Optimization of Yeast Surface-Displayed cDNA Library Screening for Low Abundance Targets

Juhyung Kim, Hyung Kyu Kim, Hye Jeong Jang, Eunkyung Kim, and Moon Kyu Kim*

Department of Immunology, School of Medicine, Kyungpook National University, Daegu 700-422, Republic of Korea

The yeast surface-displayed cDNA library has been used to identify unknown antigens. However, when unknown target antigens show moderate-to-low abundance, some modifications are needed in the screening process. In this study, a directional random-primed cDNA library was used to increase the number of candidates for the unknown antigen. To avoid the loss of target yeast clones that express proteins at a low frequency in the cDNA library, a comprehensive monitoring system based on magnetic-activated cell sorting, fluorescence-activated cell sorting, and immunofluorescence was established, and a small number of target yeast cells was successfully enriched. These results showed that our optimized method has potential application for identifying rare unknown antigens of the human monoclonal antibody.

Keywords: cDNA library, yeast surface display, tumor-associated antigen, hepatocellular carcinoma

Introduction

The yeast surface display technique has been used for a wide range of applications in antibody engineering and antigen identification [15, 16]. Affinity maturation of single-chain antibody fragments (scFv) via the yeast surface display system has become a popular tool for the development of novel therapeutics [13]. This system can also be used to screen for target yeast cells displaying desired proteins in a combinatorial random library [3]. The yeast surface-displayed cDNA library generated from tumor cells or tissues can be used for the screening of tumor-related antigens [18]. Compared with conventional serological expression cloning (SEREX), which is based on bacterial or phage display systems, the identification of tumor-specific antigens using the yeast surface display has many advantages, such as post-translational modifications [2]. Several novel tumor-specific antigens have been found, not only with the sera of cancer patients but also with phage-displayed antibodies [1, 14]. However, optimization of several aspects of the yeast surface display system is essential for detecting target proteins that show moderate-to-low expression on the tumor cell surface.

To detect target yeast cells expressing moderate-to-low abundance proteins, it is necessary to increase the size and diversity of the cDNA library and cloning efficiency. Previous studies usually applied the oligo-dT primed cDNA library or non-directional random-primed cDNA library. However, an oligo-dT primed cDNA library does not represent the mRNA uniformly, because it is likely to miss the 5' region of the mRNA [10]. Moreover, in a non-directional random-primed cDNA library, only 2–3% of the independent yeast cells can usually display the in-frame fusion protein [2]. Thus, we applied a directional random-primed cDNA library in this study to maximize the size and diversity of the cDNA library, because the quality of the cDNA library is a fundamental feature that affects the success or failure of identification of the moderate-to-low abundance antigen by the yeast surface display system [6, 10]. Furthermore, because the low abundance target yeast cells can be lost in the early phases of enrichment [4, 14], a comprehensive system for monitoring and controlling the enrichment process is essential. On the basis of magnetic-activated cell sorting (MACS) and fluorescence-activated...
cell sorting (FACS), we extensively used immunofluorescence (IF) for not only interpreting FACS data but also deciding the appropriate timing from MACS to FACS. For example, only after confirming that the target yeast cells with a clear ring pattern were concentrated to a certain range, FACS was started. Moreover, even through a series of FACS, we consistently monitored the staining pattern with IF to examine the quality of the staining and analyze enrichment patterns correlated with the sequence data. This extensively comprehensive and combinatorial analyzing system is one of the characteristics of our optimized method, which is distinguished from previous studies [1, 3, 4, 14, 18].

