METHODS AND APPLICATIONS

White and green screening with circular polymerase extension cloning for easy and reliable cloning

Elizabeth B. Speltz¹ and Lynne Regan^{1,2,3*}

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut

²Department of Chemistry, Yale University, New Haven, Connecticut

³Integrated Graduate Program in Physical and Engineering Biology, New Haven, Connecticut

Received 13 March 2013; Accepted 8 April 2013

DOI: 10.1002/pro.2268

Published online 17 April 2013 proteinscience.org

Abstract: Cloning is an essential prerequisite to test protein design and engineering ideas. However, it is often time consuming, unreliable, and therefore frustrating. Here, we present a streamlined cloning strategy that incorporates a powerful white and green screening protocol to identify colonies with inserts. We use circular polymerase extension cloning, which is both ligation and sequence independent. Furthermore, our entire procedure requires only three quick steps and one enzyme making it easy to use, inexpensive, and tractable. We anticipate that this method will be particularly useful for protein engineers who frequently subclone or make focused deletion, insertion, or substitution libraries.

Keywords: GFP; blue-white screening; circular polymerase extension cloning; Phusion® extension cloning; recombinant colony screening; protein engineering

Introduction

Molecular cloning is a necessary and often frustrating part of protein engineering. For example, ligation dependent cloning procedures require many lengthy steps such as PCR, restriction digests, dephosphorylation, and ligation. Each step requires at least one enzyme, which adds to cost. Moreover, the overall yield decreases at each step making the entire procedure inefficient. Ligation independent cloning procedures, such as circular polymerase

extension cloning (CPEC), requires fewer steps and enzymes making the procedure more cost-effective and efficient.^{1,2} However, CPEC often results in high vector background because it is difficult to completely purify linearized vector from the original plasmid (see Fig. 1). Screening colonies to identify those with inserts can be a time-consuming, costly, and laborious process.

One of the most common procedures to screen colonies for inserts is blue–white screening, which is based on the α -complementation of β -galactosidase to give an active enzyme (LacZ). LacZ hydrolyzes 5-bromo-4-chloro-indolyl- β -D-galactopyranoside (X-gal) to produce blue colonies. When DNA is inserted into an α -complementation vector, β -galactosidase is disrupted, α -complementation is not possible, and the colony appears white in the presence of X-gal.^{3–5}

Grant sponsor: Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences. Grant sponsor: National Science Foundation Graduate Research Fellowship.

*Correspondence to: Lynne Regan, 266 Whitney Avenue, Bass Center Room 323, New Haven, CT 06511. E-mail: lynne.regan@yale.edu

Figure 1. Gene cloning using White And Green Screening in combination with CPEC. In Step 1, the vector and insert are generated in parallel via PCR. Insert primers include additional vector-complementary sequences (shown as dashed black lines) that have similar T_{ms} , in the range of 55°C–60°C.⁹ In Step 2, the insert and vector PCR products are mixed and their regions of overlap (shown as solid black lines) prime one another and Phusion® Polymerase extends the recombinant plasmid. In Step 3, one microliter of the reaction is transformed into electrocompetent *E. coli*, plated on LB agar with appropriate antibiotic, incubated at 37°C overnight, and then screened for fluorescence. Nonrecombinant colonies will express GFP and appear green. Recombinant colonies will not express GFP and appear white. To illustrate, we show close-ups of plates illuminated with UV light (left) or ambient light (right). As examples, 2 nonrecombinant and 2 recombinant colonies are indicated by white and black arrows, respectively.

Although this method has extensive history, X-gal is expensive and colonies left on X-gal plates die. In addition, there are limited α -complementation vectors that may not be compatible with a given downstream application. For instance, to the best of our knowledge there are no α -complementation vectors that add a glutathione *S*-transferase tag to proteins.

