US006610472B1

(12) **United States Patent**
**Zhu et al.**

(10) **Patent No.: US 6,610,472 B1**
(45) **Date of Patent: Aug. 26, 2003**

# (54) ASSEMBLY AND SCREENING OF HIGHLY COMPLEX AND FULLY HUMAN ANTIBODY REPERTOIRE IN YEAST

(75) Inventors: **Li Zhu**, Palo Alto, CA (US); **Shaobing Benjamin Hua**, Cupertino, CA (US); **James Sheridan**, Mountain View, CA (US); **Yu-Huei Lin**, Sunnyvale, CA (US)

(73) Assignee: **Genetastix Corporation**, San Jose, CA (US)

( * ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/703,399**

(22) Filed: **Oct. 31, 2000**

(51) **Int. Cl.**$^7$ ........................ **G01N 33/53**; G01N 33/569
(52) **U.S. Cl.** .................. **435/5**; 435/7.31; 435/DIG. 47; 435/7.1; 935/28
(58) **Field of Search** ........................ 435/5, 6, 7.1, 7.15, 435/7.31, DIG. 47; 935/28

## (56) References Cited

U.S. PATENT DOCUMENTS

6,358,733 B1 * 3/2002 Motwani et al. ......... 435/320.1

OTHER PUBLICATIONS

Fusco et al. (1999) Yeast 15:715–720.*

Kostrub et al. (1998) Nucleic Acids Research 26(20):4783–4784.*

Alexey V. Terskikh. et al., "Peptabody": A new type of high avidity binding protein, Proc. Natl, Acad., vol. 94, pp. 1663–1668, Mar. 1997.

Ami Aronheim et al., "Isolation of an AP–1 Repressor by a Novel Method For Detecting Protein–Protein Interactions", Molecular and Cellular Biology, vol. 17, No. 6, Jun. 1997, pp. 3094–3102.

Yehoshua C. Broder et al., "The Ras recruitment system, a novel approach to the study of protein–protein interactions" Current Biology 1998, pp. 1121–1124.

Rainer K. Brachmann et al, "Tag games in yeast: the two–hybrid system and beyond", Current Opinion in Biotechnology, 1997, pp. 561–568.

* cited by examiner

*Primary Examiner*—Andrew Wang
*Assistant Examiner*—Thomas Friend
(74) *Attorney, Agent, or Firm*—Shirley Chen; Wilson Sonsini Goodrich & Rosati

## (57) ABSTRACT

Compositions, methods, and kits are provided for efficiently generating and screening a library of highly diverse protein complexes for their ability to bind to other proteins or oligonucleotide sequences. In one aspect of the invention, a library of expression vectors is provided for expressing the library of protein complexes. The library comprises a first nucleotide sequence encoding a first polypeptide subunit; and a second nucleotide sequence encoding a second polypeptide subunit. The first and second nucleotide sequences each independently varies within the library of expression vectors. In addition, the first and second polypeptide subunit are expressed as separate proteins which self-assemble to form a protein complex, such as a double-chain antibody fragment (dcFv or Fab) and a fully assembled antibody, in cells into which the library of expression vectors are introduced. The library of expression vectors can be efficiently generated in yeast cells through homologous recombination; and the encoded proteins complexes with high binding affinity to their target molecule can be selected by high throughput screening in vivo or in vitro.

**21 Claims, 19 Drawing Sheets**

FIGURE 1A

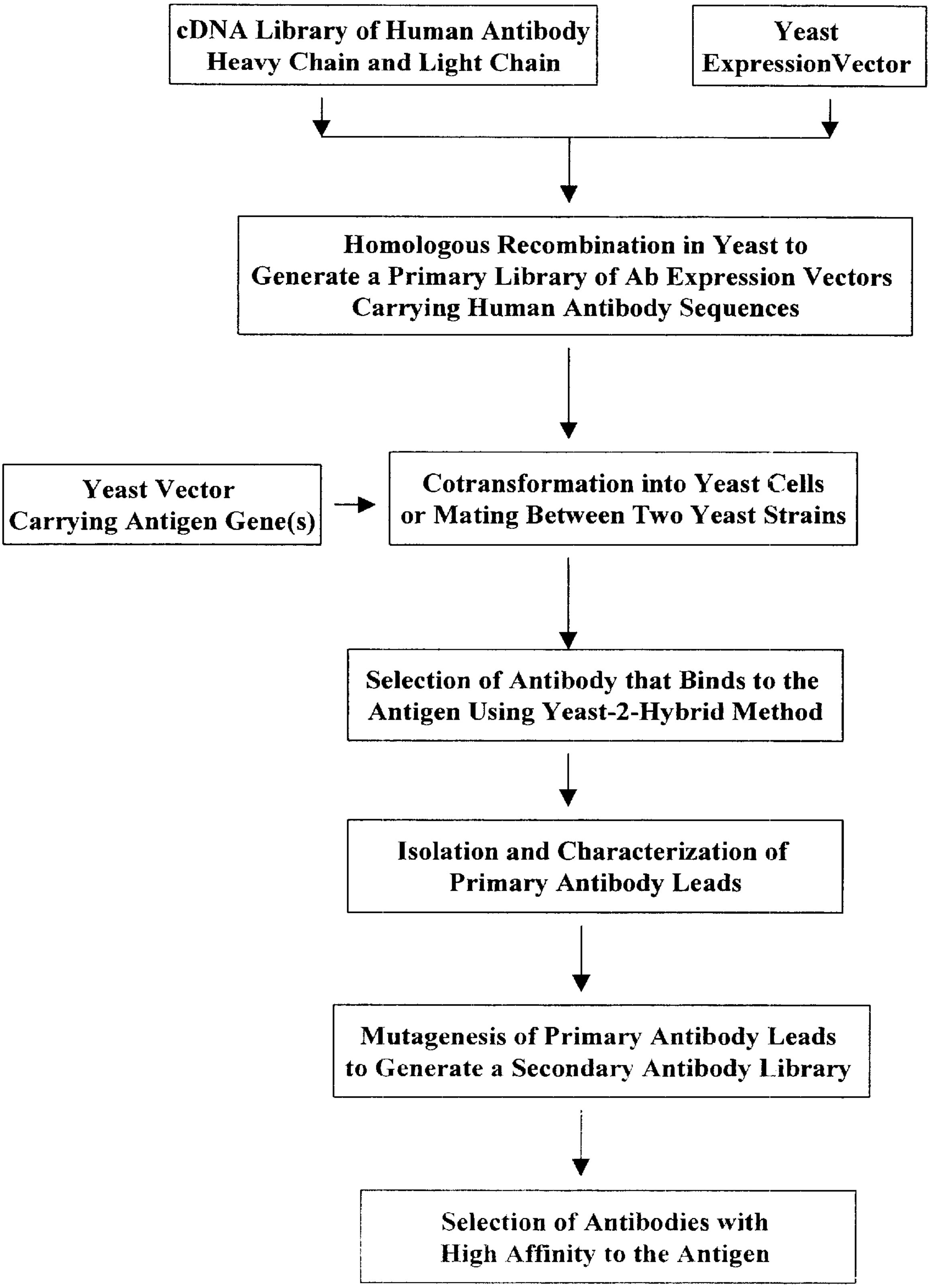

cDNA Library of Human Antibody Heavy Chain and Light Chain
Yeast ExpressionVector
Homologous Recombination in Yeast to Generate a Primary Library of Ab Expression Vectors Carrying Human Antibody Sequences
Yeast Vector Carrying Antigen Gene(s)
Cotransformation into Yeast Cells or Mating Between Two Yeast Strains
Selection of Antibody that Binds to the Antigen Using Yeast-2-Hybrid Method
Isolation and Characterization of Primary Antibody Leads
Mutagenesis of Primary Antibody Leads to Generate a Secondary Antibody Library
Selection of Antibodies with High Affinity to the Antigen

