Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection

Peggy Ho Faix¹, Michael A. Burg¹, Michelle Gonzales¹, Edward P. Ravey¹, Andrew Baird², and David Larocca¹

BioTechniques 36:1018-1029 (June 2004)

Phage display technologies are powerful tools for selecting binding ligands against purified molecular targets, live cells, and organ vasculature. However, the selection of natural ligands using phage display has been limited because of significant problems associated with the display of complex cDNA repertoires. Here we describe the use of cDNA fragmentation and open reading frame (ORF) selection to display a human placental cDNA library on the pIII coat protein of filamentous phage. The library was enriched for ORFs by selecting cDNA- β -lactamase fusion proteins on ampicillin, resulting in a cDNA population having 97% ORFs. The ORF-selected cDNAs were fused to pIII in the phagemid vector, pUCMG4CT-198, and the library was rescued with a pIII-deleted helper phage for multivalent display. The resulting phagemid particle library consisted of 87% ORFs, compared to only 6% ORFs when prepared without ORF selection. Western blot analysis indicated cDNA-pIII fusion protein expression in eight out of nine ORF clones tested, and seven of the ORF encoded peptides were displayed multivalently. The high level of cDNA expression obtained by ORF selection suggests that ORF-enriched phage cDNA libraries prepared by these methods will be useful as functional genomics tools for identifying natural ligands from various source tissues.

INTRODUCTION

The selection of natural ligands from cDNA libraries is a valuable strategy for identifying new therapeutics and targeting agents for various cell types (1). Phage display has the potential to complement high-throughput screening methods as a means of identifying cell-targeting ligands because of its ability to rapidly generate large diverse libraries from which ligands can be identified by reiterative selection on target cells or organ vasculature (2,3). However, while peptide display has been widely used (4,5), the display of cDNAs, with the exception of antibodies, has been limited (6–8). The display of cDNA libraries on filamentous phage has been challenging primarily because the cDNAs must be in the same reading frame as the phage coat protein and, when fused to the N terminus, must not contain in-frame stop codons that would prematurely terminate the display of the fusion protein. Moreover, cDNA library

display is further complicated because of the high degree of heterogeneity among the polypeptide sequences that must form functional fusions with the phage coat protein. Several alternative display systems have been developed to address these issues, such as fusion to the C terminus of pVI (6), indirect display on pIII using a c-fos/c-jun attachment (9), and cDNA fragmentation (10–13). Even in fragmented cDNA libraries, the vast majority of clones are still likely to contain inserts that are nonfunctional because of frame shifts, stop codons, or incorrect orientation. These inserts do not necessarily prevent phage particle assembly because transcriptional readthrough can result in functional pIII production (14). As a result, the selection of cDNAs from display libraries is often hindered by poor cDNA display and the low concentration of target cDNAs in the initial library population.

Although fragmented cDNA libraries have yielded binding ligands by af-

finity selection, improvements in cDNA display would increase sensitivity and make functional selections against complex cellular targets feasible. The development of vectors for phage-mediated gene delivery to mammalian cells has enabled functional selection of phage that display internalizing ligands from phage-transduced cells using selectable genes such as green fluorescent protein (GFP) and neomycin resistance (15,16). Another consideration for functional selection is the number of ligand copies displayed per particle. Multivalent display of cell-targeting ligands may be advantageous for selecting endocytosing ligands because it has been shown to increase internalization compared to monovalent display (17,18). We have previously described a phagemid vector system, pUCMG4, which was designed for receptor-mediated delivery of a functional GFP gene to mammalian cells (18). The resulting phagemid particles display a ligand multivalently or monovalently, depend-

¹Selective Genetics, Inc., San Diego, CA, USA and ²King's College London, London, UK

ing on the helper phage used. Multivalent particles are noninfectious because the vector lacks the N-terminal domain of pIII. However, selections can be performed with noninfectious phage by using amplification of phage DNA to recover selected multivalent phage from target cells (15). Here we describe the construction of a cDNA display library in pUCMG4-198 using cDNA fragmentation and open reading frame (ORF) selection to address current cDNA display limitations. The resulting library contained 87% ORFs and, when rescued with a gene III-deleted helper phage, a high percentage of the cDNAs were displayed multivalently. These data demonstrate the successful use of ORF selection to construct a phage cDNA display library with a high level of expressed cDNAs. The library will be useful for selecting cDNA-encoded ligands using traditional phage display methods or using DNA recovery when the cDNAs are displayed multivalently.

MATERIALS AND METHODS

Plasmid Construction

The pG3ORF selection vector was constructed using inverse PCR to eliminate the multiple cloning site (MCS) and the native signal peptide of the β -lactamase gene in pBAD/gIII (Invitrogen, Carlsbad, CA, USA). Two primers, amp^{inv} and pBAD^{inv} (see Table 1 for oligonucleotide sequences), were used in the following PCR: 1 cycle of 94°C for 5 min, 30 cycles at 94°C for 30 s, 60°C for 30 s, 72°C for 3 min, and 1 cycle of 72°C for 10 min. The linear PCR product (3.4 kb) was digested with *Xho*I and self-ligated to create pG3ORF. The pG3ORF selection vector contains a new MCS with the following restriction enzyme sites: *Nco*I, *Asc*I, *Fse*I, *Xho*I, *Not*I, and *Pst*I. The *Nco*I and *Not*I restriction endonuclease sites in the MCS were designed so that in-frame cDNA inserts in pG3ORF will remain in-frame when subcloned into the pUCMG4CT phagemid vector (18) and its derivatives, such as pUCMG4CT-198. The pUCMG4CT-198 vector is identical to pUCMG4CT, except that fusions to pIII are at amino acid 198 instead of 250.