A cheaper, faster, and more convenient screening procedure is to use enhanced green fluorescent protein (eGFP), which can be detected in colonies by

its intrinsic fluorescence without the addition of any substrate. A few plasmids that use GFP fluorescence to screen for insertion of DNA have been described.^{6–8} For example, pGreenscriptA, based on pBlueScript SK+, contains a multiple cloning site (MCS) upstream of GFP.⁶ Tang *et al.* developed pGreen-S, based on the pUC18 plasmid, in which eGFP is located between two MCS.⁷ Insertion of DNA in either vector disrupts GFP production resulting in non-fluorescent colonies. However, only

Table I. Summary of Plasmids and their Requirements for GFP Expression

Plasmid	Promoter	Origin	Strain	Inducer?
pZE21	P _{LTETO}	ColE1	DH10 β ^a	No
pZE22	P _{LLACO}	ColE1	DH10 β ^a	No
pMB2	araBAD	pBR322	DH10 β ^a	No
pPROEX HTaM ^b	trc	pBR322	DH10 β ^a	— ^d
pET11a	T7/LacO	pBR322	BL21(DE3) ^c	Yes

DH10 β *E. coli* are F[−] *mcrA* Δ (*mrr*[−]*hsdRMS*[−]*mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *endA1* *araD139* Δ (*ara* *leu*7697 *galU* *galK* *rpsL* *nupG* λ [−]). BL21-Gold (DE3) *E. coli* are B F[−] *ompT* *hsdS*(*r_B*[−] *m_B*[−]) *dem*⁺ Tet^r *gal* λ (DE3) *endA* Hte.

^a Can screen in any standard *E. coli* cloning or expression strain.

^b pPROEX HTaM is a modified pPROEX HTa vector with an in-frame BamHI site.

^c Can screen in any *E. coli* strain expressing T7 Polymerase.

^d Can add inducer (1 mM IPTG) or wait 24–36 hours for GFP expression.

a few plasmids that use GFP to screen for inserts have been described. Furthermore, protein scientists frequently prefer to use their own in-house plasmids for a particular application. Thus, a screening strategy that does not rely on a specific plasmid is more desirable.

We sought to develop an improved cloning protocol that is efficient, inexpensive, and allows rapid identification of correct clones in any destination vector. We envision that our method will be useful for protein engineers who frequently make libraries and different expression constructs. Our screening protocol, white and green screening (WAGS), requires only that a destination vector express a fluorescent protein. We anticipate that most labs will already have a fluorescent protein cloned into their vector of choice. DNA encoding a protein of interest is subcloned in place of the fluorescent protein, resulting in non-fluorescent (white) colonies (Fig. 1). Although WAGS can be used with any cloning protocol, it is particularly efficient with CPEC, which requires only two steps and one enzyme. In optimizing the overall protocol, we modified previously published CPEC methods enabling efficient high throughput cloning. Most importantly, our strategy is reliable. We have successfully used WAGS in combination with CPEC to construct many different recombinant plasmids.

Results

White and green screening (WAGS)

While most previous colony screening procedures relied on specific vectors, we sought to create a screening procedure that can be used with *any* destination vector. To test our method, we chose five plasmids expressing GFP under the control of different

inducible promoters (Table I). We expect that most labs will have GFP or another fluorescent protein already subcloned into their destination vector of choice.

First, it is necessary to determine the strain and induction requirements of each plasmid to produce fluorescent colonies (summarized in Table I). We transformed each plasmid into two different strains of *E. coli*—BL21-Gold(DE3) *E. coli*, which expresses T7 Polymerase or DH10 β *E. coli*, which does not—plated on LB agar containing appropriate antibiotic, and incubated at 37°C overnight. The following morning the plates are immediately screened for fluorescence using a Royal Blue LED flashlight and glasses with a 500 nm LP filter (NIGHTSEA, Bedford, MA). Colonies can also be screened using a UV light box; however, we find that the NIGHTSEA system is more sensitive and does not damage colonies. We expect that most plasmids are leaky and will not require induction for screening.

