of the most important parameters in a ligation reaction is the ratio of insert to vector molecules. Maximum yield of desired ligation products is usually achieved when this ratio is approximately 2:1 ***(2)***. Nevertheless, running several parallel ligation reactions at different values of this parameter (ligations C-E below) should increase the chances of obtaining near-maximum yield in one of the reactions.

Two different control ligation reactions lacking insert DNA should be performed. Ligation A is carried out in order to estimate the amount of background ligation products present in ligations C-E. Ligation B is performed to evaluate (by comparison to Ligation A) the effectiveness of phosphatase treatment.

1. Set up the following ligation reactions: ligation A: 100 ng phosphatase-treated vector (control for vector recircularization); ligation B: 100 ng untreated vector

(control for phosphatase activity); ligation C: 100 ng phosphatase-treated vector + equimolar amount of insert DNA; ligation D: 100 ng phosphatase-treated vector + 2-fold molar excess of insert DNA; ligation E: 100 ng phosphatase-treated vector + 3-fold molar excess of insert DNA.
2. To each ligation, add 2 µL 10X ligase buffer, 1 µL (1 U) ligase, and dH_2O to 20 µL.
3. Incubate at 16°C for 12 h (*see* **Note 8**).

3.5. Transformation of Ligations

Transformation-ready competent cells are available from suppliers such as Stratagene (La Jolla, CA). Alternatively, Zymo Research supplies an excellent kit for making one's own competent *E. coli* cells.

1. Transform Ligations A-E, into the cells, following the transformation protocol supplied with the competent cells or competent cell kit (*see* **Note 9**). As a control for the efficacy of the antibiotic, transform an equivalent amount of dH_2O to the same amount of cells. As a control for the efficiency of transformation, transform an equivalent amount (in ng) of control plasmid to the same amount of cells.
2. Spread the transformed cells onto the agar plates as instructed in the transformation protocol supplied with the competent cells or kit.
3. Incubate the plates overnight at 37°C.

4. Notes

1. The 10X ligase buffer contains ATP, which is unstable. To prevent degradation by repetitive thawing and freezing, this buffer should be divided into 10–20 µL aliquots for single use and stored at –20°C.
2. SOC media: to make 1 L, add 20 g bacto tryptone, 5 g yeast extract, 0.5 g NaCl, and 10 ml KCl (250 m*M*) to 975 mL distilled water. Autoclave for 30 min on liquid cycle. Immediately before use, add 5 mL $MgCl_2$ solution (2 *M*) and 20 mL glucose solution (1 *M*). Both solutions must be sterilized by filtration. Store refrigerated.
3. *Dpn*I is a restriction endonuclease with a four-base recognition site that digests only methylated DNA. It is used here to destroy plasmid template DNA, so that it does not contaminate the recombinant plasmid library. *Dpn*I treatment is not necessary if the PCR template is not a circular plasmid capable of replicating in *E. coli*, and will not destroy plasmid DNA purified from methylation negative bacteria.
4. Eluting with sterile distilled water warmed to 55°C increases recovery of DNA from Zymo columns.
5. If digesting with one endonuclease at a time, purify the DNA using the Zymo-5 kit after the first digestion, then digest with the other endonuclease.
6. It is preferable to begin with a vector containing an insert at the site where the library is to be ligated, rather than a vector containing only a small spacer sequence between the two cloning restriction sites. In the former case, "empty," doubly cut vector DNA is easily and efficiently isolated by preparative gel electrophoresis. In the latter case, the separation between uncut, singly cut, and dou-

bly cut vector DNA by preparative gel electrophoresis is less efficient. This can result in significant contamination of the recombinant plasmid library by uncut vector DNA.

7. Phosphatase treatment removes the 5'-phosphate groups from the linearized vector DNA, preventing recircularization of vector DNA during the ligation reaction. The insert DNA, which still possesses 5'-phosphate groups, will efficiently ligate with the phosphatase-treated vector, forming a circular DNA molecule with two nicks ***(2)***. Because circular DNA, even if nicked, transforms much more efficiently than linear DNA ***(3)***, nearly all of the transformants should harbor recombinant plasmids.
8. The low temperature enhances the stability of base pairing between the complementary cohesive ends of the insert and vector. This favors the desired intermolecular ligation reaction over vector recircularization. Incubating ligations at 16°C for more than 12 h will deleteriously affect the reactions.
9. Optimal transformation efficiencies are usually attained by mixing 1 µL of each ligation with 50 µL cells. If the protocol supplied with the cells includes an incubation step with SOC media, it is recommended that this step be performed for a maximum of 20 min. This reduces the likelihood of cell doubling, which reduces the occurrence of transformants harboring identical recombinant plasmids.

References

1. Delidow, B. C. (1993) Molecular Cloning of Polymerase Chain Reaction Fragments with Cohesive Ends, in *PCR Protocols - Current Methods and Applications* (White, B. A., ed.), Humana Press, Totowa, NJ, pp. 217–228.
2. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Conley, E. C. and Saunders, J. R. (1984) Recombination-dependent recircularization of linearized pBR322 plasmid DNA following transformation of *Escherichia coli*. *Mol. Gen. Genet.* **194,** 211–218.