FIGURE 1B

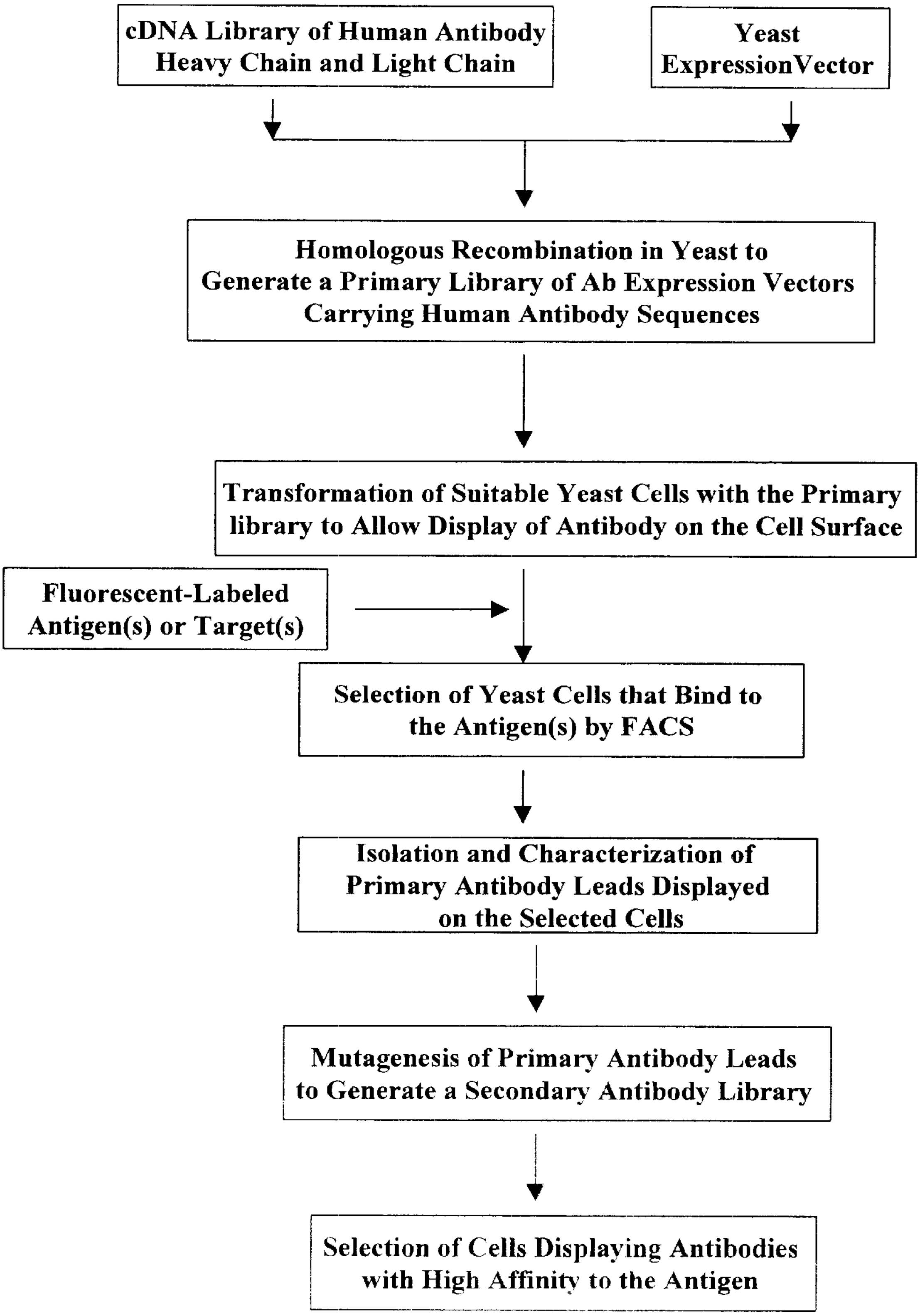

cDNA Library of Human Antibody Heavy Chain and Light Chain
Yeast ExpressionVector
Homologous Recombination in Yeast to Generate a Primary Library of Ab Expression Vectors Carrying Human Antibody Sequences
Transformation of Suitable Yeast Cells with the Primary library to Allow Display of Antibody on the Cell Surface
Fluorescent-Labeled Antigen(s) or Target(s)
Selection of Yeast Cells that Bind to the Antigen(s) by FACS
Isolation and Characterization of Primary Antibody Leads Displayed on the Selected Cells
Mutagenesis of Primary Antibody Leads to Generate a Secondary Antibody Library
Selection of Cells Displaying Antibodies with High Affinity to the Antigen

FIGURE 2

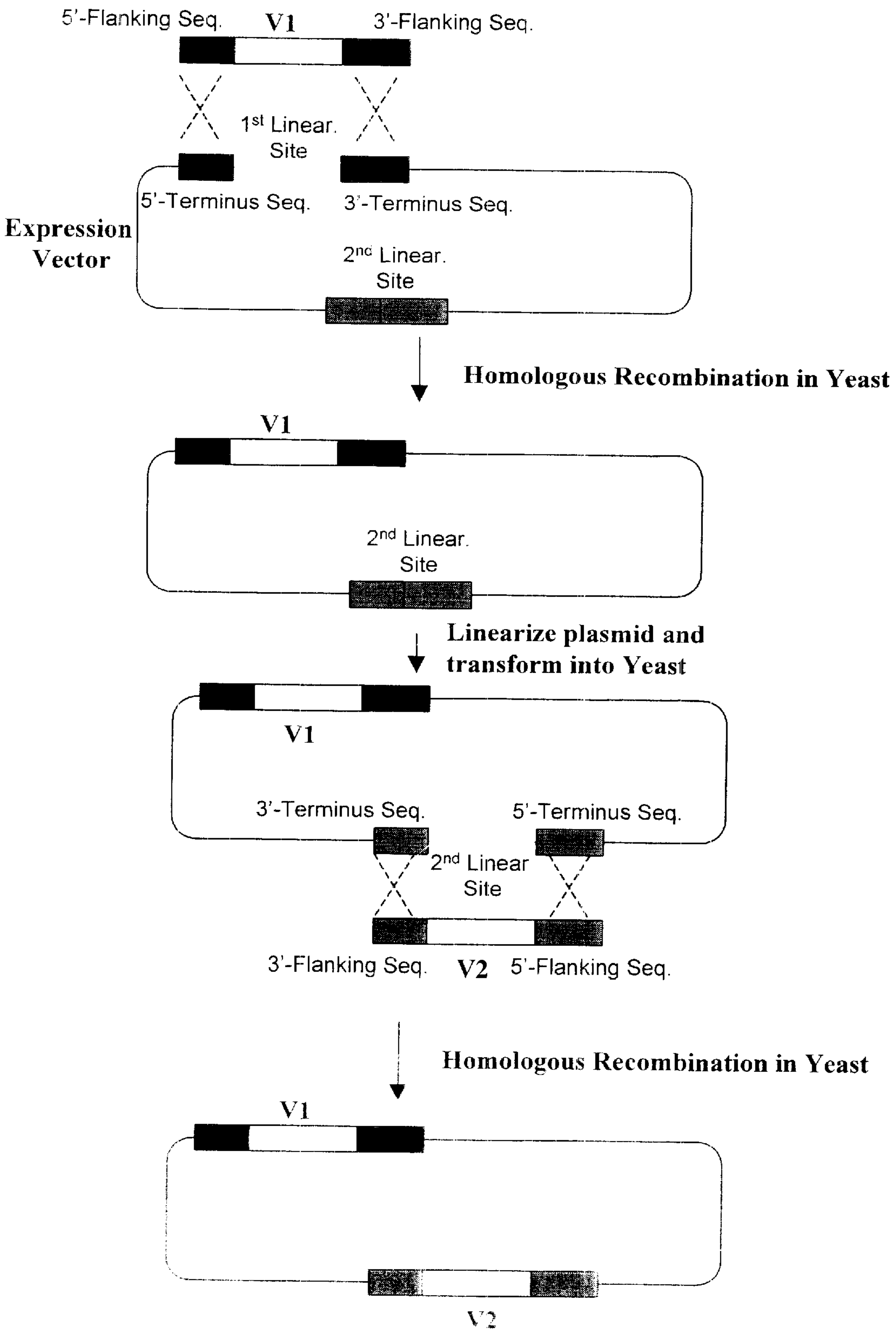

5'-Flanking Seq.
V1
3'-Flanking Seq.
1st Linear. Site
5'-Terminus Seq.
3'-Terminus Seq.
Expression Vector
2nd Linear. Site
Homologous Recombination in Yeast
V1
2nd Linear. Site
Linearize plasmid and transform into Yeast
V1
3'-Terminus Seq.
5'-Terminus Seq.
2nd Linear Site
3'-Flanking Seq.
V2
5'-Flanking Seq.
Homologous Recombination in Yeast
V1
V2

