Journal of Bioscience and Bioengineering
VOL. 119 No. 3, 251–259, 2015

ELSEVIER

www.elsevier.com/locate/jbiosc

Screening of accurate clones for gene synthesis in yeast

Tohru Yarimizu,[1,§] Mikiko Nakamura,[2] Hisashi Hoshida,[1,*] and Rinji Akada[1]

Department of Applied Molecular Bioscience, Graduate School of Medicine, Yamaguchi University, Tokiwadai 2-16-1, Ube 755-8611, Japan[1] *and Innovation Center, Yamaguchi University, 2-16-1 Tokiwadai, Ube 755-8611, Japan*[2]

Received 23 June 2014; accepted 12 August 2014
Available online 6 September 2014

Methods for error-less gene synthesis are desired because synthesized genes often contain mutations. By cloning PCR-assembled oligonucleotide fragments fused to a selection marker in yeast, we developed a novel method to screen accurate clones in gene synthesis. As a model case, the 555-bp luciferase gene from *Gaussia princeps* (*GLuc*) was synthesized to contain yeast-optimized codons (called y*GLuc* hereafter). After standard PCR-mediated oligonucleotide assembly, many clones showed no luciferase activity. Of these clones, most contained randomly located nucleotide deletions that produced frameshifts and resulted in premature termination. To exclude clones with premature termination, the synthesized y*GLuc* gene was cloned in-frame to fuse with the *URA3* coding sequence, which served as a selection marker in the yeast *Kluyveromyces marxianus*. Ura$^+$ transformation selection was expected to eliminate clones with frameshift errors. The results showed that in-frame marker selection increased the frequency of active y*GLuc* gene to 79%. We used this strategy to synthesize the 1812-bp gene from *Rhizopus oryzae* that encodes glucoamylase. Five out of seven Ura$^+$ clones exhibited amylase activity. Of the functional clones, one contained the correct sequence, and four contained sequences with nucleotide changes, suggesting that in-frame selection frequently produced functional mutants. The *K. marxianus* non-homologous end joining mediated cloning method for gene synthesis will be useful for synthetic biological studies.

© 2014, The Society for Biotechnology, Japan. All rights reserved.

[**Key words:** Gene synthesis; *Kluyveromyces marxianus*; Non-homologous end joining-mediated DNA cloning; Luciferase; Glucoamylase; Codon optimization]

Gene synthesis through the assembly of synthetic oligonucleotides is used in genetic engineering. For cloning genes or cDNA, gene synthesis is a more attractive strategy than PCR amplification from template DNA because organisms are not required for source material. In higher eukaryotes, cDNA can be difficult to prepare if expression is low or restricted to certain developmental stages (1,2). In addition, alternative splicing produces mRNA variants, some of which are not easily obtained when mRNA is prepared from tissue samples or from cDNA libraries (3–5). A similar situation is encountered in microbiology. Metagenome analyses using next-generation sequencers are generating a vast amount of sequence data without cultured microorganisms (6–8). Genes identified from this data cannot be easily obtained by PCR amplification from template DNA samples if the template microorganisms are rare or unculturable. In addition, genes from highly hazardous organisms and viruses, such as Ebola, Lassa, and Variola viruses, cannot be easily prepared because the pathogens must be handled in a high-level biosafety facility. Therefore, gene synthesis is a useful strategy for obtaining DNA of interest without living organisms.

Another useful feature of gene synthesis is codon optimization for gene expression, which enables higher expression of recombinant genes in different host organisms. For example, higher expression of codon-optimized green fluorescent protein (yEGFP) (9), red fluorescent protein (yEmRFP) (10), human growth hormone (11), lipase from *Rhizopus oryzae* (12), glucoamylase from *Aspergillus awamori* (13), and viral proteins (14,15) has been demonstrated in yeast hosts. The DNA sequences of the codon-optimized genes differ from the original sequences and require gene synthesis for preparation.

The whole genome of *Mycoplasma genitalium* has been synthesized (16–19). Since this milestone event, the pace of synthetic biological studies that design and construct biological parts, devices, and systems has accelerated. The goal of the Synthetic Yeast 2.0 project is the chemical synthesis of the whole genome of the yeast *Saccharomyces cerevisiae* (20,21). Therefore, gene synthesis has great potential in life sciences and biotechnological applications.

The basic technological concept of gene synthesis is the assembly of synthetic oligonucleotides. The polymerase chain assembly method is a primary strategy (22–24). Ligase is also used for assembly in a process called ligase chain reaction (LCR) (25). However, these strategies are not reliable because they often generate error fragments that contain incorrect sequences. The dominant source of error is incorrect oligomers (26). Therefore, highly purified oligomers obtained from high-performance liquid chromatography or polyacrylamide gel electrophoresis are often required to decrease the error frequency. As an alternative way to reduce errors, an error correction strategy has been developed, which utilizes mismatch repair mechanisms. Bacterial mismatch-

* Corresponding author. Tel.: +81 836 85 9293; fax: +81 836 85 9201.
E-mail address: hoshida@yamaguchi-u.ac.jp (H. Hoshida).
§ Present address: National Institute for Agro-Environmental Sciences, Environmental Biofunction Division, 3-1-3 Kan-nondai, Tsukuba, Ibaraki 305-8604, Japan.