We used CPEC to clone an insert into each destination vector (summarized in Fig. 1). We chose CPEC because it is a directional, positional, sequence independent, and ligation independent cloning protocol.¹ CPEC also requires only two steps and one enzyme. In parallel, we use PCR to linearize the vector and to generate an insert with an additional 21–25 bp that is complementary to the vector (Fig. 1, Step 1A and 1B, respectively). The vector complementary sequences are designed to achieve a predicted *T_m* of 55–60°C and to minimize secondary structure formation.⁹ We purified each PCR product on an agarose gel and gel extracted using a Qiagen GelExtraction kit. This step minimizes vector only background, but does not eliminate it completely. In Step 2, we mix vector and insert for CPEC. Because the vector and insert have regions of complementarity, they prime one another and Phusion® Polymerase extends the recombinant plasmid. We typically mix 100 ng vector and 15- to 30-fold molar excess insert in PCR mix (see Materials and Methods section). We use the same thermocycler conditions described in Figure 1 for each vector, regardless of size. Following 15 extension cycles, one microliter of the reaction mix is transformed into *E. coli*, plated on LB agar, and incubated at 37°C overnight.

The following morning, the colonies are then screened for fluorescence (Step 3). Colonies containing the original plasmid are green and fluorescent (white arrows in Fig. 1). Colonies containing recombinant plasmid are white and non-fluorescent (black arrows in Fig. 1). Following our typical protocol, we plate about 5–50 μ L of a transformation and obtain ≥ 100 colonies with 0–50% green colonies.

To test the efficiency of our screening procedure, we sequenced eight white and eight green colonies from a typical cloning into pZE21. All green colonies

contained original vector, whereas all white colonies contained recombinant plasmid indicating that WAGS is an indicator of successful cloning. To date, we have cloned numerous inserts (~400–900 bp in size) into each vector (~2–6 kb in size) using CPEC plus WAGS. In all cases, sequenced white colonies always contain insert indicating that our protocol is also reliable.

Although the method we describe here uses GFP to screen for inserts, we also use mCherry interchangeably. Colonies expressing mCherry typically appear pink under ambient light after a 12–30 hour incubation. We expect that our method will work with any protein that fluoresces in *E. coli*. Similar ligation independent cloning strategies, such as Transfer-PCR (as described in¹⁰) or overlap extension PCR cloning (as described in¹¹), may be substituted for CPEC. We also successfully use WAGS in combination with more traditional ligation dependent cloning procedures (data not shown). We expect that our screening protocol will be applicable to almost any cloning strategy.

Spotlight on CPEC cloning: creation of an efficient cloning protocol

We initially noticed that many of our reactions had few recombinant colonies and high vector background. Our goal was to create one efficient high-throughput cloning protocol that gives us the correct clone every time. Therefore, we also used WAGS to optimize our CPEC protocol, that is, to maximize the number of recombinant colonies while minimizing the number of vector-only colonies. For our test system, we used a 2200 bp pZE21 plasmid expressing GFP and a 585 bp insert. We ran 26 CPEC reactions in parallel, each of which used a different number of extension cycles, molar ratio of insert to vector, and amount of DNA. Here, we note significant modifications to a standard CPEC protocol that resulted from these optimizations, summarized in Figure 1.^{1,2}

We found that CPEC is very sensitive to the number of extension cycles. Our maximum efficiency was obtained at 20 cycles as shown in Figure 2(A). This result is similar to previously published studies, which demonstrated that efficiency peaks around 15–18 cycles.^{2,11} We obtained no recombinant colonies when we used only five cycles. Therefore, we recommend a minimum of 10 cycles.

We also found that efficiency strongly depends on the molar ratio of insert to vector. Our efficiency peaked when we used insert in 15-fold molar excess [Fig. 2(B)]. We found that an equimolar ratio of insert to vector, as recommended in several published protocols, is particularly inefficient.^{1,2} We frequently use insert in 15- to 30-fold molar excess over vector.

Finally, we found that efficiency also depends on the DNA concentration. When using 15 extension cycles, efficiency peaked at 100 ng of vector [Fig.

Figure 2. Optimization of CPEC. In all plots the Y axis represents the number of white colonies and the X axis is as indicated. In each plot, we also indicate the percent of white colonies for our most efficient reaction, next to the appropriate data point. At Step 2, we use 100 ng vector and a fivefold molar excess of insert in a total volume of 50 μ L, unless noted otherwise. 1 μ L of each 50 μ L reaction was transformed into *E. coli*, subsequently 1 μ L out of a 500 μ L culture volume was plated. In all plots, 15 extension cycles is represented by circles, 10 cycles is represented by triangles, and 5 cycles is represented by squares. Additionally, in plot A, 20 cycles is represented by an asterisk. A. Cloning efficiency as a function of the number of temperature cycles. B. Cloning efficiency as a function of the molar excess of insert over vector. 100 ng vector was mixed with 1, 5, 15- or 30-fold molar excess of insert as indicated. C. Cloning efficiency as a function of the DNA concentration. 50, 100, or 150 ng vector (equivalent to 1, 2 or 3 ng/ μ L vector, respectively) was mixed with a fivefold molar excess of insert.