Preparation of cDNA

A fragmented cDNA library was prepared from human placental mRNA (BD Biosciences Clontech, Palo Alto, CA, USA). A tagged random primer, NOTIRAN2, that incorporates a *Not*I site for cloning, was used for first-strand synthesis. The mRNA template (500 ng) was heated with 10 ng of NOTIRAN2 primer at 70°C for 10 min, and then chilled on ice before the addition of 200 U of SuperScript™ II reverse transcriptase (Invitrogen) in buffer containing 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 5 mM MgCl₂, 10 mM dithiothreitol (DTT), and 0.5 mM each dNTP. The reaction mixture was incubated at 25°C for 10 min, 42°C for 50 min, and 70°C for 15 min, and then purified using a QIAquick® PCR purification kit (Qiagen, Valencia, CA, USA). A second tagged random primer, NCOIRAN2, which incorporates an *Nco*I restriction enzyme site, was used for second-strand synthesis. A reaction containing first-strand cDNA, 10 ng NCOIRAN2 primer, 10 μ L of 10 $\times$ restriction enzyme buffer 2 (New England Biolabs, Beverly, MA, USA), and water to 100 μ L was boiled for 5 min and placed on ice. DNA polymerase I, Klenow fragment (20 U), and dNTPs to a final concentration of 5 mM each were added, and the reaction was incubated at 37°C for 1 h. The cDNA fragments were amplified by PCR using primers NCOIPRIM2 and NOTIPRIM2 and the Advantage™-GC2 PCR kit (BD Biosciences Clontech), following the manufacturer's protocol. The reaction conditions were 20 cycles of 94°C for 30 s, 52°C for 30 s, 72°C for 1 min, and 1 cycle at 72°C for 7 min. Purified PCR products were digested with 250 U *Nco*I and 200 U *Not*I (Roche Applied Science, Indianapolis, IN, USA) using standard conditions. The digested cDNA fragments were separated on a 1.6% agarose gel, and fragments ranging in size from 100 bp to 1.4 kb were purified by electrodialysis as previously described (19).

Selection of cDNA Inserts Enriched for ORFs

The size-selected cDNA fragments were ligated into pG3ORF at the *Nco*I

and *Not*I restriction enzyme sites, following treatment of the linearized vector with calf alkaline phosphatase (Roche Applied Science). The optimal ratio of vector to inserts was determined empirically, and the T4 DNA ligase (New England Biolabs) reaction (performed for 3 h at 16°C) was scaled up to create a library of approximately 5.5×10^5 individual clones. The ligated DNA was purified using the QIAquick PCR purification kit, eluted in water, and used to transform electrocompetent XL-1 Blue MRF bacteria (Stratagene, La Jolla, CA, USA). The library was amplified on solid media on 150 mm $\times$ 150 mm plates (Nalge Nunc International, Rochester, NY, USA) containing 2 $\times$ yeast tryptone (YT), 3% agar, 60 μ g/mL ampicillin, and 0.2% arabinose to induce the expression of the arabinose promoter. The ampicillin-resistant colonies were pooled, and an aliquot of approximately 3×10^7 bacteria (50 times the size of the library) was used to inoculate a 500-mL liquid culture from which plasmid DNA was prepared using a HiSpeed™ Plasmid Maxi kit (Qiagen). Purified plasmid DNA (100 μ g) was digested to completion with *Nco*I and *Not*I and purified on a 1.6% agarose gel to isolate fragments 150 bp to 1.0 kb. Individual colonies were picked for plasmid sequencing with the P119HIND3 primer, which anneals to sequences within gIII. Analysis of DNA sequences was performed with Lasergene software (DNASTAR, Madison, WI, USA), and Basic Local Alignment Search Tool (BLAST; <http://www.ncbi.nlm.nih.gov/blast/>) were used for DNA and protein sequence similarity searching.

Phagemid Particle Display Library Construction

The cDNA inserts from pG3ORF were ligated into a pUCMG4CT-198 phagemid at the *Nco*I and *Not*I sites, purified as described above, and used to transform electrocompetent ER2738 (New England Biolabs) host bacteria. The library (1×10^6 clones) was amplified on solid media (150 mm $\times$ 150 mm plates) (Nalge Nunc International) containing 2 $\times$ YT, 3% agarose, 2% glucose, and 60 μ g/mL kanamycin. The colonies were pooled and a 100-mL aliquot con-

taining enough cells to represent each member of the library approximately 1×10^4 times was used to inoculate a liter of 2× YT media containing 2% glucose and 60 µg/mL kanamycin. Phagemid particles were prepared by rescuing the phagemids with helper phage as previously described (20), except that the library was rescued with a pIII-deleted helper phage, hyperphage (Progen, Heidelberg, Germany) at a multiplicity of infection of 10:1 or wild-type R408 at a 20:1 ratio. The particles were purified as previously described (18). Titer [colony-forming units (cfu)/mL] was determined by the quantification of ethidium bromide-stained single-stranded phagemid DNA (21).

