Simultaneous cloning of open reading frames into several different expression vectors

Clement A. Stanyon, Thawornchai Limjindaporn, and Russell L. Finley, Jr.
Wayne State University, Detroit, MI, USA


Genomic sequencing has enabled the prediction of thousands of genes, most of which either cannot be assigned a function or can be only broadly categorized on the basis of sequence alone. High-throughput strategies for elucidating protein function are of high priority, and numerous approaches are being developed. Many of these approaches require the cloning of open reading frames (ORFs) into expression vectors that enable the encoded proteins to be tested for biological and biochemical activities. Typically, more than one type of vector must be employed, as different experiments require different conditions of protein production. Here we show that it is possible to simultaneously transfer a single ORF from a source vector to four target vectors using a commercially available in vitro recombination system. To test the approach, we constructed new vectors for expression of fusion proteins in yeast, including vectors for the LexA two-hybrid system. We show that individual ORFs can be efficiently transferred to four different vectors in a single in vitro reaction. The resulting expression plasmids can be separated using prototrophic markers specific to each vector. Using this system to produce multiple expression constructs simultaneously could greatly facilitate high-throughput subcloning and proteomic studies.

INTRODUCTION

The majority of proteins predicted from genome sequences are only poorly characterized or have unknown functions. To address this problem, numerous high-throughput approaches are being developed to obtain information about the function of many proteins at once (1). It is increasingly recognized, however, that full characterization of a proteome will require application of several different techniques. This is because any given study provides only one kind of functional data, and although many proteins are examined in parallel, the amount of information obtained about each protein is minimal. Moreover, as with conventional techniques, data obtained from high-throughput approaches often require validation. Indeed, several recent studies suggest that the use of overlapping data from multiple high-throughput approaches can increase the accuracy of functional predictions (2–4). Thus, full characterization of a proteome will require the application of several different high-throughput approaches to the same set of proteins.

The development of different techniques for proteome characterization is resulting in an increasing need for efficient methods to subclone large sets of open reading frames (ORFs) into several different expression vectors. Recent studies to characterize the yeast proteome illustrate this point. High-throughput two-hybrid studies, for example, required construction of two plasmids for each of the approximately 6000 predicted yeast ORFs, to express each protein fused to either a DNA-binding domain (BD) or an activation domain (AD) (5,6). Other studies expressed epitope-tagged versions of yeast proteins, so that protein complexes could be purified with an affinity matrix, and the components of the complexes could be determined by mass spectrometry (7). The same yeast ORFs have also been used in studies to express glutathione S-transferase (GST)- or histidine (HIS)-tagged proteins to assay for a variety of biochemical activities (8) and to use in protein microarray experiments (9). Thus, a rate-limiting step in many proteomic studies is the process of subcloning hundreds to thousands of ORFs into the necessary expression vectors.

Recently, efficient in vitro recombination approaches have been developed to subclone ORFs into expression vectors (10–12). One such approach is the commercially available Gateway™ cloning system, which is based on the λ bacteriophage recombinase, a site-specific enzyme that can differentially recombine two mutated recombination sites, enabling directional exchange of DNA between two vectors or a PCR product and a vector (10). In this system, an ORF is first cloned into an entry vector, either by traditional in vitro ligation or by recombination. The entry plasmid is then used as a substrate from which to transfer the ORF into a chosen destination vector using an in vitro recombination reaction, resulting in an expression construct. The expression plasmids are recovered by transformation of Escherichia coli with the in vitro recombination reaction. The advantage of this system is that once an entry clone is obtained, the cloned ORF can easily be subcloned into each new expression vector by an in vitro recombination reaction. The efficiency of the system has made it the choice for a number of high-throughput proteomics studies (11,13–15), including an ongoing project to construct a repository of entry plasmids containing all the ORFs for several species (16).

Given the increasing need to subclone ORFs into multiple expression vectors, it would be particularly useful if sets of ORFs could be efficiently subcloned into more than one vector at a time. We reasoned that the in vitro reaction of the Gateway system ought to be able to transfer DNA from a single entry clone to multiple destination vectors in a single reaction. This would create a mixture of expression clone constructs varying by vector backbone. If each type of expression clone had a different selectable marker, the vector types could be resolved by cloning through bacteria. Here we demonstrate that simultaneous high-fidelity transfer into at least four vectors is possible. When transformed into an appropriate bacterial strain, the constructs can be separated according to the markers present in the destination vectors.

