Efficient Library Construction by In Vivo Recombination with a Telomere-Originated Autonomously Replicating Sequence of *Hansenula polymorpha*

So-Young Kim,¹ Jung-Hoon Sohn,¹ Jung-Hoon Bae,¹ Yu-Ryang Pyun,² Michael O. Agaphonov,³
Michael D. Ter-Avanesyan,³ and Eui-Sung Choi^{1*}

Laboratory of Microbial Functions, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Yusong, Taejeon 305-333,¹
and Department of Biotechnology, Yonsei University, Seoul 120-749,² Korea, and Institute of Experimental Cardiology,
Cardiology Research Center, 121552 Moscow, Russia³

Received 27 January 2003/Accepted 8 May 2003

A high frequency of transformation and an equal gene dosage between transformants are generally required for activity-based selection of mutants from a library obtained by directed evolution. An efficient library construction method was developed by using in vivo recombination in *Hansenula polymorpha*. Various linear sets of vectors and insert fragments were transformed and analyzed to optimize the in vivo recombination system. A telomere-originated autonomously replicating sequence (ARS) of *H. polymorpha*, reported as a recombination hot spot, facilitates in vivo recombination between the linear transforming DNA and chromosomes. In vivo recombination of two linear DNA fragments containing the telomeric ARS drastically increases the transforming frequency, up to 10-fold, compared to the frequency of circular plasmids. Direct integration of the one-end-recombined linear fragment into chromosomes produced transformants with single-copy gene integration, resulting in the same expression level for the reporter protein between transformants. This newly developed in vivo recombination system of *H. polymorpha* provides a suitable library for activity-based selection of mutants after directed evolution.

Directed evolution is one of the most effective methods currently available for protein engineering to achieve specific, desirable properties. Unlike protein engineering by rational design, directed evolution does not require a priori knowledge of protein structure or the structure-function relationship. However, to achieve optimum directed evolution, there are a few requirements, including minimal protein expression bias, a sufficient number of clones in the library, and quantitative screening criteria (29). Recently, in vivo recombination for easier library construction and an alternative tool for family shuffling in *Saccharomyces cerevisiae* was suggested (1). When the linear fragments that possess overlapping DNA sequences are cotransformed into *S. cerevisiae*, the fragments undergo homologous recombination, which restores the circular topology of the plasmid (16, 19, 22). Recent reports have indicated that plasmid repair is an extremely efficient process and that recombination between overlapping sequences as short as 30 bp can be readily achieved (17, 21, 24). The advantage of in vivo recombination is twofold: efficient cloning of the resulting hybrid sequences in a large yeast expression vector and constitution of a second round of DNA shuffling not using PCR-based techniques (1, 20, 23).

The methylotrophic yeast *Hansenula polymorpha* has recently been studied as an efficient host for production of foreign proteins (6, 10, 30). It has several distinctive features as an expression host, including availability of strong promoters from genes involved in methanol metabolism, stable maintenance of

multiple copies of foreign genes in the chromosomes, and ease of growth to a high cell density of 100 to 130 g/liter (7). *H. polymorpha* is also a useful biocatalyst for metabolic pathway engineering in which expression of multiple genes in different ratios is necessary (8). Recently, a surface display system in *H. polymorpha* using novel cell wall proteins was developed (15). Surface display plays an important role in linking the genotype and phenotype in directed evolution of useful proteins. It makes possible the screening of a library by using flow cytometry for proteins that have a high affinity for a substrate or different substrate specificities (29).

The recombinant gene expression system in *H. polymorpha* DL1 uses an autonomously replicating sequence (ARS) named HARS36 (28). HARS36 is a family of telomeric ARSs residing in several ends of chromosomes in *H. polymorpha* containing the ARS domain and a telomeric repeat. The telomeric repeat of HARS36 is an 8-bp G-rich sequence (5'-GGGTGGCG-3'), and there are 18 repetitions, up a total of 144 bp. Telomeric repeats serve as a recombination hot spot (4, 14, 26). Most integration events of a transforming plasmid containing HARS36 occur near the different ends of chromosomes (26). The combination of the ARS domain and the telomeric repeats of HARS36 greatly increases the potential of HARS36 for multiple gene integration into the chromosome. The expression level of foreign proteins, however, is diverse due to a difference in the integrated gene copy number (3, 27). Thus, the integrants are not adequate for activity-based selection of a mutant library. We present in vivo recombination techniques in *H. polymorpha* for construction of a library with high efficiency and an equal gene expression level in this report. Optimizing the HARS36 as an overlapping sequence for in vivo recombination results in a transformation efficiency that is 10