Sequencing and PCR Analysis of DNA from Phagemid Particles

Single-stranded phagemid DNA was prepared from the phagemid particle library using the QIAprep® M13 kit (Qiagen) and used as a template for primer extension reactions with oligonucleotide P119ER (19). The DNA was ligated with T4 DNA ligase and used to transform electrocompetent host ER2738 bacteria. DNA was purified from individual phagemid clones and sequenced using the P119HIND3 primer as described above. The cDNA inserts were amplified by PCR using the Platinum® *Taq* Polymerase kit (Invitrogen), following the manufacturer's instructions, and 198Forward and 198Reverse primers. The reaction conditions were for 5 min at 94°C, 20 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 30 s, and 1 cycle at 72°C for 7 min. The PCR products were analyzed on a 1% agarose gel.

Immunoblot Analysis of cDNA-pIII Fusion Proteins

Immunoblot analysis was used to detect the expression of cDNA-pIII fusion proteins using methods that have been previously described (22). Purified phagemid particles (2.5×10^8 cfu/well) were boiled in 1× Tris-glycine-sodium dodecyl sulfate (SDS) sample buffer (Invitrogen) containing 5% 2-mercaptoethanol for 5 min and separated on a 10% polyacrylamide Tris-glycine gel (Novex™; Invitrogen). Immunoblots

were probed with 40 ng/mL mouse anti-pIII antibody (MoBiTec, Göttingen, Germany), and reactive bands were visualized by incubation with a biotinylated IgG-donkey anti-mouse antibody (Jackson ImmunoResearch, West Grove, PA, USA), followed by a horseradish peroxidase (HRP)-streptavidin conjugate (Vector Laboratories, Burlingame, CA, USA). The blots were developed with chemiluminescence substrate using a ChemiGlow™ West kit (Alpha Innotech, San Leandro, CA, USA).

Library Selection

The ORF-selected human placental cDNA phagemid particle library was prepared by rescuing with wild-type helper phage. Affinity selection with an antihuman placental lactogen monoclonal antibody (hPL-37; Serotech Ltd., Oxford, UK) was performed as previously described (20). Immunotubes (Nalge Nunc International) were coated with monoclonal antibody by overnight incubation at 4°C at 20 µg/mL. Phagemid particles were added at 1×10^{11} cfu/tube in 1% bovine serum albumin (BSA) in phosphate-buffered saline (PBS). The washing and eluting of bound phage were performed as described above. Eluted phage was used to infect ER2738 host bacteria to monitor output and to make libraries for further rounds of selection.

RESULTS

Preparation of a Human Placental cDNA Fragment Library

The library construction strategy is outlined in Figure 1. Fragmented human placental cDNAs were prepared using tagged random primer amplification as previously described (13), except that

distinct restriction enzyme recognition sequence tags were used for first-strand and second-strand synthesis so that the resulting cDNA fragments could be cloned directionally into the pIII display phagemid, pUCMG4CT-198 [a derivative of pUCMG4CT (18)]. We used a supE host strain to allow some C-terminal cDNAs to be displayed by

Figure 1. Strategy for displaying cDNA fragments encoding open reading frames (ORFs). A randomly fragmented human placental cDNA library was prepared using restriction enzyme site-tagged random primers for directional cloning. The fragments were PCR-amplified and size-selected by agarose gel electrophoresis. The fragmented cDNA inserts were cloned upstream of the β -lactamase gene in pG3ORF. After selection on ampicillin, the cDNA inserts were excised from pG3ORF and inserted into pUCMG4CT-198 for N-terminal display on truncated pIII(CT198). Phagemids were rescued with a gIII-deleted helper phage (hyperphage) for multivalent display.

Table 1. Oligonucleotides Used in pG3ORF and Library Construction

Name	Sequence	Function
ampinv	5'-GGCTCGAGCGGCCGCTGCAGCT-CACCCAGAAACGCTGGTGAAAG-3'	pG3ORF construction
pBADinv	5'-GGCTCGAGGGCCGCGCCGCGCGCCC-GCCATGGTGCTATGGCTATAGAACG-3'	pG3ORF construction
NOTIRAN2	5'-GCGGCCCGCAACNNNNNNNNN-3'	First-strand synthesis
NCOIRAN2	5'-TGGCCGCCGAGAACNNNNNNNNN-3'	Second-strand synthesis
NCOIPRIM2	5'-CGCACCATGGCCGCCGAGAAC-3'	PCR
NOTIPRIM2	5'-GGAGGCTCGAGCGGCCCGCAAC-3'	PCR
P119HIND3	5'-CTTCCGGCTCGTATGTTG-3'	Sequencing
P119ER	5'-pGTGCTGCAAGGCGATTAG-3'	Primer extension
198Forward	5'-GATTACGCCAAAGCTTTGGAG-3'	PCR
198Reverse	5'-CCAGCATTGACAGGAGGTC-3'	PCR