MATERIALS AND METHODS

Microbial Strains, Transformations, and Media

E. coli strain KC8 (pyrF::Tn5 hsdR leuB600 trpC9830 lacΔ74 strA galK

520 BioTechniques
hisB436) was obtained from Kevin Struhl (17). The auxotrophies of this strain for histidine, leucine, tryptophan, and uracil, are complemented by the yeast markers HIS3, LEU2, TRP1, and URA3, respectively, present on the destination vectors. KC8 bacteria, transformed with plasmids bearing one of the four yeast markers, were selected on M9 minimal media (17) with 100 µg/mL ampicillin and supplemented with three of the following four nutrients, appropriate to the marker being selected: histidine, leucine, tryptophan, and uracil, each at 12 µg/mL. The strain was transformed by the polyethylene glycol dimethyl sulfoxide (PEG-DMSO) protocol (11) in 96-well microtiter plates. For each transformation, 50 µL cells were mixed with 5 µL of the in vitro LR recombination reaction [containing destination vector(s), entry vector clone, and LR Clonase] in 1 x KCM buffer (0.1 M KCl, 30 mM CaCl2, 50 mM MgCl2), and aliquots were plated onto the appropriate minimal media. Yeast strains RFY231 (MATa, his3, trp1Δ::hisG, ura3, 3LexAop-LEU2::let2 mal+) and RFY206 (MATa, trp1Δ::hisG, his3Δ200 leu2-3 lys2Δ201 ura3-52 mal-) were previously described (18,19).

Plasmids

Additional information, including details of plasmid construction, and Supplementary Tables and Figures can be found at (http://www.biotechniques.com/Sept03/FinleySupplementary.html). Gateway destination vectors (Invitrogen, Carlsbad, CA, USA), containing the ccdB/Cmr cassette flanked by the attR recombination sites, were constructed for five different yeast vectors (see Supplemental Table 1, Sequencing/PCR Primers, and Supplemental Figure 2, New Gateway destination vectors for expression of proteins in yeast). pLex-attR and pLexN-attR were constructed from pEG202 (20) and pNLexNLS (19), which are HIS3 vectors for expression of LexA fusions. pLexN-attR includes a nuclear localization signal. pTet-attR was constructed from pCWX200, a LEU2 vector for expression of TetR DNA-binding domain fusions (21). pB42-attR was derived from pG4-S, a TRP1 vector for expression of fusions to the B42 transcription AD (22), and pFLAG-attR is derived from pESC-URA3 (Stratagene, La Jolla, CA, USA), a URA3 vector for expression of FLAG®-tagged proteins.

Expression Clone Construction

ORFs were amplified using approximately 40-mer primers with approximately 20-mer of homology to the ORF of interest, tailed with 5′ common sequences that allowed a secondary amplification to add the complete attB recombination sites. The common tails used were 5′-ACA-AAAAAGCAGGCTTAAATG-3′ and 5′-GGGTTTTTCTCAGTACTCGA (stop codon)-3′. Following primary amplification, extension primers used were 5′-GGGACCAAGTTTGACA-AAAAAGCAGGCT-3′ and 5′-GGGACCCTTTGTCACAAGAAAA-GCTGGGCTGGTCTTTTACTCGA-3′. The attB1 and attB2 sequences are underlined. Entry plasmids were constructed by transferring the PCR products into pDONR201 (Invitrogen) using the BP in vitro recombination reaction as described by the manufacturer and verified by sequencing (Supplemental Figure 1, Method used for cloning ORFs into Gateway vectors). Campylobacter jejuni ORFs were provided by J. Parrish (J. Parrish, T. Limjindaporn, and R.L. Finley, Jr., unpublished data). Drosophila ORFs for Cdk2 and cyclin J were described elsewhere (23). Transfer of ORFs from entry clones into destination vectors using the LR in vitro recombination reaction was done as described by the manufacturer. This reaction catalyzes recombination between specific sites (attR1 and attR2) flanking the ORF of an entry clone and reciprocal sites (attL1 and attL2) flanking the insertion site of a destination vector, resulting in transfer of the ORF into the destination vector to create an expression plasmid (10). All destination vectors carry the counter-selectable ccdB gene between the attR sites, which is lethal in strains lacking a gyrase mutation (gyrA462) and which ends up in the by-product of the LR reaction. To maximize efficiency of the in vitro recombination reaction, destination vectors were linearized using restriction enzymes that cut within the ccdB/Cmr cassette. Following phenol extraction and filtration with Centricon®-100 filters (Milipore, Billerica, MA, USA), linearized vectors were diluted to 50 µg/mL. The four-way transfer LR reactions were performed in 96-well plates and contained 50 ng of each vector.

Yeast Two-Hybrid Assays

Yeast two-hybrid assays were performed as described (18,19). Briefly, strains expressing LexA-fused proteins were mated on rich media with strains expressing AD-fused proteins, and the resulting diploids were replica-plated to media to assay for expression of the two-hybrid reporters. The LEU2 reporter was assayed by growth on media lacking uracil, and the lacZ reporter was assayed on X-gal indicator plates. The indicator plates were complete minimal (CM) dropout plates (19) lacking uracil, histidine, and tryptophan and containing 2% galactose and 0.2% raffinose. Additional details and protocols can be found at (http://finlab.med.edu).