* Corresponding author. Mailing address: Laboratory of Microbial Functions, Korea Research Institute of Bioscience and Biotechnology, 52 Oun-dong, Yusong, Taejeon 305-333, Korea. Phone: 82-42-860-4453. Fax: 82-42-860-4594. E-mail: choi4162@kribb.re.kr.

A. Circular

B. Acceptor(SmaI/BamHI) + Insert(XhoI/ClaI)

C. Acceptor(SmaI/PstI) + Insert(XhoI/SphI)

FIG. 1. Expression vectors used for in vivo recombination. (A) Circular pGK-CALB-CwpF; (B) *SmaI/BamHI* fragment and *XhoI/ClaI* fragment of pGK-CALB-CwpF; (C) *SmaI/PstI* fragment and *XhoI/SphI* fragment of pGK-CALB-CwpF. Abbreviations: *pGPD*, GAPDH gene promoter; KT, killer toxin signal sequence; tUK, terminator (13).

times greater than that obtained by using the circular plasmid and construction of a library with no expression bias. Both high transformation efficiency and the library with no expression bias are practical for directed evolution.

MATERIALS AND METHODS

Strains and media. *H. polymorpha* used in this study was DL1-L, a *LEU2*-defective strain of DL1 (ATCC 26012). Cells were routinely grown in YPD medium (1% yeast extract, 2% peptone, and 2% glucose) with shaking at 37°C. All transformants were selected on synthetic defined minimal medium (0.67% Bacto yeast nitrogen base without amino acids and 2% glucose).

Plasmid construction. General DNA manipulation was performed as described by Sambrook et al. (25). Total yeast DNA was isolated according to the method of Holm et al. (11). DNA sequencing was carried out with an automatic DNA sequencer (ABI model 373A; Applied Biosystems). The plasmid pGA-GOD-CwpF (15) contained the GAPDH (glyceraldehyde-3-phosphate dehydrogenase) gene promoter (27), HARS36 for autonomous replication (28), the *LEU2* gene of *H. polymorpha* (2), and CwpF as a surface display anchor from *H. polymorpha* (15). This plasmid was used as a backbone for construction of a *Candida antarctica* lipase B (CALB) expression vector. The CALB gene was obtained from the chromosome of *C. antarctica* by PCR amplification with the primers CalBN (5'-GGCTCTTCAGCCACTCCTTTGGTGAAG-3') and CalBF (5'-GCGGATCCGGGGGTGACGATGCCGGAG-3'). The PCR-amplified CALB gene fragment was treated with *SapI*-Klenow and *BamHI* and subcloned into the *EcoRI/BamHI* site of pGA-GOD-CwpF with the PCR-amplified *Kluyveromyces fragilis* killer toxin signal sequence to construct the surface display vector pGK-CALB-CwpF (Fig. 1A). The vector pGK-CALB-CwpF without HARS36 was constructed by subcloning the 3.3-kb *XbaI/ClaI* fragment containing the CALB expression cassette and *LEU2* into the *XbaI/ClaI* site of pBlue-script SK(+).

Yeast transformation. *H. polymorpha* was transformed by the lithium acetate method described previously (28). For circular plasmid transformation, 100 ng of the plasmid was used, and for in vivo recombination, 100 ng of linear acceptor vector DNA fragment (acceptor) and insert DNA fragment (insert) was used. When pGK-CALB-CwpF was used as an acceptor, 100 ng of a 5-kb *EcoRI/PstI* fragment of the plasmid was eluted from the gel and transformed with 300 ng of PCR-amplified insert. The insert was PCR amplified by using Premix *Taq* polymerase (Bioneer, Taejon, Korea) at 94°C for 3 min; 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 2 min; and then 72°C for 7 min for extension.

CALB activity assay. To test for CALB activity on plates, transformants were picked on YPD plates containing 1% tributyrin and incubated for 24 h at 37°C. Halo formation was then observed.