We previously produced liver-cancer-specific human monoclonal antibodies by using the method established by Foungetal. [7, 20]. To facilitate the identification of their unknown antigens, we optimized the yeast surface display method. Before identifying the unknown antigens of our antibodies, it was necessary to verify that this method could efficiently isolate the target yeast cells. For this, we selected actin as the target gene from the cDNA library generated from a hepatocellular carcinoma cell line. Actin is a well-conserved ubiquitous protein with various isotypes. Human liver tissue expresses at least six unique actin isotypes. Their frequency ranges from 0.262% (β-actin) to 0.002% (α- and γ-actin). The frequency of the α-actin cDNA clones in human liver tissue is 0.002%, 100 times less than that in β-actin (0.2–0.4%). This study demonstrates how to capture low abundance target yeast cells such as α-actin that constitute only $10^3$–$10^5$ out of the entire yeast cell population ($10^7$–$10^8$) in the initial round of the enrichment process. Our optimized method has potential application for identifying low abundance unknown antigens of a therapeutically utilizable human antibody [9].

### Materials and Methods

#### Cells

HepG2, a human hepatoblastoma cell line, and SNU475, a hepatocellular carcinoma (HCC) cell line, were obtained from the American Type Culture Collection (Rockville, MD, USA) and Korean Cell Line Bank (Seoul, South Korea), respectively. The HepG2 and SNU475 cells were cultured in Dulbecco’s modified Eagle’s medium (Gibco, Rockville, MD, USA) and RPMI (Gibco) media supplemented with 10% fetal bovine serum (Gibco), respectively.

#### Directional Random-Primed cDNA Library Construction

Total RNA was extracted from the liver cancer cell lines using TRIzol (Invitrogen, Carlsbad, CA, USA). mRNA was obtained using the mRNA Purification Kit (Invitrogen). A random primer with a 3’ NotI sequence was used to synthesize the first strand of the directional random-primed cDNA library. The second strand was synthesized using E. coli RNase H, DNA polymerase, and E. coli DNA ligase. The cDNA product was rendered blunt using T4 DNA polymerase and ligated to an EcoRI adapter. After phosphorylation with T4 polynucleotide kinase, the cDNA was digested with NotI, allowing it to be directionally inserted into pYDI (Invitrogen) that acts as a shuttle vector between the E. coli and the Saccharomyces cerevisiae strain EBY100 (Invitrogen). Size-selected cDNAs were fused with the pYDI cut by EcoRI/NotI. The products were purified and transformed into ElectroMAX DH10B cells (Invitrogen) by electroporation.

#### Yeast Transformation

Before transformation, yeast strain EBY100 at $OD_{600} = 0.2$ was grown in the YPD medium (1% yeast extract, 2% peptone, and 2% dextrose) in a shaking incubator (200 rpm) at 30°C overnight. EBY100 was grown to $OD_{600} = 4$–6, diluted to $OD_{600} = 0.4$, and incubated again in a shaking incubator at 30°C for 3 h. The cells were collected, washed, incubated in 2 ml of 1× LiAc/0.5× Tris- EDTA (pH 7.5) at RT for 10 min, and then mixed with 1 µg of cDNA and 100 µg of carrier DNA (denatured sheared salmon sperm DNA; Roche, Fenzberg, Germany) per 100 µl of the yeast solution. After 5–10 sec of vigorous mixing, 700 µl of 1× LiAc/40% PEG/1× Tris-EDTA was added with vortexing and incubated in the shaking incubator at 30°C for 30 min. After adding 88 µl of dimethyl sulfoxide (Applichem, Darmstadt, Germany), the mixture was immediately heat shocked at 42°C for 7–15 min and chilled on ice for 1–2 min. The transformed cells were incubated in 1 ml of YPD at 30°C for 90 min, and grown in SD-CAA selective medium (2% dextrose, 0.67% yeast nitrogen base, 0.5% casamino acids, 1% penicillin/streptomycin, and 0.05% kanamycin).