FIGURE 3

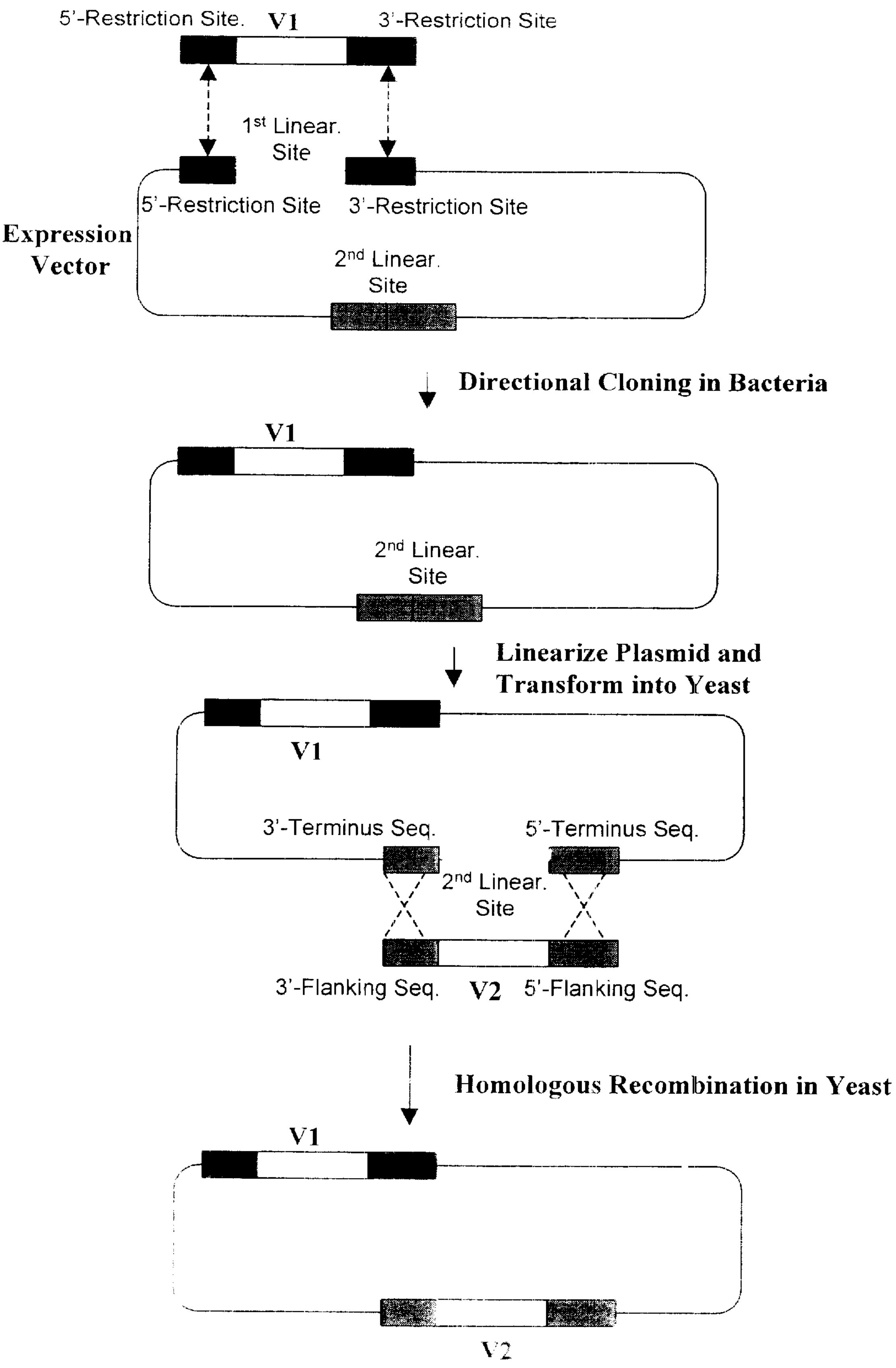
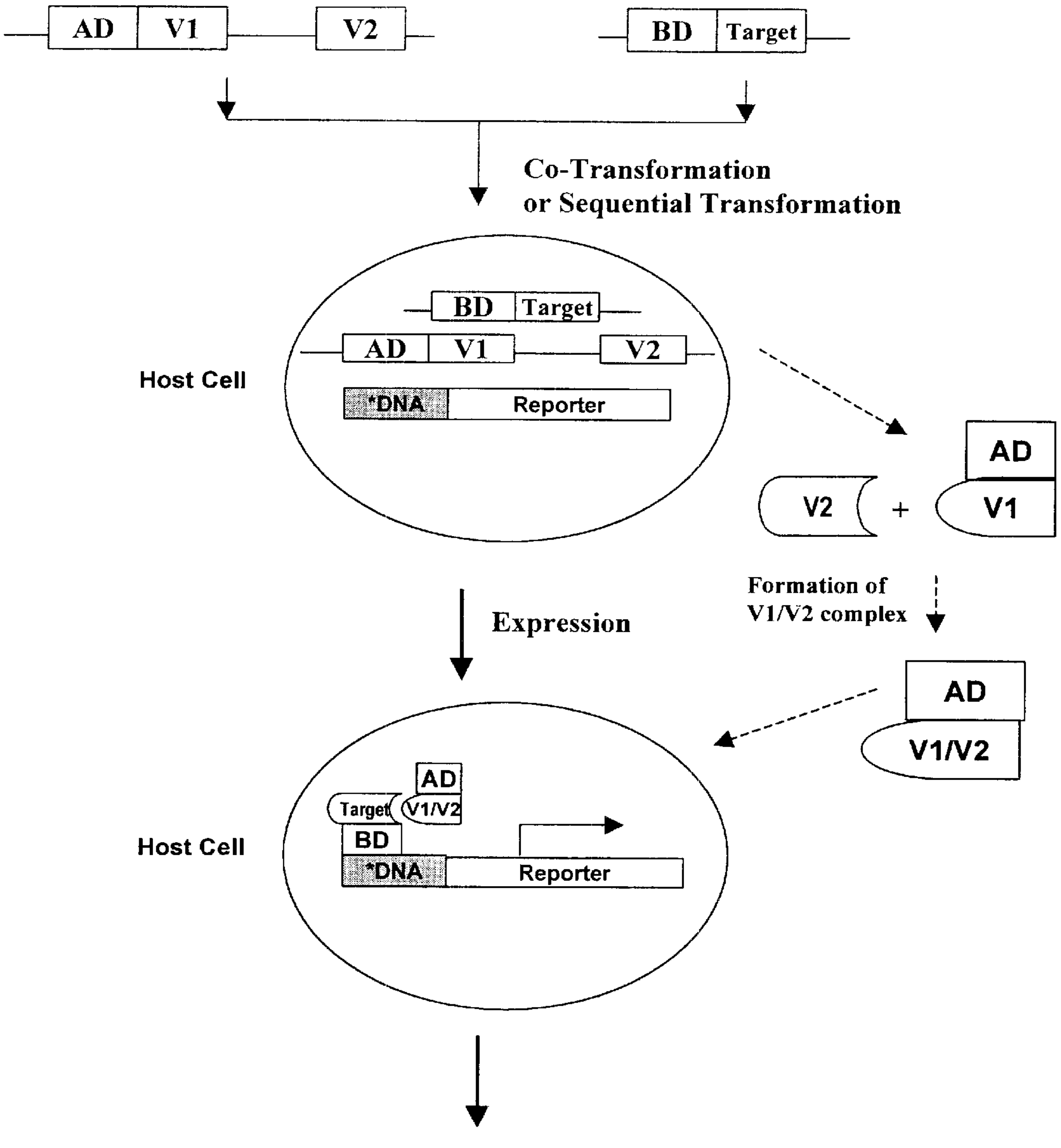
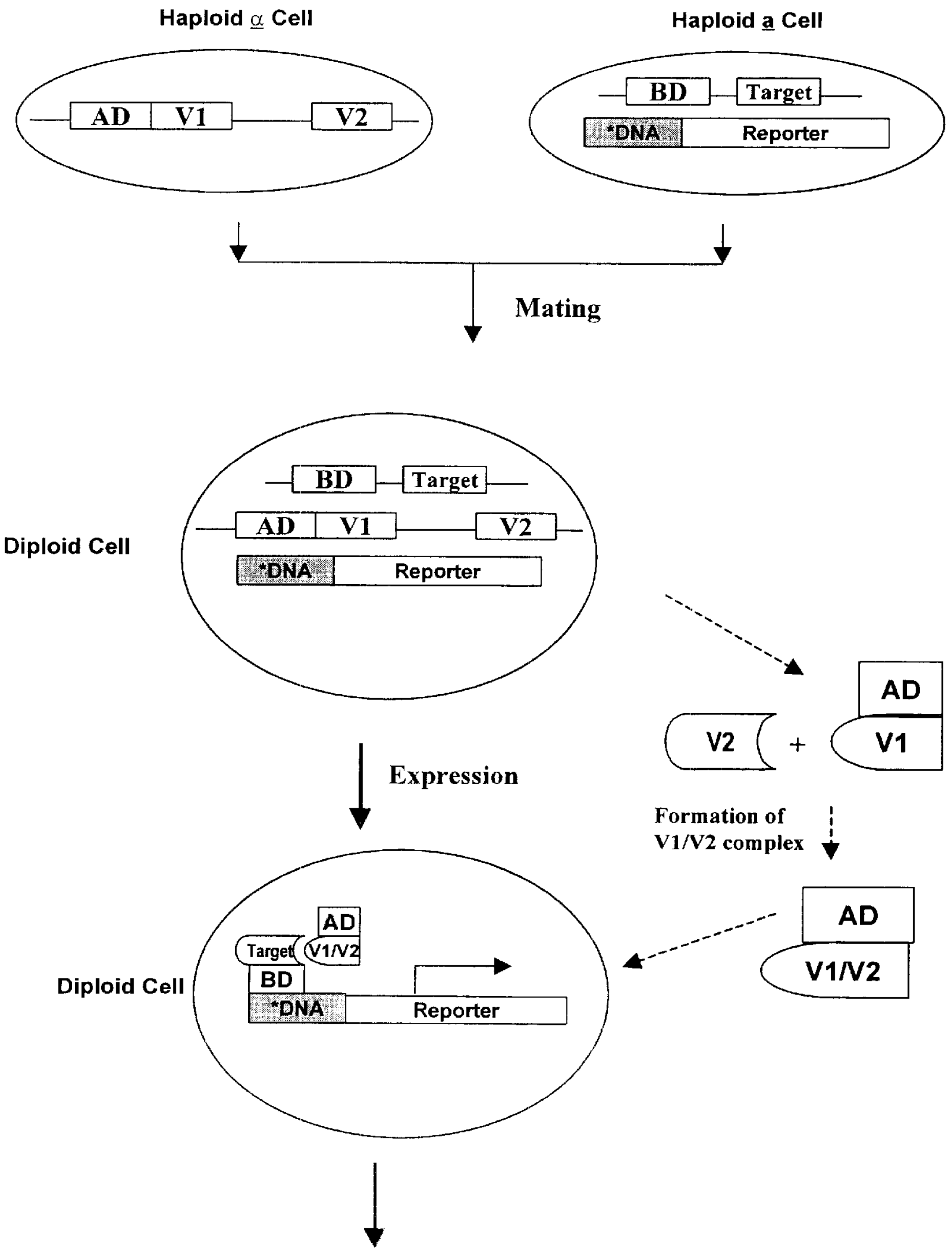

5'-Restriction Site.
V1
3'-Restriction Site
1st Linear. Site
5'-Restriction Site
3'-Restriction Site
Expression Vector
2nd Linear. Site
Directional Cloning in Bacteria
V1
2nd Linear. Site
Linearize Plasmid and Transform into Yeast
V1
3'-Terminus Seq.
5'-Terminus Seq.
2nd Linear. Site
3'-Flanking Seq.
V2
5'-Flanking Seq.
Homologous Recombination in Yeast
V1
V2