RESULTS AND DISCUSSION

To test whether an ORF can be transferred into multiple vectors in a single in vitro reaction, we constructed destination vectors with different markers that could be independently selected. We chose to construct yeast vectors that express fusions applicable to yeast two-hybrid, three-hybrid, and other studies. These vectors carry different yeast prototrophic marker genes that are effective in both yeast and bacteria: URA3, HIS3, TRP1, and LEU2. The HIS3 and TRP1 vectors correspond to the BD and AD expression vectors from the LexA-based two-hybrid system (22). The URA3 plasmid allows expression of a FLAG epitope-tagged protein from the yeast GAL10 promoter, and the LEU2 plasmid constitutively expresses Tet repressor (TetR) BD fusion proteins for use in TetR two-hybrid assays (21). We tested these vectors by transferring ORFs into them from several entry clones, using the Gateway in vitro recombination reaction (or the LR reaction). The desired expression plasmids could be isolated
from the reaction by transformation of a gyrE strain and selection on ampicillin. The markers we have employed allow an additional selection, for prototrophy in a multiply auxotrophic *E. coli* strain, such as KC8. This strain bears the mutations *leuB600*, *trpC9830*, *hisB436*, and *pyrF*::Tn5, which are complemented by yeast *LEU2*, *TRP1*, *HIS3*, and *URA3*, respectively. We transformed KC8 with individual LR in vitro recombination reactions for each of the four yeast destination vectors and plated the transformation mixture on minimal media lacking the appropriate nutrient: leucine, tryptophan, histidine, or uracil. Plasmid preparations from the colonies showed that over 90% of the cloned constructs contained the desired ORF (data not shown). Thus, the four new yeast vectors can be used to construct expression clones by the Gateway LR reaction, and recombinants can be selected in auxotrophic *E. coli*.

The LR recombination reaction creates 25-bp sites (attB1 and attB2) flanking the ORF in the expression plasmid, which introduces an additional nine amino acids into an expressed protein between the fusion moiety and protein. Previous studies in the Gal4 two-hybrid system have indicated that this does not prevent the detection of interactions (24). Expression plasmids that we constructed for the LexA two-hybrid system also showed the expected interaction phenotypes when attB1 sites were present, indicating that the Gateway cloning approach can be used to construct functional two-hybrid clones in the LexA system (Figure 1 and below). The ability to easily subclone the same set of ORFs from entry clones into two different two-hybrid system vectors could be particularly useful for constructing protein interaction maps, since the two systems can produce different false negative and false positive results.

To test whether we could clone into four vectors in one reaction, and to estimate the overall efficiency, we constructed 48 entry clones containing different ORFs from *C. jejuni* with sizes ranging from 240 to 1872 bp (see Materials and Methods). Next, we attempted to transfer the *C. jejuni* ORFs into the four destination vectors bearing different markers. Each entry clone was mixed with the four destination vectors in a single well of a 96-well plate. After incubation with LR clonase to mediate in vitro recombination, the mixtures were used to transform chemo-competent *E. coli* strain KC8, and aliquots from each reaction were plated onto four types of minimal media to select the different plasmid products in each reaction. Each transformation resulted in colonies on all four media. We isolated plasmid DNA from single colonies on each of the four media and tested for inserts by PCR using primers that were specific to each destination vector. A total of 184 out of 192 plasmids (96%) had inserts of the expected size (Figure 2). This efficiency is similar to that obtained when subcloning into a single destination vector. To examine the overall structure of the plasmids, we performed restriction enzyme digests using plasmid DNA derived from 79 of the expression clones having inserts. A total of 75 digests had the expected pattern of bands, indicating that the in vitro recombinants contained the expected insert, markers, and no gross rearrangements (Figure 3). The four other plasmid preparations had evidence of the correct plasmid plus an additional plasmid with a different marker (see below and Supplemental Figure 3, Digestion of the DNA preparations). Finally, we tested some of the expression clones for their ability to express functional proteins. To do this, we performed yeast two-hybrid assays with selected expression clones encoding known interactors, based on preliminary data from a *C. jejuni* protein interaction mapping project (J. Parrish, T. Limjindaporn, and R.L. Finley, Jr., unpublished data). We were able to detect all of the four known interactions that we tested among four BD and four AD expression clones (see Supplemental Figure 4, Interaction profiles of expression clones constructed in multiple vector transfer reactions). Combined, these results indicate that simultaneous transfer of ORFs into four vectors in a single in vitro reaction can be achieved with high efficiency and high fidelity.