#### Magnetic-Activated Cell Sorting

To induce the SNU475 yeast display library, the cell population grown in SD-CAA medium was moved to SG-CAA medium (SD-CAA with 2% galactose instead of dextrose) and incubated at 20°C for 40 h. In 100 µl of MACS buffer (PBSM; 0.8% NaCl, 0.02% KCl, 0.144% NaHPO₄, 0.024% KH₂PO₄, 0.5% BSA, and 0.0744% EDTA-2Na), $3 \times 10^7$ induced cells were added and labeled using 5 µg of polyclonal rabbit anti-actin (Invitrogen) at RT for 30 min. After washing with PBSM, 20 µl of anti-rabbit IgG Microbeads (MiltenyiBiotec GmbH, Gladbach, Germany) and 80 µl of resuspended cells were incubated at 4°C for 30 min. After washing again, 3 × 10⁷ cells in 500 µl of PBSM were loaded onto the LS column (MiltenyiBiotec GmbH, Gladbach, Germany) surrounded by magnet and washed with 10 ml of PBSM. MAC-sorted cells were incubated in 4 ml of SD-CAA medium with 100 µl of YPD at 30°C for 2 days [4].

#### Fluorescence-Activated Cell Sorting and Immunofluorescence

After three rounds of MACS, the enriched SNU475 yeast
display library was induced for FACS. From the first to the third round of FACS, $0.2 \times 10^7 – 0.8 \times 10^7$ yeast cells were sorted through FACSCalibur equipped with the sorting accessories (BD Bioscience, San Jose, CA, USA). In the fourth round of FACS, $3.1 \times 10^7$ yeast cells were run through FACS AriaIII (BD Bioscience). The induced yeast cells in 100 µl of PBSF (0.8% NaCl, 0.02% KCl, 0.144% Na2HPO4, 0.024% KH2PO4, 0.1% BSA, and 0.0744% EDTA-2Na) were labeled with 1 µg of mouse anti-Xpress (Invitrogen) and 5 µg of polyclonal rabbit anti-actin (Invitrogen) antibodies. Two-color labeling was performed with 1 µg of FITC-conjugated donkey anti-mouse IgG (Jackson Labs, PA, USA) and 1 µg of PE-conjugated goat anti-rabbit IgG (Santa Cruz Biotechnology, CA, USA). Samples thus prepared could be used for FACS and IF. For IF, an aliquot of the stained yeast cells was loaded onto a glass slide with a cover-glass. The sorted yeast cells were grown in SD-CAA medium at 30°C for 3 days.

**Plasmid Extraction from EBY100 and Sequence Analysis**

The aliquots of the sorted cells grown in SD-CAA solid medium (including 1.5% agarose) were subjected to sequencing analysis. The plasmids were extracted using a Zymoprep Yeast Plasmid MiniprepII Kit (Zymo Research, Orange, CA, USA). The extracted plasmids were transformed into DH5α cells and isolated using a Plasmid DNA Purification kit (Promega, Madison, WI, USA), because of the scarcity of the extracted plasmid DNA from yeast cells. Every insert sequence was analyzed using the NCBI database.

**Results**

**Construction of the Yeast Surface Display Library**

Two kinds of directional random-primed cDNA libraries were constructed from HepG2 and SNU475 liver cancer cell lines. To increase the efficiency of transformation, 20 transformations were performed in parallel. A total of 20 µg of cDNA was cloned into the *Saccharomyces cerevisiae* EBY100 strain (Fig. 1A). About $6 \times 10^6$ independent primary clones were obtained from the two cancer cell lines. The transformation efficiency was $3 \times 10^5$ colony forming unit (CFU) per 1 µg of cDNA. To assess the quality of the libraries, 20 yeast clones were randomly selected and sequenced. The average insert size was found to be ~0.6 kb. The independent yeast clones that displayed the in-frame fusion protein constituted 11% and 5.2% of the cDNA library of HepG2 and SNU475 cells, respectively. The percentage of the in-frame fusion protein obtained in the directional random-primed cDNA library in this study was 2–3 times higher than that obtained from the oligo-dT or non-directional random-primed cDNA library [2]. The cDNA insert size ranged from ~100 to ~1,000 bp, with an average of 700 bp in HepG2 and 430 bp in SNU475 (Table S1).