Random mutagenesis. The conditions for PCR random mutagenesis were optimized to produce two to five base substitutions per CALB gene by using a PCR random mutagenesis kit (Clontech, Palo Alto, Calif.). pGK-CALB-CwpF containing both the CALB and the CwpF genes was used as a template for the first round of PCR random mutagenesis. The forward primer was GPD-err (5'-GCAGAGCTAACCAATAAGG-3'), and the reverse primer was H150 (5'-TGCAGTTGAACACAACAC-3'). PCR was performed at 94°C for 30 s; 25 cycles of 94°C for 30 s and 68°C for 1 min; and then 68°C for 10 min. The amplified DNA fragment was purified by agarose gel elution and used as a template for a second PCR, which was performed with Premix *Taq* polymerase (Bioneer) for DNA amplification.

Southern hybridization. The integration patterns of the transforming DNA were analyzed by Southern hybridization techniques (25). Total chromosomal DNA was isolated and digested with restriction endonucleases. After electrophoresis, the DNA was transferred onto a nylon membrane (Schleicher & Schuell). A CALB gene probe was obtained from a PCR using CalBN and CalBF primers with a digoxigenin-labeling deoxynucleoside triphosphate mixture (Roche). Hybridization was performed at 42°C in a hybridization oven (Hybaid) with a hybridization solution (5× SSC [1× SSC is 0.15 M NaCl with 0.015 M sodium citrate], 0.1% [wt/vol] *N*-lauroylsarcosine, 0.02% [wt/vol] sodium dodecyl sulfate, 5% [wt/vol] blocking reagent, 50% [vol/vol] formamide), as recommended by the manufacturer.

RESULTS

Transformation of *H. polymorpha* through in vivo recombination. A single linear DNA fragment or combinations of overlapped DNA fragments from pGK-CALB-CwpF (Fig. 1A) were tested for transformation efficiency and CALB expression to develop an in vivo recombination system in *H. polymorpha*. HARS36, found in several chromosomal ends of *H. polymorpha*, has been identified as a recombination hot spot between the chromosome and a transforming plasmid containing HARS36 (26). As shown in Table 1, the transformation efficiency of the circular plasmid pGK-CALB-CwpF was much higher than the transformation efficiency of the same plasmid without HARS36. This enhancement of transformation frequency was also found in linearized fragments containing HARS36. Two linearized DNA fragments generated by the unique cutting enzymes *XhoI* in *LEU2* and *SphI* near the end of HARS36 in pGK-CALB-CwpF showed higher transformation efficiencies than the same linear fragment without HARS36. These linearized fragments were integrated into the end of the chromosome with high frequency. More than 95% of the transformants showed CALB activity, even in the cells transformed with a DNA cut in the middle of *LEU2*. The linearized DNA was apparently integrated into the chromosome via homologous recombination, probably after recirculation. A linear fragment generated by the unique cutting enzyme *SphI* near HARS36 showed a higher transformation efficiency than the circular plasmid.

Cotransformation of the two linearized DNA fragments, a *SmaI/BamHI* fragment as an acceptor vector DNA fragment (acceptor) and a *XhoI/ClaI* fragment as an insert DNA frag-

TABLE 1. Transformation efficiency and multiple integration ratios of various DNAs from pGK-CALB-CwpF

Plasmid	Enzyme	DNA status	No. of transformants/ μg of DNA	% Multiple integration ^a
pGK-CALB-CwpF without HARS36	None	Circular vector	10	ND
	<i>XhoI</i>	Single linearized fragment	50	ND
pGK-CALB-CwpF	None	Circular vector (Fig. 1A)	2,920	>50
	<i>XhoI</i>	Single linearized fragment	1,800	ND
	<i>SphI</i>	Single linearized fragment	3,800	ND
	<i>SmaI/BamHI</i> + <i>XhoI/ClaI</i>	Two linearized fragments (Fig. 1B)	2,480	ND
	<i>SmaI/PstI</i> + <i>XhoI/SphI</i>	Two linearized fragments (Fig. 1C)	21,040	<15

^a The ratio of multiple integration was determined by genomic Southern blotting of 40 randomly selected transformants. ND, not determined.