Consideration of the reactants, products, and unresolved intermediates suggests that transfer between destination vectors should not occur. To assure that this is indeed the case, we transformed yeast with plasmid DNA prepared from 48 individual KC8 transformant colonies, corresponding to 12 ORFs in four different vectors. All 48 transformations produced yeast that grew on the selective media corresponding to the marker present on each plasmid. Surprisingly, some of the transformations appeared to contain yeast transformed with heterologous plasmids, with four samples giving large numbers of heterologous transformants, indicating that the yeast had acquired two different marker genes (see Supplemental Figure 5, Yeast transformed with KC8 bacterial DNA preparations). As noted above, restriction digests of these four DNA preparations suggested that they had two different plasmids. To confirm this, we
transformed bacteria with these DNA samples and analyzed plasmids from single colonies. Colonies from each transformation had one or the other of the two different plasmids with different markers, indicating that the different plasmids or markers were not covalently linked (not shown). Thus, in 4 out of the 48 plasmid preparations tested, there were two different expression clones.

Figure 2. Efficiency of transfer of 48 open reading frames (ORFs) into four destination vectors. Inserts were amplified by PCR from bacterial plasmid DNA using primers specific to each destination vector. Plasmids were prepared from single colonies that resulted from transformation with 48 LR in vitro recombination reactions. Each LR reaction had one of 48 different entry clones plus all four destination vectors, pB42-attR (TRP1), pTetR-attR (LEU2), pLex-attR (HIS3), and pFLAG-attR (URA3). Each transformation was plated onto four different media to select for the marker in one of the destination vectors: (A) TRP1; (B) LEU2; (C) HIS3; (D) URA3. Of 192 expression plasmids, 184 (approximately 96%) had inserts of the expected size; product size differences between vectors are due to the different amounts of vector in each PCR product (see Supplemental Table 1, Sequencing/PCR Primers). The 1-kb Plus marker (Invitrogen) was used as a DNA ladder.

Figure 3. Restriction profiles of expression clones. Expression clones showed the expected restriction enzyme digestion profiles. Representative XmnI digests of 20 plasmids derived from cloning five open reading frames (ORFs) into each of the four expression vectors containing different yeast markers: (A) HIS3; (B) TRP1; (C) LEU2; (D) URA3. Arrows indicate the bands containing the ORF inserts. Expression clones from lanes 1, 2, 3, 4, and 5 were originally derived from the Campylobacter jejuni ORFs, cj0056c (422 bp), cj0036 (1308 bp), cj0036 (1350 bp), cj1190c (1380 bp), and cj0695 (1389 bp), respectively. These ORFs do not have XmnI sites. All digests resulted in the predicted pattern of XmnI digestion, which supported the integrity of the vectors after the LR reactions and selection in Escherichia coli. For example, the TRP1 vector has three XmnI sites, so that XmnI digestion leads to three bands, one of which varies with the size of the insert. Each gel also contains the 1-kb DNA ladder size marker.
from two different vector types. Additional testing of DNA samples capable of yielding heterologous transfectants indicates that approximately 1 in 50 yeast will, on average, be transformed by a heterologous plasmid, when the highly contaminated samples are omitted from consideration (see Supplemental Table 2, Contamination by Heterologous Plasmids). These data underscore the need for a characterization step following creation of expression constructs with this method, as with any plasmid construction method.

Current strategies for proteome-wide protein characterization would benefit from this technique. A key feature of the simultaneous transfer approach is that it would allow high-throughput construction of several expression constructs for each of many ORFs. We have shown that the in vitro recombination reaction and subsequent E. coli transformation can be performed in standard 96-well microtiter plates, so that 96 different ORFs can be transferred into four different vectors, to create 384 expression constructs in one plate of reactions. Although we only tried up to four vectors at one time, the fact that we did not see a significant decrease in efficiency with multiple vectors suggests that the in vitro reactions may be able to support subcloning into more than four vectors simultaneously. One limiting factor would be the ability independently to select the different expression constructs in E. coli. Other antibiotic resistance markers might be used to increase the number and variety of vectors that may be used in a single reaction. Finally, although we have chosen to test this approach with yeast vectors, it is likely that other vectors would work as well.

ACKNOWLEDGMENTS

The first two authors contributed equally to this work. We thank members of the laboratory for helpful discussions, reagents, and technical assistance. We are especially grateful to Dr. Jodi Parish for providing primary PCR products for C. jejuni genes, for preliminary data on C. jejuni protein interactions, and for critical comments on the manuscript. This work was supported by National Institutes of Health grant no. ROI HG01536 and the Michigan Proteome Consortium.

REFERENCES


Received 9 January 2003; accepted 4 June 2003.

Address correspondence to Russell L. Finley, Jr., Center for Molecular Medicine and Genetics, Wayne State University, 540 E. Canfield Ave., Detroit, MI, 48201, USA. (http://proteome.wayne.edu) e-mail: rfinley@genetics.wayne.edu