**Concentration of the Target Yeast Cells with MACS**

The frequency of the α-actin cDNA clones is 0.002% (5/205,291 via NCBI EST database) of the yeast cells expressing the in-frame fusion protein (i.e., 5–11% of the total yeast cells). Thus, the initial number of in-framed α-actin yeast clones before the first round of MACS was estimated to be $10^3$–$10^4$, which is about 0.0001–0.0002% of the entire yeast cell population (Fig. 1B). In the first two rounds of MACS, there was no obvious increase in the targeted yeast cells in the FACS analysis (Fig. 2A). However, after the third round of MACS, prominent enrichment of the targeted yeast cells was observed in the FACS analysis (Fig. 2A). After the first round of MACS, about $10^3$–$10^5$ out of $3 \times 10^7$ yeast cells were obtained from various concentrations of the anti-actin antibody (Fig. 2B). The yeast cell population enriched to ~1% was used for the next round of MACS. In three rounds of MACS, the yeast cells with a clear ring pattern were enriched to 0.5–1.0% (Fig 2C). Only after confirming that the target yeast cells were concentrated to 0.1–1.0% was the enrichment device switched from MACS to FACS.

**Fig. 1.** The system for displaying the cDNA library on the yeast surface and the process for enriching the target clones. (A) Schematic representation of the yeast surface display library system. The directional random-primed cDNA library from hepatocellular carcinoma (HCC) was transformed into *Saccharomyces cerevisiae* strain EBY100. The unknown candidate antigens (red) were expressed as fusion proteins next to the C-terminus of the Aga2p linked with Aga1p through double disulfide bonds on the surface of EBY100. Express tag was used as the indicator of induction (green). (B) Schematic representation of the enrichment process. The number of target yeast cells enriched through several rounds of MACS and FACS is indicated.
Enrichment of the Target Yeast Cells with FACS

Two-color labeling was performed with FITC as an indicator of induction and with phycoerythrin (PE) for the target protein. The gating was performed in the double-positive region. The specific region for gating could be expected on the basis of the enrichment pattern obtained through titration (Fig. S1A). Until the third round of FACS, we gated 0.3% of the total population (Fig. 3A, a-c) and then the gating was narrowed to 0.2% in the fourth round of FACS (Fig. 3A, d). Interestingly, although ~30% of the total yeast cells were double-positive in the third round of FACS (data not shown), only 2-5% of the yeast cells displayed the clear ring pattern during the fluorescence microscopic examination (Fig. 3B, e). As shown in Table 1, after the second round of FACS, only ~6% of the α-actin clones remained. Thus, in 30% of the total yeast cells at the third round of FACS, the nonspecifically stained yeast cells were mixed. After changing the secondary antibody in the fourth round of FACS, the nonspecific population was decreased to 5% (Table 1) and the percentage of the bright ring pattern increased to 25–30% (Fig. 3B, f). Considering that the percentage of induced cells stained with FITC was 50–70% (data not shown), the actual number of clones with the bright ring pattern is expected to be more than 50% of the entire induced cell population, which is equivalent to the percentage of the actin clones (ACTC 1) obtained in the sequencing analysis (Table 1).

DNA Sequence Analysis of the Isolated Yeast Clones

After the second round of FACS, the target yeast cells were increased to ~5%, which appeared as 1 α-actin clone out of 18 clones (Fig. 3A, b and Table 1). The sequencing analysis performed after the third round of FACS showed that only 18% of the population comprised the target α-actin clones, whereas 71% was nonspecific (Table 1). Thus, FACS analysis alone cannot detect the enrichment of the nonspecific population. To monitor the progress of the enrichment more accurately, repeated DNA analysis and IF data are needed. When the secondary antibody was changed from PE-conjugated goat anti-rabbit to PE-conjugated...
donkey anti-rabbit, the 71% of the nonspecific population decreased to 5% (Table 1). After the fourth FACS sorting, three different kinds of muscular actin clones, including ACTC1 (alpha, cardiac muscle, 65%), ACTA2 (alpha, aorta smooth muscle, 5%), and ACTG2 (gamma 2, enteric smooth muscle, 5%) were detected (Table 1).