ment (Fig. 1B) resulted in a transformation efficiency similar to that of the single linearized DNA fragment (Table 1). More than 95% of the transformants showed CALB activity. These two fragments contained the 5'-overlapping regions in *LEU2* and the 3'-overlapping regions in the CALB and CwpF genes. Transformation of the acceptor only did not show any transformant in the leucine-lacking plate. The results indicated that two linear fragments efficiently recombined through in vivo recombination. Cotransformation of the *SmaI/PstI* fragment as an acceptor and the *XhoI/SphI* fragment as an insert, which both have the 5'-overlapping sequence in *LEU2* and the 3'-overlapping sequence in HARS36 (Fig. 1C), showed an eightfold-higher transformation efficiency than the *XhoI/ClaI* fragment harboring the 3'-overlapping sequence in the CALB and CwpF genes. All transformants from the *SmaI/PstI* and the *XhoI/SphI* fragments of pGK-CALB-CwpF exhibited CALB activity and the typical genomic Southern pattern of telomeric integration (as analyzed by the previously described method of Sohn et al. [26]; data not shown). Therefore, it is probable that the exposed HARS36 at the end of the acceptor and the insert enhances both reconstitution of the circular plasmid and integration into the chromosomal end. Transformation with two linearized fragments through in vivo recombination also reduced the occurrence of multiple integration events of transforming DNA observed with a circular vector (from 50 to 15%) (Table 1). It will be helpful to avoid the false selection of mutants with high activity due to the high gene dosage.

Optimization of in vivo recombination in *H. polymorpha*. In our vector system of *H. polymorpha*, *LEU2* and HARS36 were found to be effective as the 5'- and 3'-overlapping sequences

for in vivo recombination, respectively. The size and component of the insert, however, are both important, especially when the insert must be generated by error-prone PCR techniques for directed evolution. The overlapping sequence was optimized to reduce the size of the insert and to avoid the unwanted mutations in genes other than the target. Since *LEU2* was far from the CALB target gene, another 5'-overlapping sequence was tested in the GAPDH promoter near the initiation codon of the CALB open reading frame. When the H625 fragment that was amplified by PCR with the primer pair GPD-err and H625 was transformed with 100 ng of the *EcoRI/PstI* fragment from pGK-CALB-CwpF as an acceptor, 3.4×10^4 transformants were obtained (Fig. 2). Transformation with the acceptor resulted in less than 5% of the number of transformants obtained with the acceptor and the H625 insert. Hence, most transformants appearing on leucine-lacking plates came from reconstitution of the plasmid between the acceptor and the insert via in vivo recombination.

For size reduction of the insert near the 3'-overlapping sequence, the minimum required HARS36 for in vivo recombination was determined. Three important domains in HARS36 were reported to be (i) a bent DNA domain, (ii) an AT-rich ARS core domain, and (iii) a telomeric repeat domain (28). HARS36 was gradually deleted from the insert (Fig. 3) by PCR using the primers shown in Table 2. H625 contains the full length of HARS36. The telomeric repeat domain was deleted in H454. All three domains of HARS36 were deleted in H150. The C domain left in H150 is also a part of HARS36 but is known to have no function after deletion for autonomous replication (28). Equal amounts of each PCR fragment, H625,

FIG. 2. In vivo recombination using the GAPDH gene promoter as a 5'-overlapping sequence and HARS36 as a 3'-overlapping sequence. The *EcoRI/PstI* fragment of pGK-CALB-CwpF was used as an acceptor, and the PCR fragment H625 amplified by the primer pair GPD-err and H625 was used as an insert.

FIG. 3. Transformation efficiency of different insert fragments harboring various lengths of 3'-overlapping sequence in HARS36. C, flanking domain of ARS; BD, bent DNA domain; ARS, AT-rich ARS core domain; Rep, telomeric repeat domain.

H454, and H150, were cotransformed with the acceptor (Fig. 3 and Table 3). The transformation efficiencies of the inserts were then compared. Surprisingly, the efficiencies were similar even in H150, which does not contain the three important domains of HARS36. Therefore, all three domains of HARS36 in the insert are not necessary for a high frequency of in vivo recombination. Reassembly of two fragments for circularization was impossible for H150 since H150 and the acceptor do not share the 3'-overlapping sequence. However, more than 95% of the transformants exhibited CALB activity, indicating correct in vivo recombination between the two fragments with just one overlapping sequence in the 5' end of H150. Accordingly, it is conceivable that the insert and acceptor are first recombined near the left end of the insert and the resulting linear long piece of DNA is directly integrated into the chromosome without recircularization (Fig. 4A).