2(C)]. Less vector resulted in fewer recombinant plasmids. More than 100 ng vector, also gives fewer recombinant plasmids. To follow up on the reason for this decrease, we analyzed each reaction mix on an agarose gel. Higher molecular weight bands were visible in the reaction using 150 ng vector. Thus, it appears that too much DNA results in unwanted side products and possibly concatenation of

sequences.¹¹ However, the optimal DNA concentration also depends on the number of cycles. If the researcher chooses to use 10 cycles to save time, a compensatory increase in DNA concentration to 150 ng can improve efficiency.

Application of CPEC and WAGS to create focused libraries

We frequently use WAGS and CPEC to make focused libraries including insertions, deletions, or substitutions. In this case, we modify the protocol as shown in Figure 1 to encode the desired mutations either in the insert's forward or reverse primers or in additional internal primers.

Discussion

Here, we provide one complete cloning protocol that enables an investigator to rapidly, efficiently, and inexpensively create and identify recombinant clones. We have successfully used WAGS in combination with CPEC to create numerous recombinant plasmids in all six vectors. Our CPEC protocol is also extremely efficient. Indeed our most efficient reaction resulted in more than 70 white colonies and no green colonies when we plated only 0.2% of a transformation culture. We frequently see efficiencies in this range. WAGS can be used with nearly any cloning protocol; however, it may not be suitable for cloning very toxic proteins as it does result in expression of the target protein. In theory it is also possible to obtain original vector from a white colony if the fluorescent protein does not express well enough to be visualized. In contrast, we have never sequenced original plasmid from a white colony so we believe the probability of obtaining false positives is relatively low. Importantly, our protocol is easy to run in parallel and we commonly clone many constructs at a given time. We believe that our protocol will be particularly useful for protein engineers because it enables the user to successfully clone many constructs simultaneously.

Materials and Methods

Standard molecular biology

For a summary of the plasmids used in this study refer to Table I. pPROEX HTa was purchased from Invitrogen (Grand Island, NY) and modified to include an in-frame BamHI site. pZE21 and pZE22 were used as described in¹². pET11a was purchased from Novagen (Madison, WI). All plasmids contained GFP. Vector was linearized via PCR following standard methods.¹³ Insert was also linearized via PCR. All DNA fragments were analyzed on 0.8% agarose gels using a 1 kb DNA ladder (New England BioLabs, Ipswich, MA) as a size marker. Linearized vector and insert were gel extracted using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). Recombinant plasmids were purified using a QIAprep Spin Miniprep Kit (Qiagen,

Valencia, CA) and sequenced at the W.M. Keck Foundation Biotechnology Resource Laboratory (New Haven, CT). Poor vector amplification can sometimes be solved by doubling the nucleotide concentration, adding 3% DMSO, modifying the annealing temperature or increasing the extension time.

Primer design

All primers used in this study were purchased from W.M. Keck Foundation Biotechnology Resource Laboratory (New Haven, CT). To amplify vector sequences, we design primers about 21–30 bp in length with a predicted T_m of 65°C or greater.⁹ As an example, see Vector Forward and Vector Reverse primers shown below for a pMB2 vector. The forward primer binds after the open reading frame of GFP, whereas the reverse primer binds before GFP's open reading frame.