**Discussion**

The yeast surface display method is frequently used for identifying unknown antigens from cDNA libraries. However, it needs to be optimized for isolating low abundance target cells. The results of this study showed that our directional random-primed cDNA library is even more appropriate to increase the functional library size and diversity of the cDNA library than an oligo-dT-primed cDNA library. However, a limitation of this study is that our random-primed cDNA library with an average insert size of 700 bp is likely to miss bigger functional domains. To overcome this limitation, it is possible to mix the oligo-dT-primed library, which has a larger average size of DNA fragments, with the directional random-primed cDNA.

**Table 1.** List of isolated cDNA clones.

<table>
<thead>
<tr>
<th>GenBank Accession No.</th>
<th>Description of clones</th>
<th># Colonies*</th>
<th>% Colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>After second FACS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_005159.4</td>
<td>Actin, alpha, cardiac muscle 1 (ACTC1)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>NM_003486.5</td>
<td>Solute carrier family 7 (amino acid transporter light chain, L system), member 5 (SLC7A5)</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>NM_138820.2</td>
<td>HIG1 hypoxia inducible domain family member 2A (HIGD2A)</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>NM_002631.2</td>
<td>Phosphoglucomutase dehydrogenase (PGD)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>NM_002869.4</td>
<td>Rab5a, member RAS oncogene family (RAB5a)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>NR_003287.2</td>
<td>Homo sapiens RNA, 28S ribosomal 1 (RN28S1), ribosomal RNA</td>
<td>10</td>
<td>56</td>
</tr>
<tr>
<td><strong>After third FACS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_005159.4</td>
<td>Actin, alpha, cardiac muscle 1 (ACTC1)</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>NM_003486.5</td>
<td>Solute carrier family 7 (amino acid transporter light chain, L system), member 5 (SLC7A5)</td>
<td>12</td>
<td>71</td>
</tr>
<tr>
<td>NM_002869.4</td>
<td>Rab5a, member RAS oncogene family (RAB5a)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>NR_003287.2</td>
<td>Homo sapiens RNA, 28S ribosomal 1 (RN28S1), ribosomal RNA</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td><strong>After fourth FACS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_005159.4</td>
<td>Actin, alpha, cardiac muscle 1 (ACTC1)</td>
<td>13</td>
<td>65</td>
</tr>
<tr>
<td>NM_001615.3</td>
<td>Actin, gamma 2, smooth muscle, enteric (ACTG2)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>NM_001613.2</td>
<td>Actin, alpha 2, smooth muscle, aorta (ACTA2)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>NM_003486.5</td>
<td>Solute carrier family 7 (amino acid transporter light chain, L system), member 5 (SLC7A5)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>NM_002869.4</td>
<td>Rab5a, member RAS oncogene family (RAB5a)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>NR_003287.2</td>
<td>Homo sapiens RNA, 28S ribosomal 1 (RN28S1), ribosomal RNA</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

*The number of colonies.
library. The percentage of the membrane proteins was 7% (4 cDNA sequences encoding membrane proteins out of total 53 cDNA sequences) as compared with 93% of soluble proteins (Table S1). Although they were not in in-frame state, if a greater number of yeast cells are analyzed, there will be yeast cells displaying the in-frame membrane proteins. Thus, our directional random-primed cDNA library can display not only cytosolic proteins but also membrane proteins in the cells.