FIGURE 4



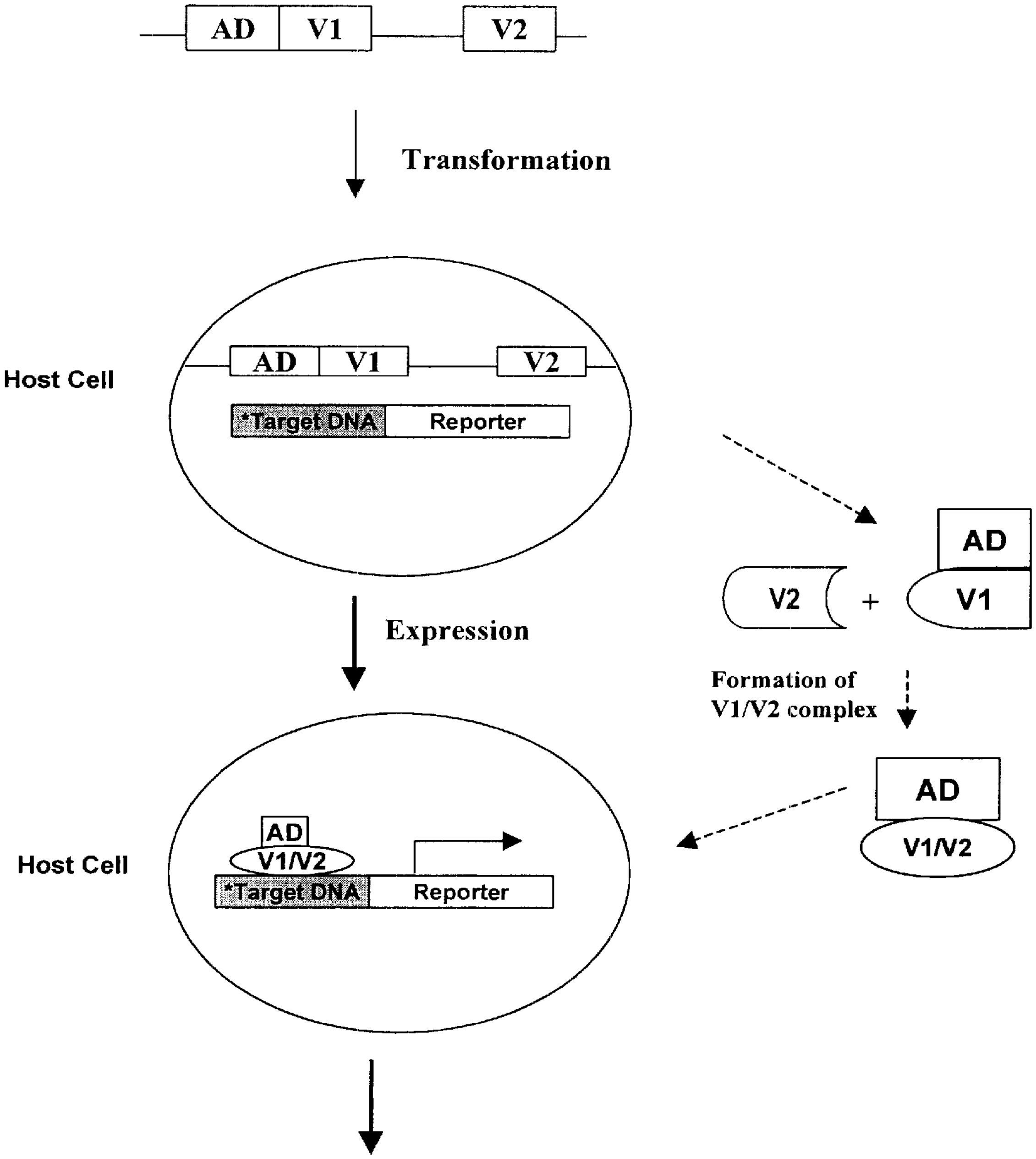
Selection of Clones Indicating
Positive Binding Between
the V1/V2 Protein Complex and the Target

FIGURE 5



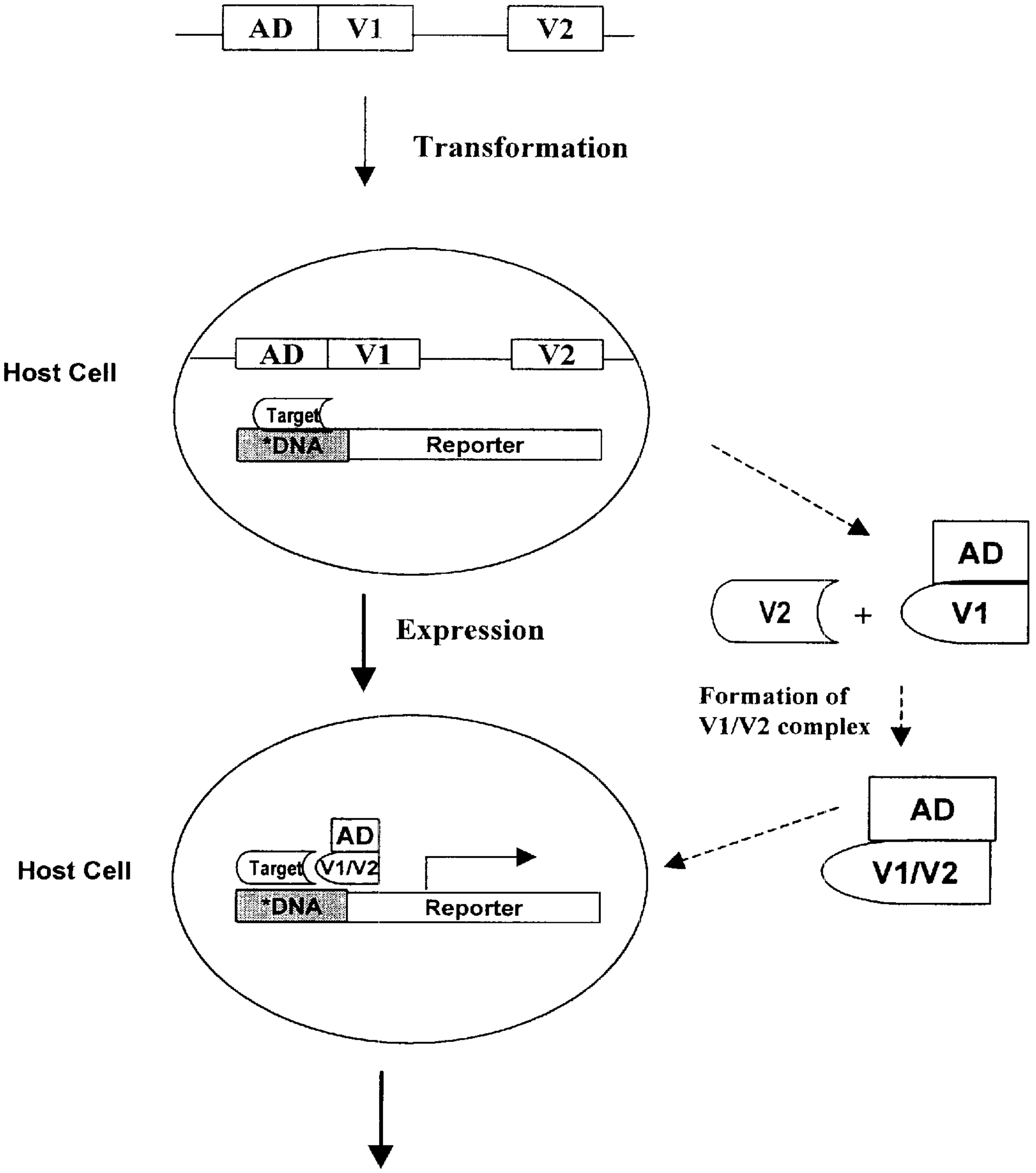
Selection of Clones Indicating
Positive Binding Between
the V1/V2 Protein Complex and the Target

FIGURE 6



Selection of Clones Indicating
Positive Binding Between
the V1/V2 Protein Complex and the Target DNA

FIGURE 7



Selection of Clones Indicating
Positive Binding Between
the V1/V2 Protein Complex and the Target Protein