We tested five further shortened inserts (H98, H39, T2, T1, and L) (Fig. 3) to determine the minimum sequence of HARS36 required for in vivo recombination. As shown in Table 3, although H98 still showed an efficiency comparable to that of H150 and H454, H39 exhibited an efficiency four times lower than that of H98. The transformation efficiencies of inserts T2, T1, and L, which contain further deletions, were sharply decreased to the value obtained with the acceptor only.

Complete deletion of HARS36 in the acceptor also resulted in a basal level of transformation efficiency, even when H625 was used for transformation. It suggests that a part of HARS36 sequence is necessary in both strands for high-frequency transformation through in vivo recombination. The large difference in transformation efficiency between H98 and H39 indicated that the minimum length of region C of HARS36 is around 100 bp for the efficient integration of one-end-recombined DNA into the chromosome.

Mechanism of in vivo recombination in *H. polymorpha*. The mechanism of integration into the chromosome after in vivo recombination was studied with the transformants from the acceptor (*EcoRI/PstI* fragment of pGK-CALB-CwpF) and H150. Twelve randomly selected transformants exhibiting CALB activity were analyzed by Southern blotting (Fig. 4B) using the digoxigenin-labeled CALB gene as a probe. Total genomic DNA was isolated and digested with *EcoRI* and *SacII*, respectively. All 12 transformants showed a 2-kb single band (Fig. 4B) in the case of *EcoRI*, indicating a typical pattern of single-copy integration into the same locus. When genomic DNA digested with *SacII* that cuts once in the CALB gene was used, the 1.2-kb common band found in all tested transformants and a variable-sized band due to *SacII* polymorphism of different chromosomes were identified (Fig. 4B). We previ-

TABLE 2. Primers used for the construction of insert fragments

Name	Sequence	Priming site ^a
GPD-err	5'-GCAGAGCTAACCAATAAGG-3'	
H625	5'-GTTGGCAATGTTGTACCCGATGATC-3'	601-625
H454	5'-TCAGACCGCATCCTAAAACC-3'	446-454
H150	5'-TGCAGTTGAACACAACCAC-3'	132-150
H98	5'-GCCTTCGTACACCTTGCTG-3'	80-98
H39	5'-CGTCTCCACAGCACCAGG-3'	22-39
T2	5'-GAGGTCGATCTAAGGAGTC-3'	
T1	5'-AATCAGAGGCGGTGTGTGC-3'	
L	5'-GCGGATCCGGGGGTGACGATGCCGGAG-3'	

^a The nucleotide position from the first nucleotide of HARS36.

TABLE 3. Transformation efficiency of different insert fragments harboring various lengths of 3'-overlapping sequence in HARS36

Insert fragment	No. of transformants/ μg of DNA
H625	33,670
H454	30,500
H150	31,920
H98	30,440
H39	6,820
T2	1,360
T1	1,730
L	1,400
Circular vector	1,989
Acceptor vector only	1,421

ously reported that five chromosomal ends among 12 ends from six chromosomes of *H. polymorpha* consisted of a sequence that was highly homologous with HARS36 (26). The 1.2-kb bands found in all transformants were presumably new chromosome ends produced after the integration of transforming DNA. A variable number of telomeric repeats generates a small size variation in the end fragments between transformants. The results indicate a mechanism of in vivo recombination between the chromosome and two transforming DNA fragments (Fig. 4A). Such high frequency of transformation through in vivo recombination can be caused by the multiple targets existing in the different chromosomes of *H. polymorpha*.

FIG. 4. Chromosomal integration of the acceptor and the insert fragments into the end of chromosome. (A) Schematic diagram of integration into the end of a chromosome. (B) Southern analysis of 12 transformants randomly selected from the library that was constructed by in vivo recombination. Chromosome DNA was digested with *EcoRI* or *SacII* and hybridized by using the *CALB* gene as a probe. C, DL1 wild-type strain; V, transformant with the acceptor only. "Transformants" refers to transformants with the acceptor and the insert. TE, 2-kb single band obtained with *EcoRI*; T, 1.2-kb common band; F, variable-sized band due to *SacII* polymorphism of different chromosomes.