Furthermore, MACS proved to be very advantageous so as to avoid the loss of the low abundance target yeast cells in the early phases of enrichment [4, 14]. Concentrating the rare target yeast cells by MACS before FACS can increase the probability of isolating these cells [4, 19]. Besides this, the required number of FACS rounds was reduced owing to MACS, thus minimizing the exposure to nonspecific weak binders, which otherwise occurs during several rounds of FACS. Moreover, IF provides supplementary information to facilitate the interpretation of the results obtained from MACS and FACS. Thus, a comprehensive monitoring system based on MACS, FACS, and IF enables the detection of target yeast cells expressing moderate-to-low abundance proteins. Before enriching the target yeast cells with MACS, it was necessary to estimate how many rounds of MACS were to be performed, because the required rounds of MACS may differ according to the rarity of the target antigen. The target yeast clones should be concentrated at least up to 0.1–1.0% to avoid the loss of the rare clones during the gating of the first round of FACS (Fig. S1A). Before gating in the FACS window, IF was performed to differentiate between the signal of the specifically targeted clones and the nonspecific signal, because the target signal that is less than 0.1% can be easily contaminated by the nonspecific aggregation of the fluorescent chromophore and the cell debris (Fig. 2C). Usually, nonspecific aggregation does not display a clear ring pattern (Fig. 2C, enlarged window). To confirm that the rare target yeast cells were enriched enough to safely switch from MACS to FACS, plating assay, FACS analysis, and IF were performed simultaneously.

When 0.1–0.3% of the gating is required, disclosing the identity of the nonspecific FACS signal can be a critical issue for FACS gating. If the unwanted positive signal arises from the aggregation of the chromophore and cell debris alone, this problem can be solved by creating an extensive gating (e.g., 1–5% gating). However, if the positive signal is caused by the secondary antibody, then this antibody should be changed (Fig. 3 and Table 1). In the single actin clone population, using one round of FACS sorting can increase the target yeast cells about 300 times (Fig. S1C), whereas a mixed population, like our yeast surface-displayed cDNA library, shows only a 3-fold increase (Table 1). From these data, it is apparent that nonspecific binding can disrupt the specific enrichment. However, even in the case of a nonspecific signal, the higher intensity of the specifically targeted signal can overcome a weaker nonspecific signal [5]. The signal intensity of the target yeast clones increased with a gradual increase in the concentration of the primary anti-actin antibody (Fig. S1B). However, unless the positive signal intensity increases with the increased concentration of the primary antibody, alternate signal amplification methods such as biotin-conjugated albumin and chromophore-conjugated streptavidin may be considered. For the successful enrichment of the target clone, both reducing the nonspecific binding and increasing the specific binding against the unknown target antigen are important [17].

Previously, we obtained the fully human monoclonal antibody secreted from the hybridoma cell line generated by fusing B cells from the lymph nodes of liver cancer patients and partner cells. This method was established by Foung et al. [8] for generating human monoclonal antibodies for HCV treatment. These human antibodies were screened by their selective binding to the surface of the liver cancer cells. Although the antibodies used in our study are also tumor-specific, it was necessary to identify their antigens. Some of these were identified by immunoprecipitation; however, because they were membrane proteins, it was difficult to solubilize them using detergents [11]. Thus, by using this yeast surface display library system, we constructed an antigen library for identifying unknown antigens of the fully developed human antibody [12]. The main objective of this study was to develop an efficient method for identifying low abundance unknown antigens using the yeast surface display library system. The isolated actins in this study were all low abundance muscular type actins such as α-cardiac (ACTC1), α-smooth muscle (ACTA2), and γ-smooth muscle (ACTG2). The polyclonal antibody that we used for this study was generated by chicken muscle, and it might have a preference for muscular actins such as α-actin, which is more abundant than nonmuscular isotypes such as β-actin in chicken muscle. Nevertheless, the α-actin isolated by the polyclonal anti-actin used in this study was 0.0001%, which is very low in abundance. Therefore, the results of this research support the idea that our method based on the directional random-primed
cDNA library, MACS, FACS, and IF can be effectively applied for the identification of rare and novel antigens of human therapeutic antibodies.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (NRF-2012R1A1B3001047). We are grateful to Hyun Woo Joo for the schematic illustrations.

References