FIGURE 8

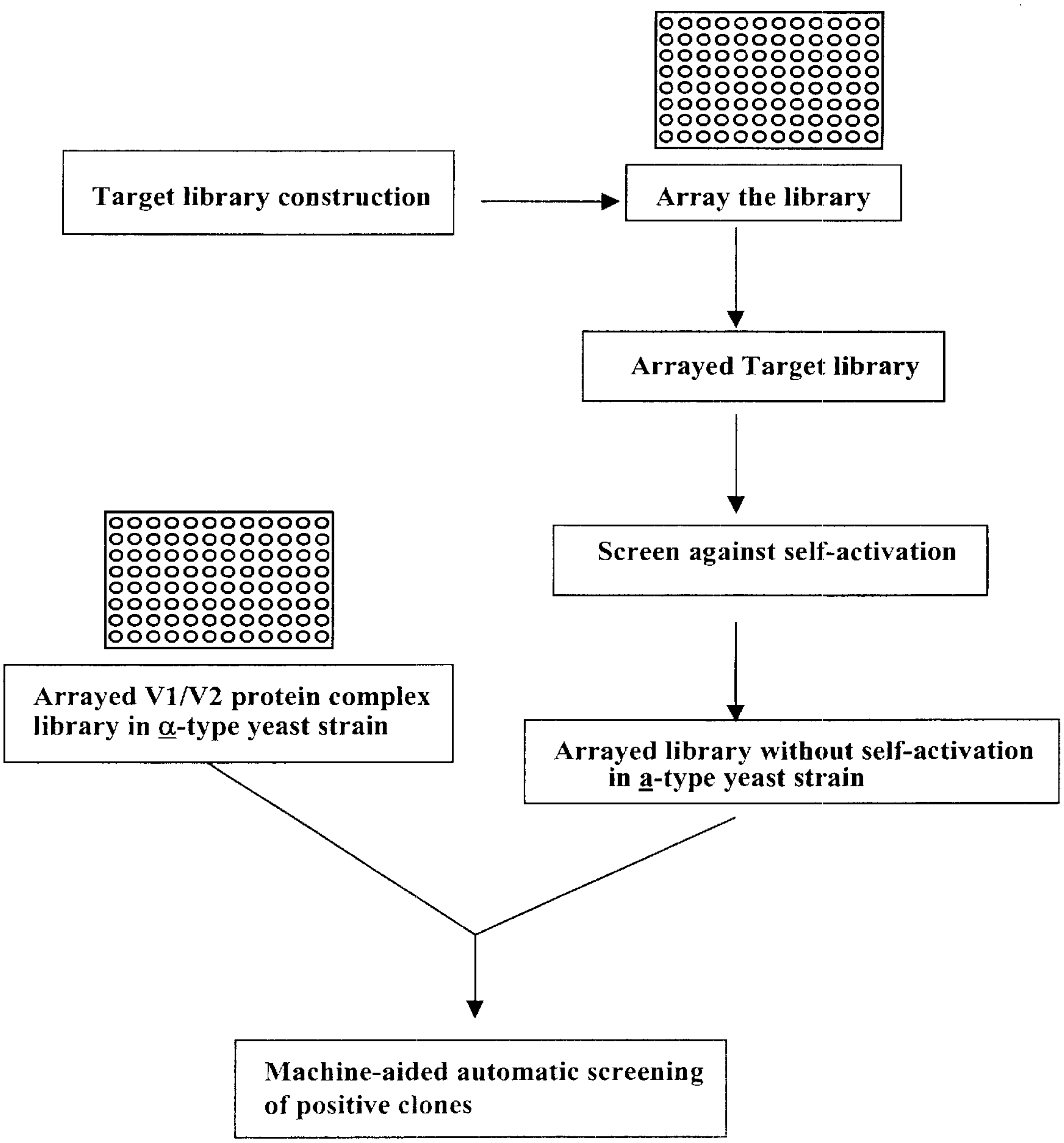

Target library construction
Array the library
Arrayed Target library
Screen against self-activation
Arrayed V1/V2 protein complex library in α-type yeast strain
Arrayed library without self-activation in a-type yeast strain
Machine-aided automatic screening of positive clones

FIGURE 9

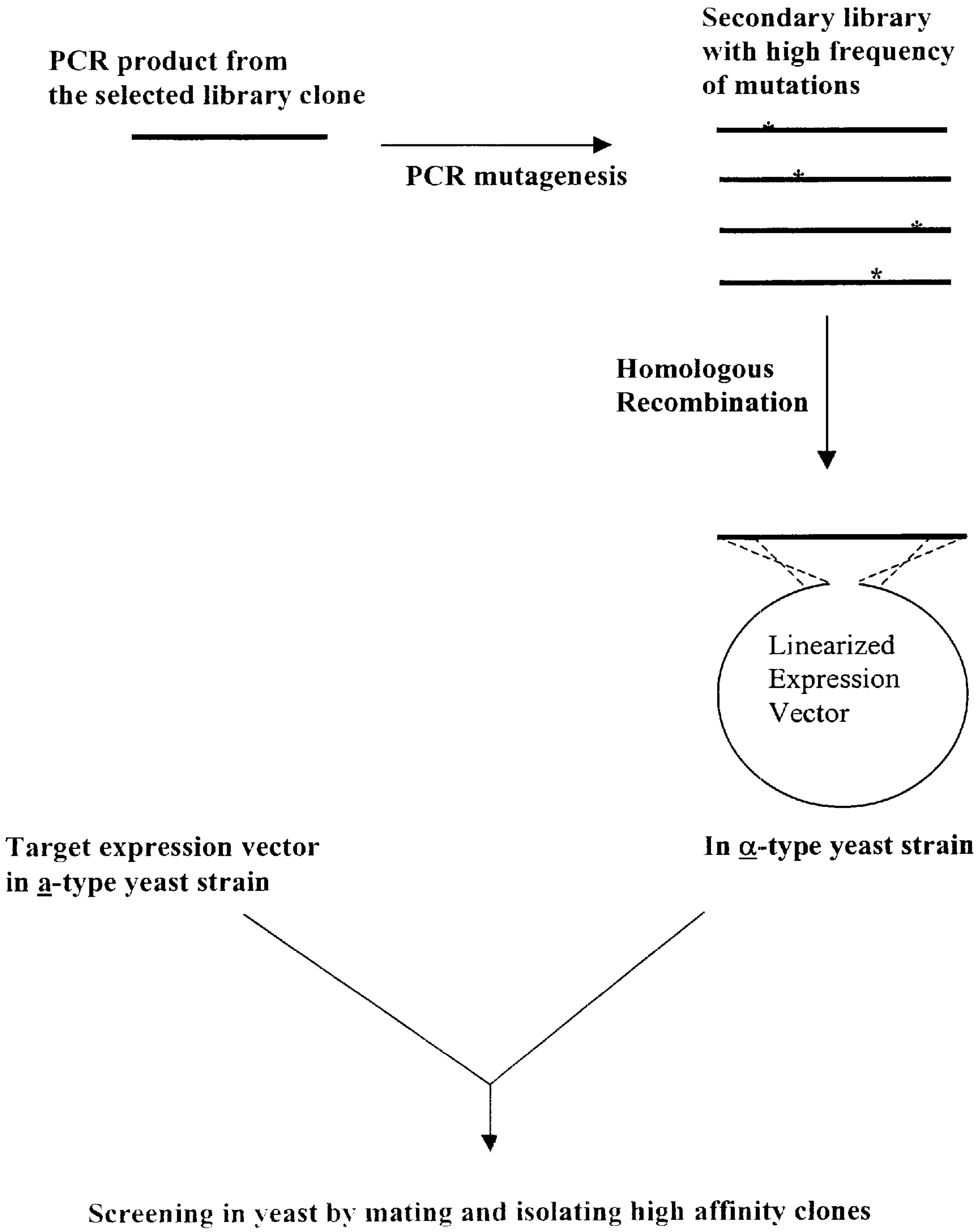
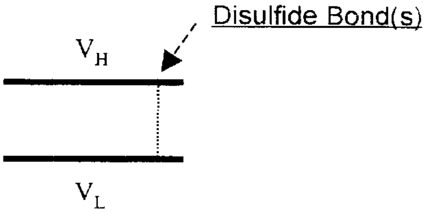
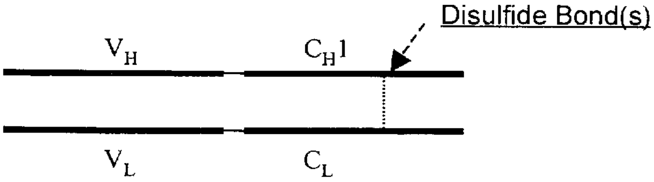

PCR product from
the selected library clone
PCR mutagenesis
Secondary library
with high frequency
of mutations
Homologous
Recombination
Linearized
Expression
Vector
In α-type yeast strain
Target expression vector
in a-type yeast strain
Screening in yeast by mating and isolating high affinity clones

FIGURE 10A

**a) dcFv**



**b) Fab**



**c) Fully assembled Antibody**

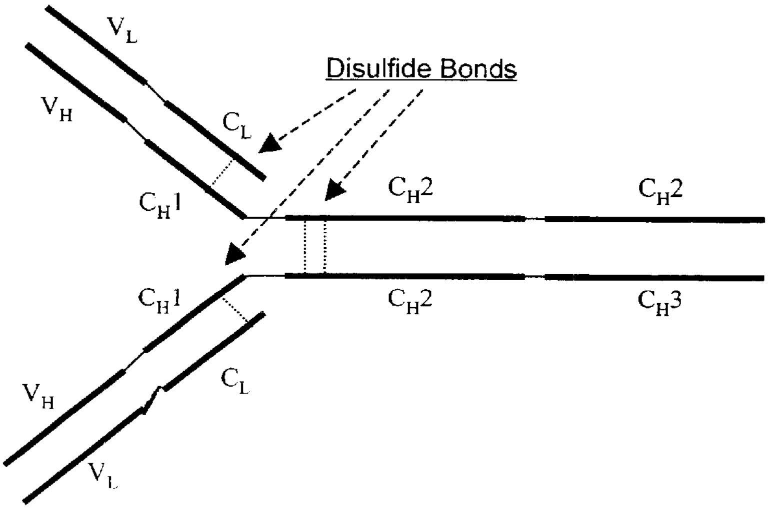
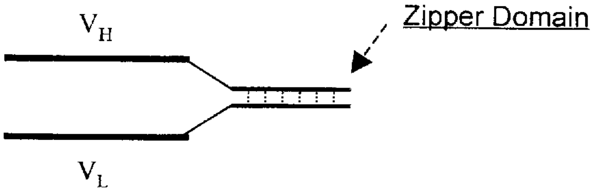
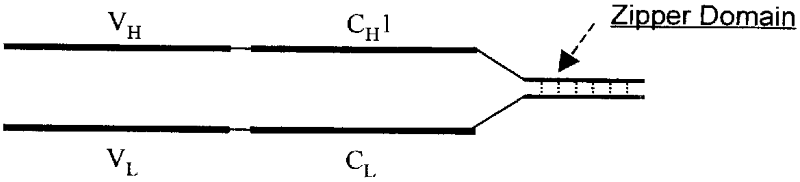