the readthrough of TAG stop codons. As shown in Table 1, 86% of the sampled cDNA inserts ($n = 36$) encoded peptides that were fused in-frame with the pIII signal peptide but were terminated prematurely by nonsuppressed stop codons prior to the cDNA-pIII junction. The *Nco*I site is designed to encode a short five amino acid peptide (MAAEN) fused in-frame with the signal peptide, thus frame shifts at the *Nco*I site were rarely observed (1/36; 3%). The truncated in-

serts encode either 3' termini of genes or ORFs that contain one or more stop codons and are not normally translated. There were two additional cDNA inserts (2/36; 6%) that were in the same ORF as the signal peptide but were frame-shifted at their 3' ends at the junction with the pIII gene. Thus, without ORF selection, 94% of the cDNA inserts surveyed formed fusion proteins that were either truncated prematurely or out-of-frame.

The percentage of cDNA inserts

encoding continuous ORFs in the resulting phagemid particles was not significantly improved after the rescue of the phagemid clones with a pIII-deleted helper phage. Without ORF selection, the phagemid particle library displayed relatively little cDNA-pIII fusion protein and contained a high proportion of large polyphage particles (data not shown), which form when there is an insufficient amount of pIII C-terminal (CT) domain (23,24). Therefore, we developed methods to select continuous ORF inserts before the insertion of cDNA fragments into the pUCMG4CT-198 display phagemid and library production.

ORF Selection

The pG3ORF vector map is shown in Figure 2A. The fusion of cDNA fragments to the C terminus of the native pIII signal sequence and to the N terminus of β -lactamase allows expression of the cDNAs in a similar context as the cDNA-pIII display fusion because they are both exported to the periplasm. The

Figure 2. Plasmid map and linker sequence of pG3ORF. (A) The plasmid map of pG3ORF. Restriction sites in the polylinker are boxed. Amp^r, ampicillin resistance gene; pBAD, promoter sequence; araC, gene-encoding AraC regulatory protein, gIII ss, gIII signal sequence; ORI, pBR322 origin of replication. (B) The polylinker sequence between the gIII signal sequence and the ampicillin resistance gene. The predicted amino acid sequence is given above the DNA sequence. The sequences of the *Nco*I and *Not*I restrictions sites are underlined. (C) Ligation of cDNA fragments into pG3ORF. Sequences derived from oligonucleotides in library construction are in bold.

pG3ORF vector was designed so that cDNAs producing functional fusions to β -lactamase could be subcloned directly into the pUCMG4CT-198 phagemid for display as a fusion to the pIII CT domain (CT198) while maintaining the same ORF (Figure 2C). The fragmented human placental cDNAs were inserted into the pG3ORF vector, and the resulting library was grown on solid ampicillin-containing media for the selection of cDNA inserts with continuous ORFs. Sequencing of representative clones indicated that the selection for functional cDNA- β -lactamase fusions resulted in successful enrichment of ORF-encoding cDNAs to 97% ($n = 32$) (Table 1).

Display of ORF-Enriched cDNAs on pIII(CT198)

The pG3ORF-selected cDNAs were removed from the vector and inserted into pUCMG4CT-198. Sequencing of random samples of phagemid clones showed a 12-fold increase in the percentage of cDNA inserts with ORFs from 6% ($n = 36$) in the unselected phagemid library to 70% ($n = 17$) in the selected library (Table 1). Of the remaining 30% of the clones, 18% encoded polypeptides that were fused in-frame with the pIII signal-encoding sequence but were truncated by one or more TAA or TGA stop codons, and 12% were frame-shifted at the junction

with the pIII structural gene. Rescue of the phagemid population with helper phage increased the population of phagemid particles that contained continuous ORF inserts to 87% ($n = 53$). Almost half of these ORF inserts (48%) were in the ORF predicted to encode a known or hypothetical protein. The remaining ORFs were sequences that are not known to be coding: including 5'- and 3'-untranslated regions, alternate ORFs, low-complexity DNA (5 clones), minus-strand DNA (7 clones), or ribosomal DNA (1 clone).

We analyzed the cDNA fragment length in the ORF-selected phagemid particles using PCR to amplify the insert population. The PCR product contained the insert plus

an additional 150 bp, which were contributed by the vector and primer sequences. When adjusted for the additional sequence, the fragment population in the library was estimated to range from 800 to 50 bp (Figure 3A), with most of the population in the 350 to 150 bp size range, indicating a bias toward this size class of cDNA fragments. This bias was confirmed by the analysis of a random sample of 20 clones, which showed that the average insert length was approximately 250 bp (Figure 3B).