FIG. 5. Comparison of CALB activities of randomly selected transformants with different vectors. (A) Transformants with the circular plasmid pGK-CALB-CwpF; (B) transformants obtained by in vivo recombination with the acceptor vector and the insert amplified by normal PCR; (C) transformants obtained by in vivo recombination with the acceptor and the insert amplified by error-prone PCR.

Library construction using in vivo recombination. An equal gene dosage between transformants is preferable for activity-based selection of mutants from a library obtained by directed evolution. Transformation using such one-end-overlap fragments, the *EcoRI/PstI* fragment of pGK-CALB-CwpF and H150, resulted in single-copy integration into chromosome of all tested transformants (Fig. 4B). As mentioned in the previous section, in vivo recombination could reduce the frequency of multiple gene integration from 50 to 15% of all transformants (Table 1). In vivo recombination with one-end-overlap fragments could further reduce the frequency of multiple integrations. To check the activity variation between transformants, randomly selected transformants from two groups transformed with the circular plasmid pGK-CALB-CwpF and two optimized fragments of an acceptor and H150 were compared for CALB activity on YPD-tributyryn plates (Fig. 5A and B). Transformation of the circular plasmid resulted in a considerable activity variation between transformants due to different integration copy numbers (data not shown). On the other hand, cotransformation of the acceptor and H150 through in vivo recombination produced even activity halos. Therefore, in vivo recombination greatly reduces the possibility of selecting a false-positive clone from a mutant library after directed evolution.

For a practical test of the system, a mutant CALB gene library was constructed by using error-prone PCR techniques under conditions for the introduction of two to five base substitutions on the CALB gene using the primers GPD-err and H150. Cotransformation of the PCR product with the acceptor resulted in approximately 3×10^4 transformants per μg of acceptor. The transformants were randomly selected and picked on YPD-tributyryn plates. As shown in Fig. 5C, the CALB activities of the transformants were quite diverse. Ten colonies from the mutant pool were randomly selected and subjected to DNA sequencing to test the efficiency of library construction. Total DNA was isolated from each transformant, and the mutant CALB gene integrated into the chromosome was recovered by PCR with the primers GPD-err and H150. From the sequencing of a total of 4.7 kb of the sense strand from 10 transformants, 34 base substitutions were identified

(Table 4). Three clones carried three substitutions, and the remaining clones contained 0, 1, 2, 4, 5, 6, and 7 substitutions in the 470 bp of the CALB gene. The data presented above showed that each transformant contained a different insert fragment and the full-length expression cassette.

DISCUSSION

Recently, in vivo recombination techniques have been used as an alternative tool for subcloning and library construction in *S. cerevisiae*. Library construction using in vivo recombination in yeast eliminates the gene manipulation steps in *Escherichia coli* and the possibility of expression bias for eukaryote-originated proteins in *E. coli*. Family shuffling can also be anticipated with in vivo recombination (1). To construct a mutant library using directed evolution in *H. polymorpha*, we developed an in vivo recombination system using various combinations of DNA fragments from the CALB expression vector pGK-CALB-CwpF in *H. polymorpha*.

Among the various DNA fragments tested, the acceptor and the insert both containing HARS36 at the end of the fragments exhibited dramatically higher transformation efficiencies than the circular plasmid. Transforming DNA was always found in the chromosome instead of a plasmid. The fate of two overlapping DNA fragments was different in *S. cerevisiae* in which the fragments formed an episomal plasmid (17). Recent reports have indicated that the GC content of the targeting

TABLE 4. DNA mutations identified in the sense strand of mutants of the CALB gene after PCR random mutagenesis (from sequencing of 4.7 kb from 10 randomly selected transformants).

Wild type	No. of mutations				Total
	T	C	A	G	
T		11	3	1	15
C	6		2	2	10
A	2			2	4
G	2		3		5
Total	10	11	8	5	34

sequence is important for recombination due to increased stability of the pairing, which supports the exceptionally high recombination efficiency of telomeric repeats (9). Frequent recombination between telomeres (12) and between plasmids and chromosomes (5) has also been reported. Therefore, the high frequency of *in vivo* recombination obtained with HARS36 probably results from direct integration of transforming DNA fragments into the multiple and recombinogenic targets in the chromosome of *H. polymorpha*.