FIGURE 10B

**a) dcFv-zipper**



**b) Fab-zipper**



**c) Fully Assembled Antibody with Zipper Domain Attached**

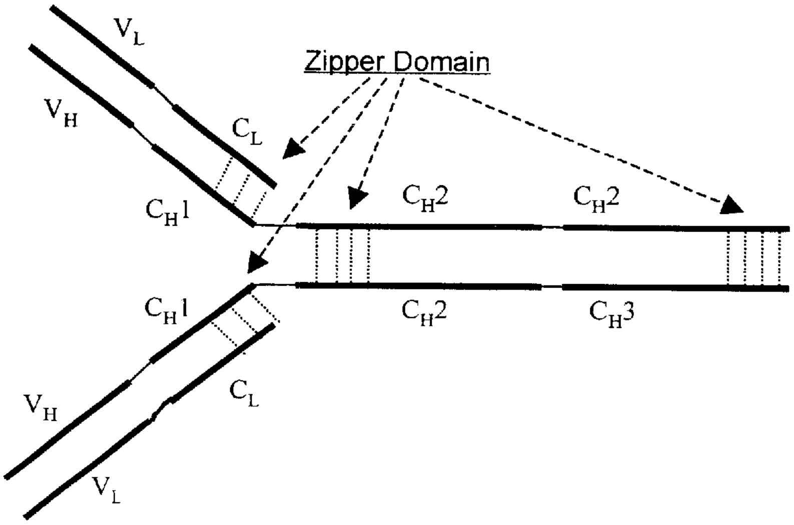


FIGURE 10C

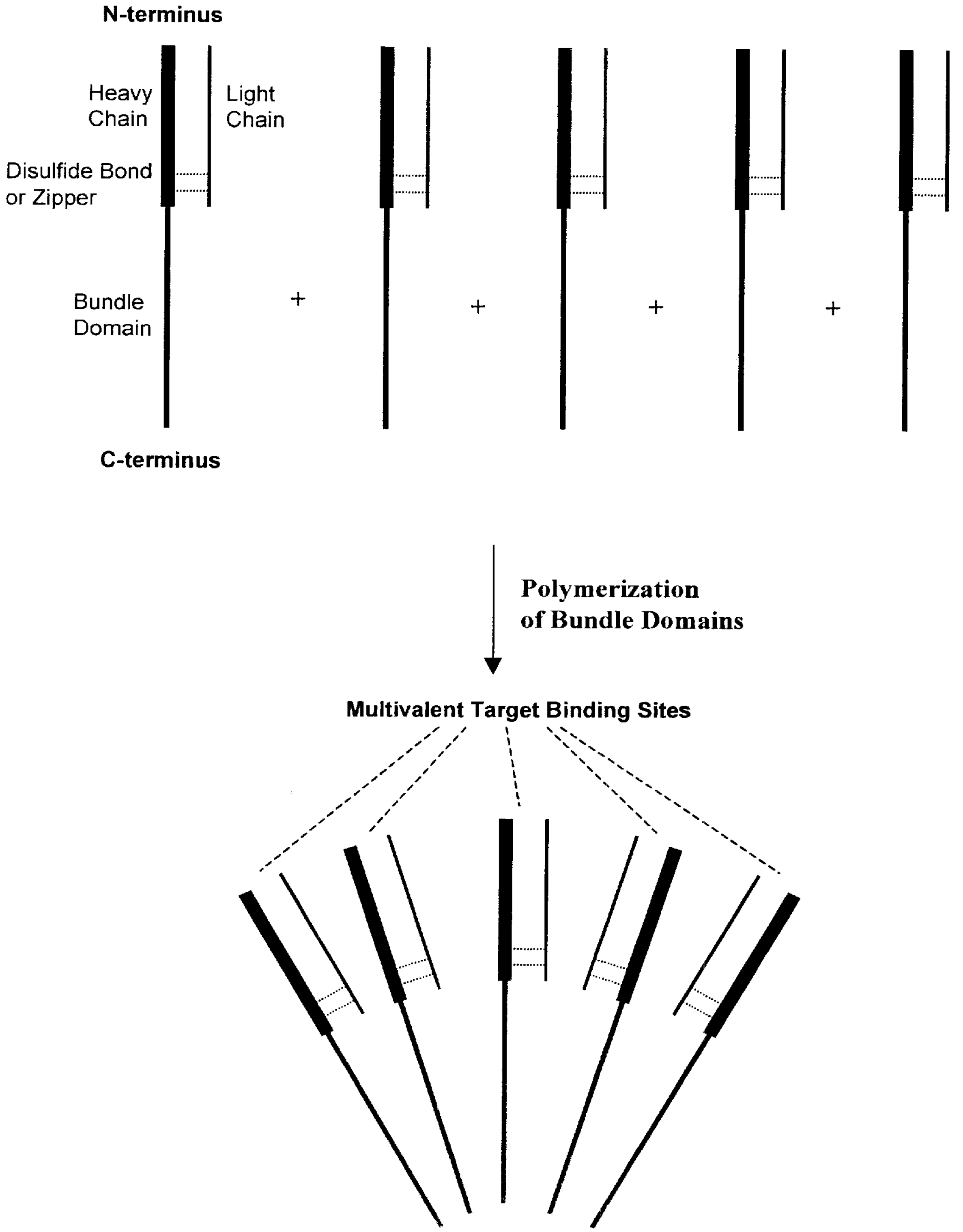

N-terminus
Heavy Chain
Light Chain
Disulfide Bond or Zipper
Bundle Domain
+
+
+
+
C-terminus
Polymerization of Bundle Domains
Multivalent Target Binding Sites

FIGURE 10D

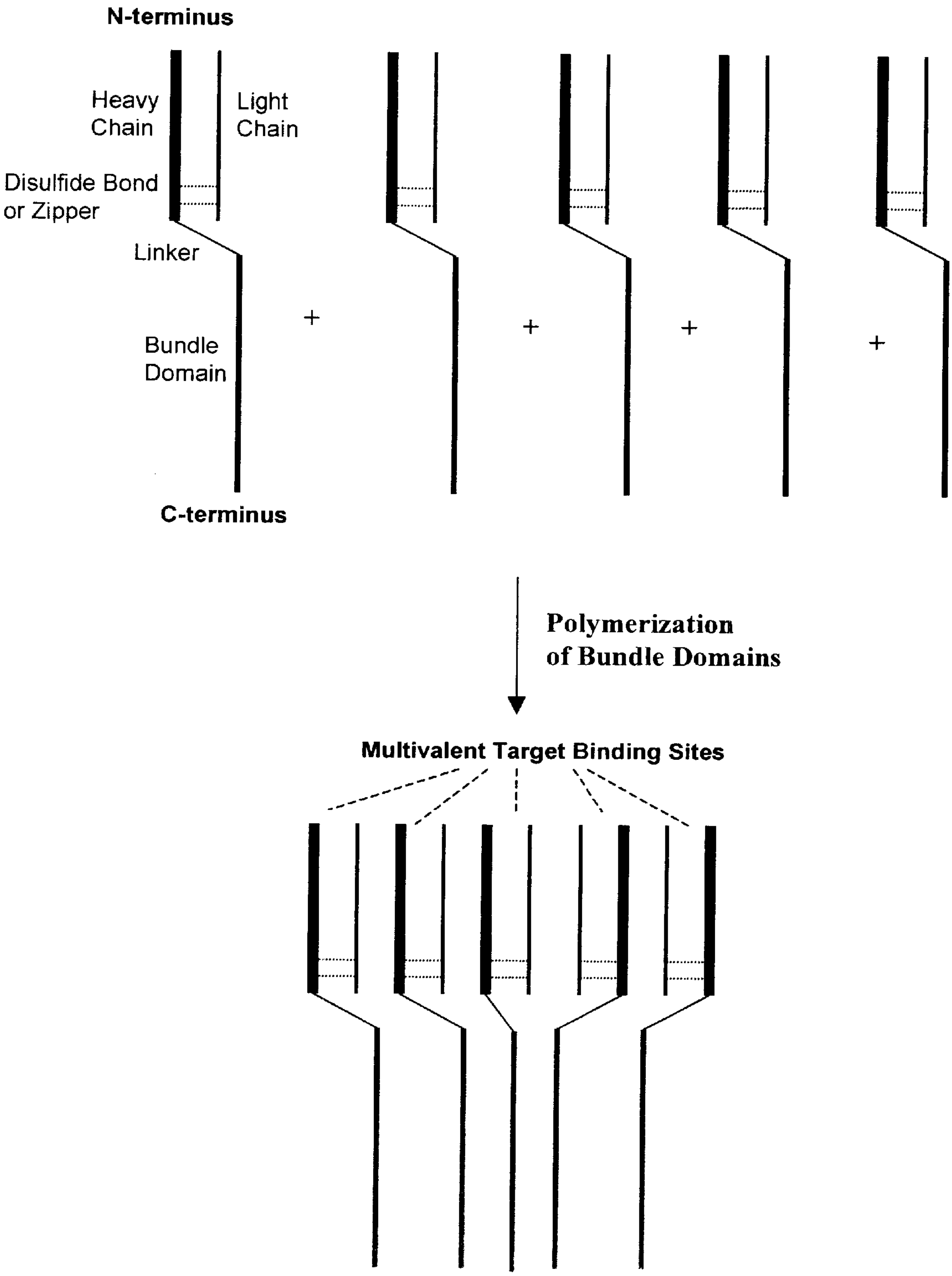

N-terminus
Heavy Chain
Light Chain
Disulfide Bond or Zipper
Linker
Bundle Domain
C-terminus
+
+
+
+
Polymerization of Bundle Domains
Multivalent Target Binding Sites