Immunoblot analysis was used to analyze the display of cDNA-pIII fusion proteins in the ORF-selected library (Figure 3C). The truncated pIII(CT198) protein from pUCMG4CT-198 phagemid particles migrated as a doublet at 30–32 kDa, while full-length wild-type pIII from helper phage migrated at approximately 60 kDa

Figure 3. Analysis of the placental cDNA phagemid particle library and individual clones. (A) Analysis of cDNA fragment size in the phagemid particle library by PCR. The single-stranded library DNA was amplified using oligonucleotide primers that flanked the cloning site. The positive control (Ctrl.) is the PCR product of a single clone containing a 450-bp insert. Molecular weights (bp) of DNA size markers (Hi-Lo™ DNA marker; Minnesota Molecular, Minneapolis, MN, USA) are shown on the left side of the gel. (B) PCR amplification of inserts from randomly selected clones. (C) Detection of cDNA-pIII(CT198) fusion proteins in the library. Lanes containing helper phage (WT) and pUCMG4CT-198 phagemid particles [pIII(CT198)] indicate the size of the full-length pIII and the C-terminal domain of pIII, respectively. Molecular weights (kDa) of prestained protein size markers are indicated on the left side of the blot. (D) Detection of fusion proteins from individual cDNA clones. The identity of each clone is listed in Table 2. Extracts from purified phagemid clones [2.5×10^8 colony-forming units (cfu)/well] were separated and immunoblotted as described in the text.

Table 2. Characterization of Placental cDNA Library Before and After ORF Selection

	Sample Size (n)	Average Length (bp)	ORF (%)	Truncated ^a (%)	3' Frame-Shift (%)	5' Frame-Shift (%)	Mitochondrial (%)
Phagemid library (no selection)	36	273	6	86	6	3	22
pG3ORF	32	207	97	3	0	0	3
Phagemid library (after selection)	17	233	70	18	12	0	12
Particle library (after selection)	53	189	87	13	0	0	2

Sequence analysis of randomly selected clones from phagemid (pUCMG4CT-198) library before open reading frame (ORF) selection and from plasmid (pG3ORF), phagemid, and particle libraries after ORF selection.

^aContains 1 nonsuppressed stop codon that is in-frame with the pIII signal sequence.

(Figure 3C). The pUCMG4CT-198 vector displays a small peptide fused to amino acid 198 of the pIII CT domain. The doublet presumably results from preparations that contain both the fusion protein migrating at 32 kDa and a degradation product, pIII(CT198), at 30 kDa. The phagemid particles prepared from the library displayed anti-pIII reactive proteins that migrated as a smear that extended 30–40 kDa, with a predominant band at approximately 38 kDa. This protein size range is consistent with the cDNA insert size as determined by PCR analysis of the phagemid library (Figure 3, A and B).

Nine ORF clones with homology to known proteins and one clone with an insert containing four in-frame stop codons (Table 2) were chosen from 53 randomly selected clones for immunoblot analysis of cDNA-pIII fusion protein expression. The presence of anti-pIII reactive proteins that migrated slower than the 30 kDa pIII(CT198) protein indicated a successful display of all or part of the ORF-encoded polypeptides as fusions with pIII(CT198) in all the ORF clones except 6A (Figure 3D). The anti-pIII reactive protein in extracts from clone 6A migrated at approximately the same molecular weight as the 30 kDa CT (198) protein, indicating that most or all of the cDNA was not displayed. More than half of the ORF clones (clones 10A, 2B, 7B, 9B, and 10B) expressed cDNA-pIII(CT198) fusion proteins that migrated near or at their predicted molecular weights of 37–39 kDa (Figure 3D and Table 2). This size class correlates well with the predominant pIII-reactive band in the total library immunoblot (Figure 3C). This broad band likely represents simi-

larly sized cDNA-pIII fusions from a diverse cDNA fragment population as indicated by the distinct ORF sequences (Table 2). The cDNA-pIII fusion proteins from clones 5A and 4B migrated slower than their predicted molecular weights, while the cDNA-pIII fusion protein from clone 4A migrated faster. The phagemid particle extracts contained some additional lower molecular weight anti-pIII reactive proteins, indicating partial protein degradation or mistranslation of the transcript (Figure 3D). A 30-kDa anti-pIII reactive protein that comigrated with the 30-kDa pIII(CT198) was detected in the extract from clone 1A phagemid particles even though the cDNA contained four in-frame stop codons. Thus, the presence of a nonfunctional insert did not prevent pIII(CT198) production.

The pUCMG4CT-198 phagemid system rescued with a pIII-deleted helper phage results in phage particles that display only the cDNA-pIII(CT198) fusion protein and no wild-type pIII. The display of the cDNA is multivalent at about five copies of fusion protein per phage unless the cDNA is cleaved from the pIII(CT198) or skipped over during translation. Thus, the absence or relatively low levels of the 30-kDa pIII(CT198) protein in seven of the nine ORF clones suggests that these cDNA fusion proteins were displayed multivalently (Figure 3D; clones 4A, 5A, 10A, 2B, 4B, 7B, and 9B). Although the predominant pIII reactive bands (33–34 kDa) in clones 4A and 7B were smaller than the predicted full-length fusion proteins because of protein cleavage or mistranslation, the absence of the pIII(CT198) band indicates multivalent display of at least part of the cDNA-encoded peptide.

In contrast, clone 10B contained predominantly the 30-kDa anti-pIII reactive protein, indicating monovalent display of the cDNA encoded peptide. Taken together, these data indicate that all but one ORF clone tested displayed all or part of the cDNA insert, and seven of the nine ORF cDNA-pIII fusions were displayed multivalently.