The insert is amplified using standard PCR primers (18–25 bp) designed to achieve a predicted $T_m \approx 60^\circ\text{C}$ (sequences shown in black below). An additional 18–21 bp complementary to the vector are added to the 5' end of each insert primer (shown as bold and italics text below). These sequences are designed to minimize secondary structure formation and achieve a predicted T_m of 55–60°C.⁹

Vector Forward Primer

5' **TTGGCTGTTTTGGCGGATGAGAGAAGATT**T

Vector Reverse Primer

5' **GACGTCATTCTCTCTGTTAGCCCCAAAAA**

Insert Forward Primer

5' **GCTAACAGGAGGAATGACGTCATGGCTAA**
GCAGGCACTGAAA

Insert Reverse Primer

5' **CTCATCCGCCAAAAACAGCCAATTAGTTGTA**
CAGTTCATCCATGC

CPEC and WAGS

Vector and insert concentration were measured using a Nanodrop 2000c spectrophotometer (Thermo Scientific, Wilmington, DE). Unless noted, all reactions were carried out using 100 ng vector and a 15- to 30-fold molar excess insert in a 50 μL total volume of PCR mix (containing 100 Units Phusion® Polymerase, 0.2 mM dNTPs, 1x Phusion® HF Buffer, and 3% DMSO) (New England BioLabs, Ipswich, MA). The thermocycler settings were as follows: an initial denaturation step at 94°C for 2 minutes, 5–20 temperature cycles of a (denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 6 minutes) and a final extension step at 72°C for 8 minutes. All reactions were carried out in an Eppendorf mastercycler (Hauppauge, NY).

A total of 1 μL of each reaction was transformed into 45 μL electrocompetent DH10 β or BL21-Gold (DE3) cells using a BioRad micropulser (Hercules, CA). The cells were outgrown in 500 μL LB for 1

hour at 30°C. Subsequently, 5–50 μ L were plated on LB agar plates with 100 μ g/mL Ampicillin or 50 μ g/mL Kanamycin. For inducing plates, we also included either 1 mM IPTG or 0.2% Arabinose (Sigma, St Louis, MO). The plates were incubated at 37°C overnight. The following morning they were immediately screened for fluorescence using a 1W Royal Blue LED flashlight and glasses containing a 500 nm LP filter (<http://www.nightsea.com>, NIGHTSEA, Bedford, MA). Alternatively, colonies are imaged using a HighPerformance UV transilluminator (UVP, Upland, CA) and a Kodak DC290 Zoom camera (Rochester, NY) with a 535DF100 camera ring filter.

Acknowledgments

The authors sincerely thank E. White and Dr. K. Inampudi for introducing us to CPEC. They also thank Dr. F. Isaacs for pZE21, pZE22, and pMB2 plasmids and members of the Regan lab for critical reading of this manuscript. E.B.S. is supported by a National Science Foundation Graduate Research Fellowship.

References

1. Quan J, Tian J (2011) Circular polymerase extension cloning for highthroughput cloning of complex and combinatorial DNA libraries. *Nat Protoc*6:242–251.
2. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One*4:e6441.
3. Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*33:103–119.
4. Short JM, Fernandez JM, Sorge JA, Huse WD (1988) λ ZAP: A bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res*16:7583–7600.
5. Slilaty SN, Lebel S (1998) Accurate insertional inactivation of lacZa: construction of pTrueBlue and M13TrueBlue cloning vectors. *Gene*213:83–91.
6. Inouye S, Ogawa H, Yasuda K, Umesono K, Tsuji FI (1997) A bacterial cloning vector using a mutated *Aequorea* green fluorescent protein as an indicator. *Gene*189:159–162.
7. Tang J, Liang S, Zhang J, Gao Z, Zhang S (2009) pGreen-S: A clone vector bearing absence of enhanced green fluorescent protein for screening recombinants. *Anal Biochem*388:173–174.
8. Banjeree S, Apte-Deshpande A, Padmanabhan S (2010) A novel prokaryotic vector for identification and selection of recombinants: Direct use of the vector for expression studies in *E. coli*. *Microb Cell Fact*9:30–38.
9. Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res*35:W43–W46.
10. Erijman A, Dantes A, Bernheim R, Shifman JM, Peleg Y (2011) Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *J Struct Biol*175:171–177.
11. Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques*48:463–465.
12. Lutz, Bujard (1997) Independent and tight regulation of transcriptional units in *Escherichia Coli* via the LacR/O, the TetR/O and AraC/I₁-I₂ regulatory elements. *Nucleic Acids Res*25:1203–1210.
13. Mullis KB, Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol*155:335–350.