FIGURE 11

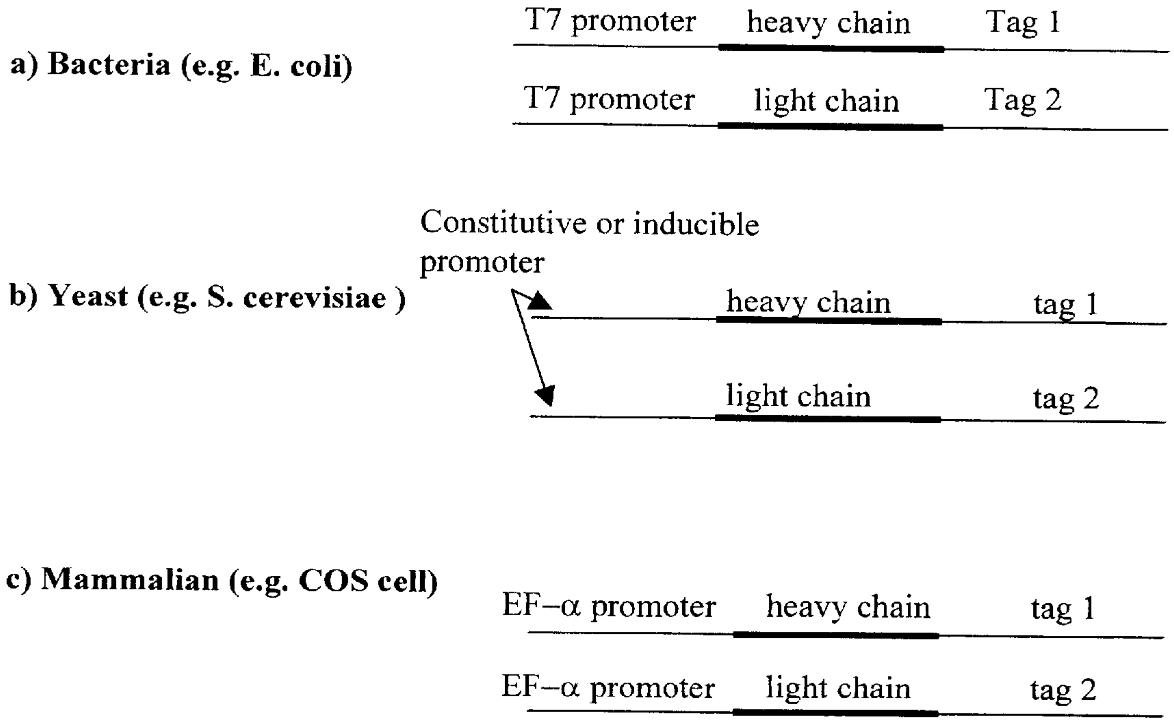

T7 promoter
heavy chain
Tag 1
a) Bacteria (e.g. E. coli)
T7 promoter
light chain
Tag 2
Constitutive or inducible promoter
b) Yeast (e.g. S. cerevisiae )
heavy chain
tag 1
light chain
tag 2
c) Mammalian (e.g. COS cell)
EF–α promoter
heavy chain
tag 1
EF–α promoter
light chain
tag 2

FIGURE 12A

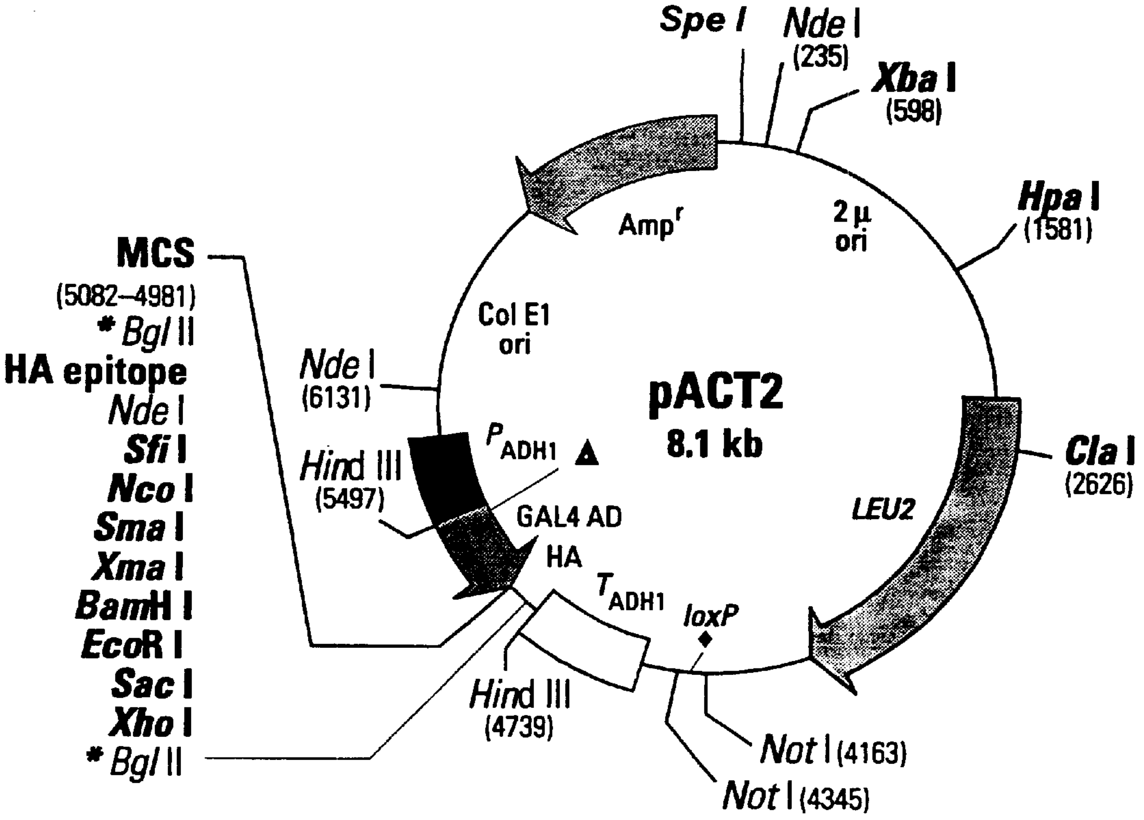

Spe I
Nde I
(235)
Xba I
(598)
Hpa I
(1581)
Amp^r
2 µ
ori
MCS
(5082–4981)
* Bgl II
HA epitope
Nde I
Sfi I
Nco I
Sma I
Xma I
BamH I
EcoR I
Sac I
Xho I
* Bgl II
Col E1
ori
Nde I
(6131)
pACT2
8.1 kb
P ADH1
Hind III
(5497)
GAL4 AD
HA
T ADH1
loxP
LEU2
Cla I
(2626)
Hind III
(4739)
Not I (4163)
Not I (4345)

**FIGURE 12B**

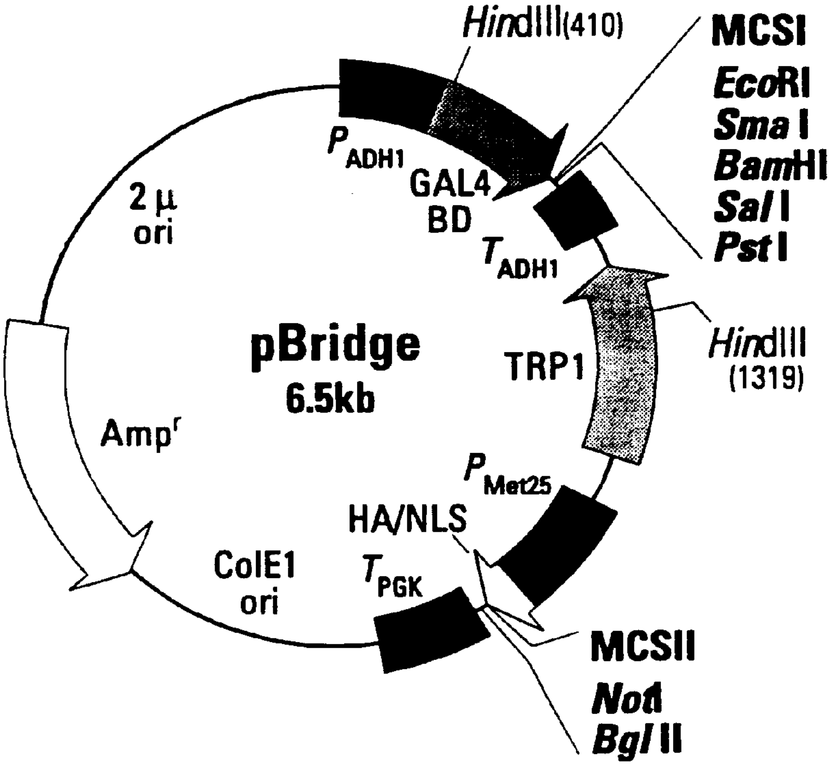

*Hin*dIII(410)
MCSI
*Eco*RI
*Sma* I
*Bam*HI
*Sal* I
*Pst* I
$P_{ADH1}$
GAL4
BD
$T_{ADH1}$
2 µ
ori
pBridge
6.5kb
TRP1
*Hin*dIII
(1319)
Amp$^r$
$P_{Met25}$
HA/NLS
ColE1
ori
$T_{PGK}$
MCSII
*Not*I
*Bgl* II