A mutant library for directed evolution can be constructed by *in vivo* recombination of PCR-amplified DNA fragments with the acceptor vector. In *Bacillus subtilis* and *Acinetobacter calcoaceticus*, PCR-amplified DNA was efficiently captured by marker replacement recombination (18). Also, in *S. cerevisiae*, shuffled P450 gene fragments were recombined with the vector fragment in the cell to constitute a library (1). To use PCR-based gene fragments for library construction through *in vivo* recombination in *H. polymorpha*, the overlapping sequences of the acceptor and the insert were optimized in this study. The use of *in vivo* recombination diminished the problems usually encountered in library construction in *H. polymorpha*, such as variations in the copy number and the integration locus. Transformants from the circular vector with HARS36 usually showed over 50% multiple integrations (3). When the acceptor and the insert shared the 475-bp 3'-overlapping sequence in HARS36, the portion of the transformants with multiple gene integration was approximately 15%. After deletion of the shared HARS36 from each fragment, all tested clones showed single-copy gene integration and an even level of CALB activity, probably caused by recircularization and multimerization inabilities of the two linearized transforming DNAs. This fact will be useful for activity-based selection of mutant proteins after directed evolution. Moreover, the transformation efficiency was enhanced more than 10 times compared to that obtained with the circular plasmid. It will be sufficient for construction of a highly diverse library. The cloning steps can also be greatly simplified by eliminating the library-cloning step in *E. coli*.

In this work, we constructed a mutant library for *C. antarctica* lipase B using *in vivo* recombination and error-prone PCR techniques. We recovered diverse lipase sequences from randomly selected clones, indicating efficient construction of a random library. The *in vivo* recombination technique using HARS36 will be useful for generation of diverse gene libraries for directed evolution in *H. polymorpha*.