1389-1723/$ – see front matter © 2014, The Society for Biotechnology, Japan. All rights reserved.
http://dx.doi.org/10.1016/j.jbiosc.2014.08.006

cleaving enzymes have been used to remove errors in chloramphenicol-acetyltransferase gene synthesis (27). In addition, the mismatch detection and binding protein MutS from *Thermus aquatic* has been used for error correction during green fluorescent protein synthesis (28). However, the use of mismatch repair proteins for error correction has been limited, possibly because enzyme conditions suitable for repair are difficult to achieve (26). To overcome these problems, another approach has been developed that detects clones with correctly synthesized fragments. In this approach, green fluorescence protein (GFP) is fused in frame to the synthesized gene (29). Kim et al. (29) successfully used this post-synthesis error detection approach to synthesize six genes in *Escherichia coli*.

In this study, we used a simple yeast cloning system (30) to improve the post-synthesis error detection approach. We cloned PCR-assembled fragments into the yeast *Kluyveromyces marxianus* by fusing them in-frame with the *URA3* coding sequence, placed downstream as a selection marker. Premature termination due to sequence errors resulted in the loss of *URA3* function, which inhibited colony growth on selection plates. Thus, selected transformants were expected to express fragments without premature terminations. We applied this strategy to secretory luciferase and glucoamylase genes and obtained correct clones and active mutant clones at a high frequency. The direct functional selection of synthetic genes in yeast will be useful for genetic engineering and synthetic biological studies.

MATERIALS AND METHODS

Yeast strains, media, and primers The *S. cerevisiae* and *K. marxianus* yeast strains used in this study are listed in Table 1. Cells were grown in YPD medium (1% yeast extract, 2% polypeptone, and 2% glucose) and synthetic drop-out media (0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, 2% glucose, and required nutrients) (30). If necessary, 2% agar was added. Starch plates contained 1% soluble starch in YPD. To check for plasmid loss, 5-fluoroorotic acid (FOA) plates were prepared as described previously (31).

Gene synthesis and cloning in *S. cerevisiae* The luciferase gene from *Gaussia princeps* (G*Luc*: 558 bp with stop codon) (11) and a glucoamylase gene from *R. oryzae* (Ro*GLU1*: 1815 bp with stop codon) (32) were selected as synthesis targets. The DNA sequences of yeast codon-optimized G*Luc* (yG*Luc*) and Ro*GLU1* (yRo*GLU1*) were determined from the amino acid sequences (Supplementary Fig. S1). The yeast codons used for sequence design are shown in Supplementary Table S1. The oligonucleotides used are listed in Supplementary Table S2. The oligonucleotides were dissolved in sterilized water to a concentration of 10 μM. For yG*Luc* synthesis, 80-mer or 40-mer oligonucleotides from both strands were designed and purchased from Fasmac Co., Ltd. (Atsugi, Japan). The same 40-mer oligonucleotides were also purchased from Hokkaido System Science Co., Ltd. (Sapporo, Japan). For assembly, three different oligonucleotide mixtures, named F80/C40, F80/C80, and F40/C40, were prepared (Fig. 1A). F80/C40 was a mixture of 80-mer forward oligonucleotides and 40-mer complementary oligonucleotides containing 20 base overlaps in the forward strand. F80/C80 was a mixture of 80-mer forward and 80-mer complementary oligonucleotides. F40/C40 was a mixture of 40-mer forward and 40-mer complementary oligonucleotides. The concentration of each oligonucleotide was adjusted to 0.5 μM in the mixtures. For yRo*GLU1* synthesis, 50-mer oligonucleotides of both strands without gaps were designed and purchased from Hokkaido System Science. The concentration of each oligonucleotide in the mixture was 0.27 μM.

For oligonucleotide assembly, a PCR mixture was prepared by mixing 1 μl of 10× KOD -Plus- buffer, 1 μl of 2 mM dNTPs, 0.4 μl of

TABLE 1. Yeast strains used in this study.