Affinity Selection of cDNAs

The ability to select a cDNA clone from the library was assessed by affinity selection using a monoclonal antibody against human placental lactogen. After two rounds of biopanning using an infectious library (rescued with wild-type helper phage), all of the clones that were sampled ($n = 8$) had homology with human placental lactogen (GenBank® accession no. NM020991). Of the eight clones, six contained continuous ORFs encoding nearly full-length placental lactogen (amino acid 30 through 216). The ORFs also included the TAG stop codon, which is suppressed in the host strain, allowing translational fusion with pIII. Two distinct alleles (each one recovered three times) were identified as indicated by sequence polymorphisms in the 3' end. The remaining two inserts were frame-shifted but may have expressed part or all of the encoded placental lactogen by translational readthrough. The larger placental lactogen clones selected by the antibody were not previously detected in a random sample of 53 clones from the unselected library. Although the epitope of the anti-placental lactogen monoclonal antibody has not been reported, the lack of shorter placental lactogen cDNA inserts in the antibody-selected population suggests that the epitope may be conformational. These results demonstrate the feasibility of selecting cDNAs from ORF-enriched libraries prepared using the methods described here.

DISCUSSION

We have used cDNA fragmentation, directional cloning, and ORF selec-

Table 3. Identity of Placental cDNAs from the ORF-Enriched Library

Clone	Length	Identity	GenBank Accession No.	Start ^b	End
1A ^a	163	BAC clone RP11-329A6	AC064859.6	59078	59215
4A	171	Dedicator of cyto-kinesis 1 (DOCK1)	D50857.1	5246	5324
5A	201	Insulin-like growth factor 2	X07868.1	1527	1704
6A	264	mRNA for nidogen-2	NM_007361.2	450	685
10A	156	Glycogen synthase kinase 3 α	BC051865.2	1885	2020
2B	138	Placental protein 5	NM_006528.2	239	360
4B	219	Similar to ssDNA-binding protein	NM_032627.1	1162	1341
7B	177	DEAD box polypeptide 17	NM_030881.1	1419	1573
9B	219	Sequestosome 1 (SQSTM1)	NM_003900.2	1040	1252
10B	183	Placental lactogen	BC002717.2	416	583

BAC, bacterial artificial chromosome; ssDNA, single-stranded DNA.
^aClone 1A contains a frame shift at the 3' end of the insert and four stop codons.
^bRegion of GenBank sequence that aligns with clone.

tain the native orientation of the fragments and therefore double the probability of obtaining inserts with continuous ORFs. Our data indicate that in the absence of ORF selection, the level of expression of cloned cDNAs in M13 libraries may be even lower than anticipated because of a bias toward cDNAs with premature stop codons (31/36; 86%). This bias may be the result of selective pressure against the overproduction of pIII protein, as suggested by previous studies (25,26). The low percentage of functional clones did not prevent particle formation; however,

tion to improve the expression of a pIII coat protein cDNA display library on fd phage and validated this approach by affinity selection of an expressed cDNA from the library. Fragmentation of the cDNA was used to increase the probability of expressing functional domains, al-

though, as in all phage display systems, the expression of such domains will be limited by a lack of posttranslational modification and the folding capacity of the bacterial host. Directional cloning was accomplished by random priming from mRNA instead of cDNA to main-

the particle library consisted primarily of large polyphage particles (data not shown), which form in the absence of pIII (24). In some cases, readthrough of out-of-frame inserts could also allow particle formation as we have observed here, confirming previous reports of

translational readthrough in the M13 system (14,27). We therefore applied ORF selection to improve the quality of the library by enriching for cDNA expression, with the goal of facilitating the identification of cDNAs by affinity and/or functional selection. The combined methods resulted in a cDNA library containing 87% ORFs with a high level of successfully displayed cDNAs.

We applied ORF selection, which has been used to create small genomic libraries in M13, to make an M13 display library from human placental cDNA. The selection of ORFs using ampicillin resistance in the pG3ORF plasmid compares favorably with previously reported ORF selection strategies, including the fusion of DNA fragments to selectable genes such as HIS3 and GFP in yeast (28,29) and antibiotic resistance genes in bacteria (29,30), which resulted in ORF-enriched cDNA libraries containing from 54% to 100% ORF inserts. In the present study, we obtained 97% ORF inserts following selection in pG3ORF. Zacchi et al. (30) fused DNA fragments from a 7.3-kb plasmid to β -lactamase and obtained enrichment of ORFs to 100% (43/43). Even with selection from a much larger and more complex population of cDNA fragments derived from human placenta, the initial selection of ORFs in pG3ORF (97%) was similar to that previously reported (30). Although the incidence of ORF clones decreased to 70% when the ORF-selected library was recloned into the pUCMG4CT-198 phagemid, it increased to 87% when the library was packaged into phagemid particles, suggesting that phagemid particle production of the ORF-selected cDNAs might overcome phagemid bias toward truncated inserts. ORF selection resulted in increased cDNA expression, which allowed the production of a multivalent library with a high percentage of cDNAs displayed at 3–5 copies per particle.