REFERENCES

1. Abecassis, V., D. Pompon, and G. Truan. 2000. High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **28**:E88.
2. Agapionov, M. O., A. I. Poznyakovski, A. I. Bogdanova, and M. D. Ter-Avanesyan. 1994. Isolation and characterization of the *LEU2* gene of *Hansenula polymorpha*. *Yeast* **10**:509–513.
3. Agapionov, M. O., P. M. Trushkina, J. H. Sohn, E. S. Choi, S. K. Rhee, and M. D. Ter-Avanesyan. 1999. Vectors for rapid selection of integrants with different plasmid copy numbers in the yeast *Hansenula polymorpha* DL1. *Yeast* **15**:541–551.
4. Brahmachari, S. K., G. Meera, P. S. Sarkar, P. Balagurumoorthy, J. Tripathi, S. Raghavan, U. Shaligram, and S. Pataskar. 1995. Simple repetitive sequences in the genome: structure and functional significance. *Electrophoresis* **16**:1705–1714.
5. Dunn, B., P. Szauter, M. L. Pardue, and J. W. Szostak. 1984. Transfer of yeast telomeres to linear plasmids by recombination. *Cell* **39**:191–201.
6. Gellissen, G. 2000. Heterologous protein production in methylotrophic yeasts. *Appl. Microbiol. Biotechnol.* **54**:741–750.
7. Gellissen, G., Z. A. Janowicz, A. Merckelbach, M. Piontek, P. Keup, U. Weydemann, C. P. Hollenberg, and A. W. Strasser. 1991. Heterologous gene expression in *Hansenula polymorpha*: efficient secretion of glucoamylase. *Bio/Technology* **9**:291–295.
8. Gellissen, G., M. Piontek, U. Dahlems, V. Jenzelewski, J. E. Gavagan, R. DiCosimo, D. L. Anton, and Z. A. Janowicz. 1996. Recombinant *Hansenula polymorpha* as a biocatalyst: coexpression of the spinach glycolate oxidase (GO) and the *S. cerevisiae* catalase T (*CTT1*) gene. *Appl. Microbiol. Biotechnol.* **46**:46–54.
9. Gray, M., and S. M. Honigberg. 2001. Effect of chromosomal locus, GC content and length of homology on PCR-mediated targeted gene replacement in *Saccharomyces*. *Nucleic Acids Res.* **29**:5156–5162.
10. Hollenberg, C. P., and G. Gellissen. 1997. Production of recombinant proteins by methylotrophic yeasts. *Curr. Opin. Biotechnol.* **8**:554–560.
11. Holm, C., D. W. Meeks-Wagner, W. L. Fangman, and D. Botstein. 1986. A rapid, efficient method for isolating DNA from yeast. *Gene* **42**:169–173.
12. Horowitz, H., P. Thorburn, and J. E. Haber. 1984. Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**:2509–2517.
13. Kang, H. A., J. H. Sohn, E. S. Choi, B. H. Chung, M. H. Yu, and S. K. Rhee. 1998. Glycosylation of human α 1-antitrypsin in *Saccharomyces cerevisiae* and methylotrophic yeasts. *Yeast* **14**:371–381.
14. Katinka, M. D., and F. M. Bourgain. 1992. Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*. *EMBO J.* **11**:725–732.
15. Kim, S. Y., J. H. Sohn, Y. R. Pyun, and E. S. Choi. 2002. A cell surface display system using novel GPI-anchored proteins in *Hansenula polymorpha*. *Yeast* **19**:1153–1163.
16. Kunes, S., H. Ma, K. Overbye, M. S. Fox, and D. Botstein. 1987. Fine structure recombinational analysis of cloned genes using yeast transformation. *Genetics* **115**:73–81.
17. Marykwas, D. L., and S. E. Passmore. 1995. Mapping by multifragment cloning *in vivo*. *Proc. Natl. Acad. Sci. USA* **92**:11701–11705.
18. Melnikov, A., and P. J. Youngman. 1999. Random mutagenesis by recombinational capture of PCR products in *Bacillus subtilis* and *Acinetobacter calcoaceticus*. *Nucleic Acids Res.* **27**:1056–1062.
19. Mezard, C., and A. Nicolas. 1994. Homologous, homeologous, and illegitimate repair of double-strand breaks during transformation of a wild-type strain and a *rad52* mutant strain of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14**:1278–1292.
20. Mezard, C., D. Pompon, and A. Nicolas. 1992. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. *Cell* **70**:659–670.
21. Oldenburg, K. R., K. T. Vo, S. Michaelis, and C. Paddon. 1997. Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. *Nucleic Acids Res.* **25**:451–452.
22. Orr-Weaver, T. L., and J. W. Szostak. 1985. Fungal recombination. *Microbiol. Rev.* **49**:33–58.
23. Pompon, D., and A. Nicolas. 1989. Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions. *Gene* **83**:15–24.
24. Raymond, C. K., T. A. Pownder, and S. L. Sexson. 1999. General method for plasmid construction using homologous recombination. *BioTechniques* **26**:134–138, 140–141.
25. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. Sohn, J. H., E. S. Choi, H. A. Kang, J. S. Rhee, and S. K. Rhee. 1999. A family of telomere-associated autonomously replicating sequences and their functions in targeted recombination in *Hansenula polymorpha* DL-1. *J. Bacteriol.* **181**:1005–1013.
27. Sohn, J. H., E. S. Choi, H. A. Kang, J. S. Rhee, M. O. Agapionov, M. D. Ter-Avanesyan, and S. K. Rhee. 1999. A dominant selection system designed for copy-number-controlled gene integration in *Hansenula polymorpha* DL-1. *Appl. Microbiol. Biotechnol.* **51**:800–807.
28. Sohn, J. H., E. S. Choi, C. H. Kim, M. O. Agapionov, M. D. Ter-Avanesyan, J. S. Rhee, and S. K. Rhee. 1996. A novel autonomously replicating sequence (ARS) for multiple integration in the yeast *Hansenula polymorpha* DL-1. *J. Bacteriol.* **178**:4420–4428.
29. Van Antwerp, J. J., and K. D. Wittrup. 2000. Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.* **16**:31–37.
30. Veale, R. A., M. L. Giuseppe, H. M. van Eijk, P. E. Sudbery, and C. T. Verrips. 1992. Development of a strain of *Hansenula polymorpha* for the efficient expression of guar α -galactosidase. *Yeast* **8**:361–372. (Erratum, **8**:813.)