Strain name	Species[a]	GenGenotype	Reference
BY4743	Sc	*MAT***a**/α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0*	44
RAK3614	Sc	*MAT***a** *ade2Δ0::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0::LEU2*	35
RAK3625	Sc	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::*Sc*TDH3*p$_{Ascl}$Ao*TAA*$_{NotI}$*PGK1*ter-*LEU2*	35
RAK3908	Km	*ura3-1 ade2-1*	36
RAK4296	Sc	*MAT***a** *his3Δ200 leu2Δ0 met15Δ0 trpΔ63 ura3Δ0::*Sc*GAL10*p-yCLuc$_{15C}$*URA3*	35
RAK4314	Sc	*MAT***a** *ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::*Sc*TDH3*p$_{15C}$*URA3*	35
RAK4920	Sc	*MAT***a** *ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::*Sc*TDH3*p-yCLuc$_{15C}$*LEU2*	This study
RAK4960	Sc	*MAT***a** *ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::*Sc*TDH3*p-yCLuc$_{15C}$*URA3*	This study
RAK5125	Sc	*MAT***a** *ade2Δ0A his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::*Sc*TDH3*p-yG*Luc*$_{15C}$*LEU2*	This study
RAK5986	Km	*ura3-1 ade2-1* pKM019 [Sc*ADE2-KmARS7-*Sc*URA3*]	This study
RAK6140	Km	*ura3-1 ade2-1* pKM030 [KmCenD-Sc*ADE2*-KmARS7-Sc*URA3*]	This study
RAK6202	Km	*ura3-1 ade2-1* pKM149 [KmCenD-Sc*ADE2*-KmARS7-Sc*TDH3*p-yCLuc$_{15c}$Sc*URA3*]	This study
RAK9817	Km	*ura3-1 ade2-1* pKM288 [Sc*ADE2-*Sc*TDH3*p-PhEG-Sc*URA3*-KmCenD-KmARS7]	This study
RAK7889	Km	*ura3-1 ade2-1* [KmCenD-Sc*ADE2*-KmARS7-Sc*TDH3*p-yRo*GLU1*$^{\Delta N-1}$ = Sc*URA3*]	This study
RAK7890	Km	*ura3-1 ade2-1* [KmCenD-Sc*ADE2*-KmARS7-Sc*TDH3*p-yRo*GLU1*$^{\Delta N-2}$ = Sc*URA3*]	This study
RAK10276	Km	*ura3-1 ade2-1* pKM407 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1*$^{T138P\ A139L\ T140P}$ = Sc*URA3*][d]	This study
RAK10277	Km	*ura3-1 ade2-1* pKM408 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1*$^{[S84L-D120T]\ L418F}$ = Sc*URA3*][b,d]	This study
RAK10278	Km	*ura3-1 ade2-1* pKM409 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1*^{T136S} = Sc*URA3*][d]	This study
RAK10279	Km	*ura3-1 ade2-1* pKM410 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1*^{L422WI504VN506T} = Sc*URA3*][d]	This study
RAK10280	Km	*ura3-1 ade2-1* pKM411 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1*$^{I101N\ [K105R-T152P]}$ = Sc*URA3*][c,d]	This study
RAK10281	Km	*ura3-1 ade2-1* pKM412 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1* wild type = Sc*URA3*][d]	This study
RAK10282	Km	*ura3-1 ade2-1* pKM413 [KmCenD-Sc*ADE2*-KmARS7- Sc*TDH3*p-Δ42yRo*GLU1*$^{W338\Delta}$ = Sc*URA3*][d]	This study
RAK10470	Km	*ura3-1 ade2-1* pKM429 [Sc*ADE2-*Sc*TDH3*p-yRo*GLU1* wild type-Sc*URA3*-KmCenD-KmARS7]	This study
RAK13613	Km	*ura3-1 ade2-1* pKM935 [Sc*ADE2-*Sc*TDH3*p-yRo*GLU1*^{T138PA139LT140P}-Sc*URA3*-KmCenD-KmARS7]	This study
RAK13615	Km	*ura3-1 ade2-1* pKM936 [ScADE2-Sc*TDH3*p-yRo*GLU1*$^{[S84L-D120T]L418F}$-Sc*URA3*-KmCenD-KmARS7][b]	This study
RAK13617	Km	*ura3-1 ade2-1* pKM937 [Sc*ADE2-*Sc*TDH3*p-yRo*GLU1*^{T136S}-Sc*URA3*-KmCenD-KmARS7]	This study
RAK13619	Km	*ura3-1 ade2-1* pKM938 [Sc*ADE2-*Sc*TDH3*p-yRo*GLU1*^{L422WI504VN506T}-Sc*URA3*-KmCenD-KmARS7]	This study
RAK13621	Km	*ura3-1 ade2-1* pKM939 [Sc*ADE2-*Sc*TDH3*p-yRo*GLU1*$^{I101N\ [K105R-T152P]}$-Sc*URA3*-KmCenD-KmARS7][c]	This study
RAK13623	Km	*ura3-1 ade2-1* pKM940 [Sc*ADE2-*Sc*TDH3*p-yRo*GLU1*$^{W338\Delta}$-Sc*URA3*-KmCenD-KmARS7]	This study

[a] Sc, *Saccharomyces cerevisiae*; Km, *Kluyveromyces marxianus*.
[b] Amino acid sequence from 84 to 120: LLQSLVLTTNTGPSLLLSTVSRNSTSSTKSLVRPTTT.
[c] Amino acid sequence from 105 to 152: RNSTSSTKSLVRPTTTTTTLLTTKSLPLSQPPPPLPLPPPPLHLPLPP.
[d] Equals sign (=) indicates in-frame fusion of coding sequences.

Download English Version:

https://daneshyari.com/en/article/20399

Download Persian Version:

https://daneshyari.com/article/20399

Daneshyari.com