In the phagemid display system, the valency of the displayed ligand can be regulated by the type of helper phage used to rescue the particles (18). When the library is rescued with wild-type helper to produce monovalent infective phage particles, it can be selected using standard phage display methods. For example, we selected a placental lactogen cDNA from the monovalent library with an anti-placental lactogen

antibody in two rounds of affinity binding and recovery of infective phage. The library can also be rescued with pIII-deleted helper to maximize the valency of the displayed cDNAs, as we have shown by immunoblot analysis. Multivalent display may facilitate receptor-mediated internalization (17,18) and thus may be advantageous for cell-based selections. Selection of multivalent ORF-enriched libraries in the pUCMG4CT-198 phagemid will require alternative phage recovery methods because the libraries are noninfective. Toward this end, we have developed phage recovery methods that employ *in vitro* amplification of phage DNA in place of traditional amplification *in vivo* following the infection of host bacteria. For example, we used genetic selection to identify cell-internalizing phagemid particles based on the GFP expression and recovered internalized phagemid particles from GFP-positive cells by PCR amplification of phagemid DNA from which new phagemid particles were prepared (15). We have also used rolling circle amplification to amplify and recover phage DNA as a method of recovering cell-internalizing phage during the selection of peptide display phage on human prostate carcinoma cells (31). Thus, using libraries prepared as described here, it may be possible to increase the diversity of ligands that can be selected against cellular targets by varying the valency of the library and the method of phage recovery accordingly.

The degree of representation of the original mRNA repertoire in the cDNA display library will determine how useful the library will be for selecting natural polypeptides that are expressed in the source tissue (i.e., human placenta). Clearly, some filtering out of ORFs may occur because not all of the correctly fused cDNA sequences will be able to be exported to the periplasm and/or result in active β -lactamase. Moreover, additional selection bias will occur in the phagemid vector for sequences that promote host growth and phage assembly. Nevertheless, a sample of 53 cDNA inserts from the phagemid particle library showed no obvious bias except for an overrepresentation of ORFs that encoded fragments of placental lactogen (3/53; two identical and

one overlapping clone) and ORFs from a low complexity region of insulin-like growth factor II that is not known to be coding (4/53). The high incidence of these sequences may reflect their abundance in human placenta RNA or, alternatively, they may confer a selective growth advantage to the phage or host bacteria. Nearly half of the sampled ORF inserts encoded known or predicted proteins in-frame with pIII, indicating significant representation of biologically relevant polypeptides from the human placental proteome. The incidence of mitochondrial and ribosomal sequences in our initial unselected library was consistent with previous reports (13) but was reduced in the ORF-selected library, which suggested that this may be an additional benefit of ORF selection. Taken together, these data suggest that sequence diversity within the library is representative of the cDNA repertoire but not without some bias resulting from the ORF selection and display vector systems that were used. Further development of the cDNA synthesis, ORF selection, and phage display methods described here is being explored to reduce bias. The use of normalized libraries or libraries prepared from defined PCR-amplified ORFs would also provide improved representation for functional as well as affinity selection of biologically relevant ligands.

Current approaches to functional genomics largely involve the cloning and expression of the ORFs within a genome and subsequent functional analysis. Phage display is an alternative approach for gene discovery that is potentially faster and more convenient because it takes advantage of the power of selection from a large pool of gene fragments of unknown function. The cDNA display methods described here, together with functional selection, will make it feasible to probe cDNA repertoires from different cell types for novel ligands that act on a desirable target cell or organ (1,15). Such ligands will provide useful functional genomic information and serve as targeting agents and/or leads for new protein therapeutics. ORF-selected cDNA display libraries could also prove valuable for a variety of other applications, including antibody development against cellular

proteomes, the identification of immunoreactive proteins involved in cancer and autoimmune disease, and vaccine development.

ACKNOWLEDGMENTS

The authors thank Kristen Jensen-Pergakes and Emelie Amburn for their excellent technical contributions to this work. This work was supported by small business innovative research grants from the National Institutes of Health (grant nos. 2R44DK and AR57985).

REFERENCES

- Larocca, D. and A. Baird. 2001. Receptor-mediated gene transfer by phage-display vectors: applications in functional genomics and gene therapy. *Drug Discov. Today* 6:793-801.
- Brown, K.C. 2000. New approaches for cell-specific targeting: identification of cell-selective peptides from combinatorial libraries. *Curr. Opin. Chem. Biol.* 4:16-21.
- Pasqualini, R. 1999. Vascular targeting with phage peptide libraries. *Q. J. Nucl. Med.* 43:159-162.
- Smith, G.P. and V.A. Petrenko. 1997. Phage display. *Chemical Reviews* 97:391-410.
- Sidhu, S.S., H.B. Lowman, B.C. Cunningham, and J.A. Wells. 2000. Phage display for selection of novel binding peptides. *Methods Enzymol.* 328:333-363.
- Jespersen, L.S., J.H. Messens, A. De Keyser, D. Eeckhout, I. Van den Brande, Y.G. Ganssemans, M.J. Lauwereys, G.P. Vlasuk, et al. 1995. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology (NY)* 13:378-382.
- Hufton, S.E., P.T. Moerkerk, E.V. Meulemans, A. de Bruine, J.W. Arends, and H.R. Hoogenboom. 1999. Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. *J. Immunol. Methods* 231:39-51.
- Cramer, R., R. Jaussi, G. Menz, and K. Blaser. 1994. Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction. *Eur. J. Biochem.* 226:53-58.
- Cramer, R., S. Hemmann, and K. Blaser. 1996. PJuFo: a phagemid for display of cDNA libraries on phage surface suitable for selective isolation of clones expressing allergens. *Adv. Exp. Med. Biol.* 409:103-110.
- Cochrane, D., C. Webster, G. Masih, and J. McCafferty. 2000. Identification of natural ligands for SH2 domains from a phage display cDNA library. *J. Mol. Biol.* 297:89-97.
- Butteroni, C., M. De Felici, H.R. Scholer, and M. Pesce. 2000. Phage display screening reveals an association between germline-specific transcription factor oct-4 and multiple cellular proteins. *J. Mol. Biol.* 304:529-540.
- Redl, B., P. Merschak, B. Abt, and P. Wojnar. 1999. Phage display reveals a novel interaction of human tear lipocalin and thioredoxin which is relevant for ligand binding. *FEBS Lett.* 460:182-186.
- Santi, E., S. Capone, C. Mennuni, A. Lahm, A. Tramontano, A. Luzzago, and A. Nicosia. 2000. Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics. *J. Mol. Biol.* 296:497-508.
- Carcamo, J., M.W. Ravera, R. Brissette, O. Dedova, J.R. Beasley, A. Alam-Moghe, C. Wan, A. Blume et al. 1998. Unexpected frame-shifts from gene to expressed protein in a phage-displayed peptide library. *Proc. Natl. Acad. Sci. USA* 95:11146-11151.
- Kassner, P.D., M.A. Burg, A. Baird, and D. Larocca. 1999. Genetic selection of phage engineered for receptor-mediated gene transfer to mammalian cells. *Biochem. Biophys. Res. Commun.* 264:921-928.
- Legendre, D. and J. Fastrez. 2002. Construction and exploitation in model experiments of functional selection of a landscape library expressed from a phagemid. *Gene* 290:203-215.
- Becerril, B., M.A. Poul, and J.D. Marks. 1999. Toward selection of internalizing antibodies from phage libraries. *Biochem. Biophys. Res. Commun.* 255:386-393.
- Larocca, D., K. Jensen-Pergakes, M. Burg, and A. Baird. 2001. Receptor-targeted gene delivery using multivalent phagemid particles. *Mol. Ther.* 3:476-484.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. CSH Laboratory Press, Cold Spring Harbor, NY.
- Kay, B.K., J. Winter, and J. McCafferty. 1996. *Phage Display of Peptides and Proteins: A Laboratory Manual*. Academic Press, San Diego, CA.
- Rakonjac, J., G. Jovanovic, and P. Model. 1997. Filamentous phage infection-mediated gene expression: construction and propagation of the *gIII* deletion mutant helper phage R408d3. *Gene* 198:99-103.
- Larocca, D., P. Kassner, A. Witte, R. Ladner, G.F. Pierce, and A. Baird. 1999. Gene transfer to mammalian cells using genetically targeted filamentous bacteriophage. *FASEB J.* 13:727-734.
- Rakonjac, J. and P. Model. 1998. Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* 282:25-41.
- Rakonjac, J., J. Feng, and P. Model. 1999. Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J. Mol. Biol.* 289:1253-1265.
- Beekwilder, J., J. Rakonjac, M. Jongsma, and D. Bosch. 1999. A phagemid vector using the E. coli phage shock promoter facilitates phage display of toxic proteins. *Gene* 228:23-31.
- Huang, W., M. McKevitt and T. Palzkill. 2000. Use of the arabinose p(bad) promoter for tightly regulated display of proteins on bacteriophage. *Gene* 251:187-197.
- Goldman, E., M. Korus, and W. Mandeck. 2000. Efficiencies of translation in three reading frames of unusual non-ORF sequences isolated from phage display. *FASEB J.* 14:603-611.
- Holz, C., A. Lueking, L. Bovekamp, C. Gutjahr, N. Bolotina, H. Lehrach, and D.J. Cahill. 2001. A human cDNA expression library in yeast enriched for open reading frames. *Genome Res.* 11:1730-1735.
- Rombel, I.T., K.F. Sykes, S. Rayner, and S.A. Johnston. 2002. ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282:33-41.
- Zacchi, P., D. Sblattero, F. Florian, R. Marzari, and A.R. Bradbury. 2003. Selecting open reading frames from DNA. *Genome Res.* 13:980-990.
- Burg, M.A., E.P. Ravey, M. Gonzalez, A. Amburn, P.H. Faix, A. Baird, and D. Larocca. Selection of internalizing ligand display phage using rolling circle amplification for phage recovery. *DNA Cell Biol.* (In press).

Received 6 January 2004; accepted 24 March 2004.

Address correspondence to:

David Larocca
Selective Genetics, Inc.
11588 Sorrento Valley Road, Suite 21
San Diego, CA 92121, USA
e-mail: laroccad@selectivegenetics.com