FIGURE 12C

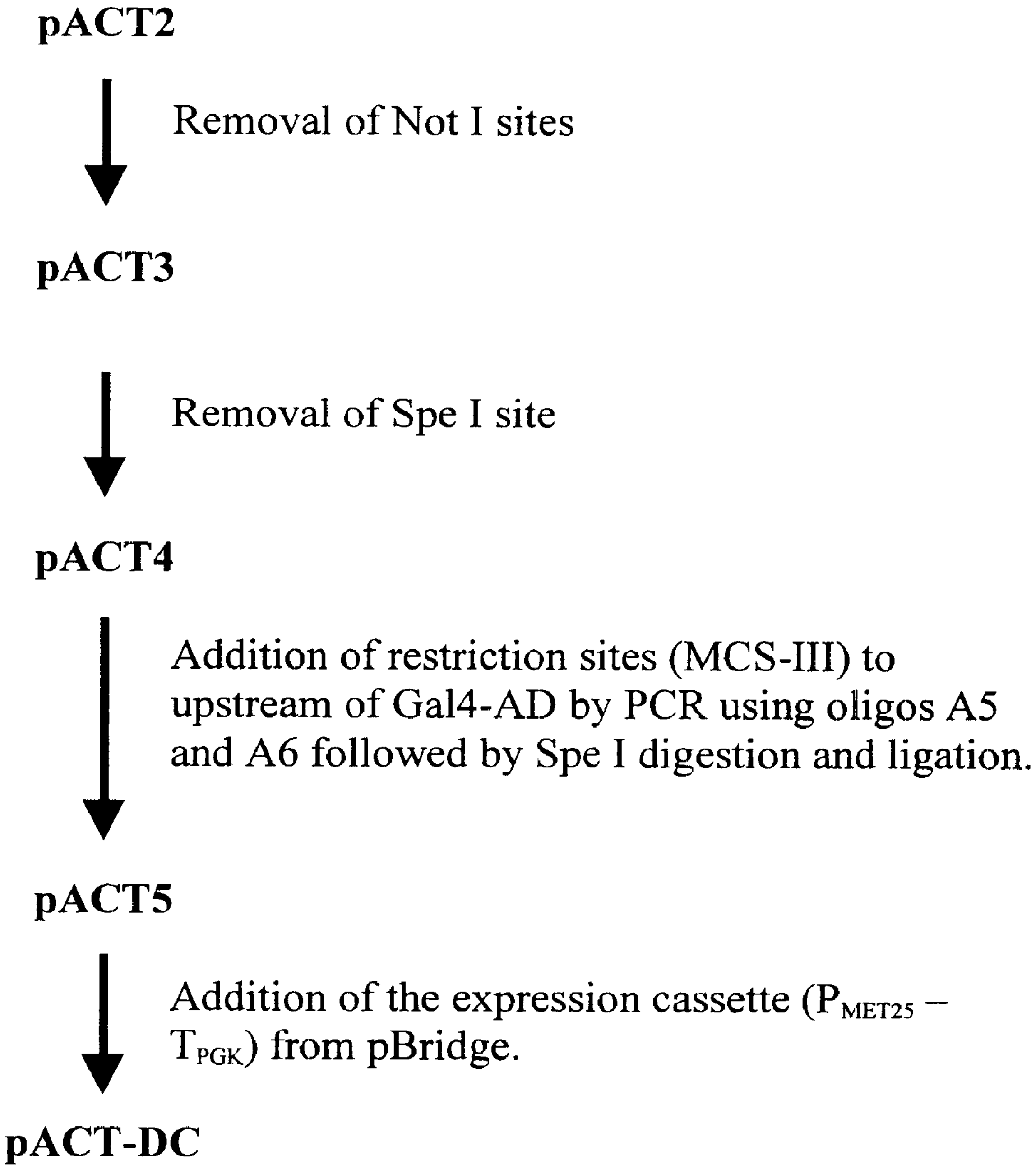
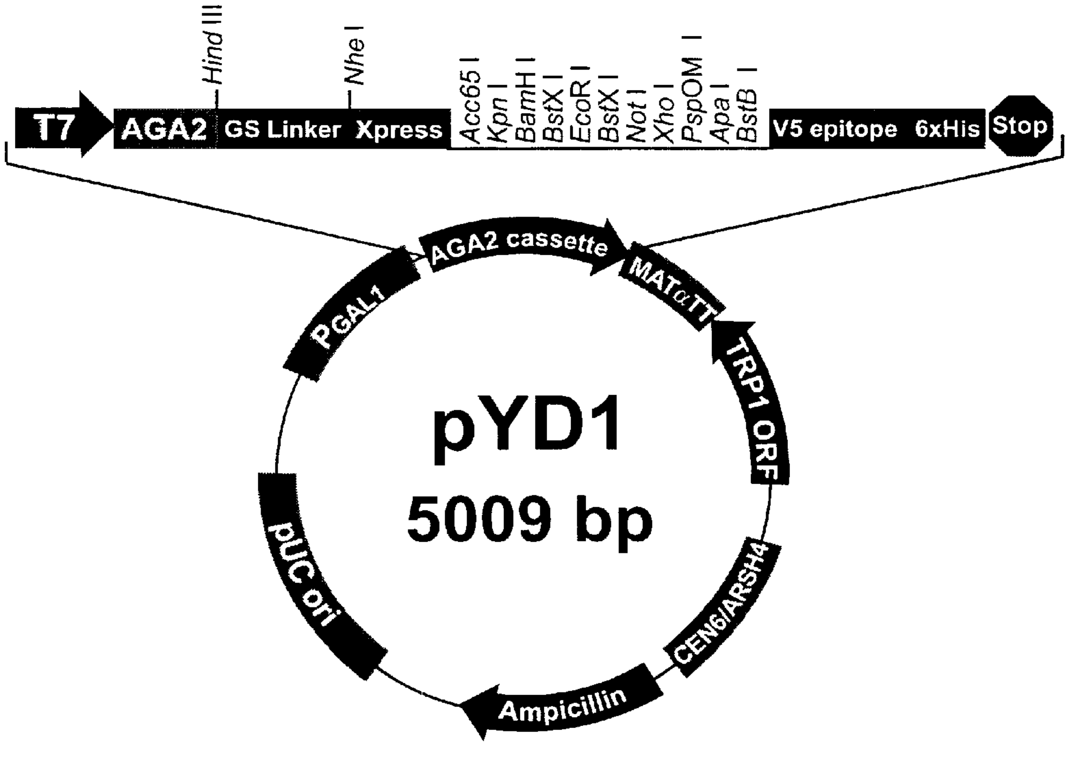

pACT2
Removal of Not I sites
pACT3
Removal of Spe I site
pACT4
Addition of restriction sites (MCS-III) to upstream of Gal4-AD by PCR using oligos A5 and A6 followed by Spe I digestion and ligation.
pACT5
Addition of the expression cassette ($P_{MET25}$ – $T_{PGK}$) from pBridge.
pACT-DC

**FIGURE 12D**


Hind III
Nhe I
T7
AGA2
GS Linker
Xpress
Acc65 I
Kpn I
BamH I
BstX I
EcoR I
BstX I
Not I
Xho I
PspOM I
Apa I
BstB I
V5 epitope
6xHis
Stop
AGA2 cassette
MATαTT
TRP1 ORF
CEN6/ARSH4
Ampicillin
pUC ori
PGAL1
pYD1
5009 bp

# ASSEMBLY AND SCREENING OF HIGHLY COMPLEX AND FULLY HUMAN ANTIBODY REPERTOIRE IN YEAST

## BACKGROUND OF THE INVENTION

### 1. Field of the Invention

This invention relates to compositions, methods and kits for generating libraries of recombinant expression vectors and using these libraries in screening of affinity-binding pairs, and, more particularly, for generating libraries of recombinant human antibodies and screening for their affinity binding with target antigens.

### 2. Description of Related Art

Antibodies are a diverse class of molecules. Delves, P. J. (1997) "Antibody production: essential techniques", New York, John Wiley & Sons, pp. 90–113. It is estimated that even in the absence of antigen stimulation a human makes at least $10^{15}$ different antibody molecules—its Permian antibody repertoire. The antigen-binding sites of many antibodies can cross-react with a variety of related but different antigenic determinants, and the Permian repertoire is apparently large enough to ensure that there will be an antigen-binding site to fit almost any potential antigenic determinant, albeit with low affinity.

Structurally, antibodies or immunoglobulins (Igs) are composed of one or more Y-shaped units. For example, immunoglobulin G (IgG) has a molecular weight of 150 kDa and consists of just one of these units. Typically, an antibody can be proteolytically cleaved by the proteinase papain into two identical Fab (fragment antigen binding) fragments and one Fc (fragment crystallizable) fragment. Each Fab contains one binding site for antigen, and the Fc portion of the antibodies mediates other aspects of the immune response.

A typical antibody contains four polypeptides-two identical copies of a heavy (H) chain and two copies of a light (L) chain, forming a general formula $H_2L_2$. Each L chain is attached to one H chain by a disulfide bond. The two H chains are also attached to each other by disulfide bonds. Papain cleaves N-terminal to the disulfide bonds that hold the H chains together. Each of the resulting Fabs consists of an entire L chain plus the N-terminal half of an H chain; the Fc is composed of the C-terminal halves of two H chains. Pepsin cleaves at numerous sites C-terminal to the inter-H disulfide bonds, resulting in the formation of a divalent fragment [F(ab')] and many small fragments of the Fc portion. IgG heavy chains contain one N-terminal variable ($V_H$) plus three C-terminal constant ($C_H1$, $C_H2$ and $C_H3$) regions. Light chains contain one N-terminal variable ($V_L$) and one C-terminal constant ($C_L$) region each. The different variable and constant regions of either heavy or light chains are of roughly equal length (about 110 amino residues per region). Fabs consist of one $V_L$, $V_H$, $C_H1$, and $C_L$ region each. The $V_L$ and $V_H$ portions contain hypervariable segments (complementarity-determining regions or CDR) that form the antibody combining site.

The $V_L$ and $V_H$ portions of a monoclonal antibody have also been linked by a synthetic linker to form a single chain protein (scFv) which retains the same specificity and affinity for the antigen as the monoclonal antibody itself. Bird, R. E., et al. (1988) "Single-chain antigen-binding proteins" Science 242:423–426. A typical scFv is a recombinant polypeptide composed of a $V_L$ tethered to a $V_H$ by a designed peptide, such as $(Gly_4\text{-}Ser)_3$ (SEQ ID NO: 80), that links the carboxyl terminus of the $V_L$ to the amino terminus of the $V_H$ sequence. The construction of the DNA sequence encoding a scFv can be achieved by using a universal primer encoding the $(Gly_4\text{-}Ser)_3$ linker by polymerase chain reactions (PCR). Lake, D. F., et al. (1995) "Generation of diverse single-chain proteins using a universal $(Gly_4\text{-}Ser)_3$ (SEQ ID NO: 80) encoding oligonucleotide" Biotechniques 19:700–702.

The mammalian immune system has evolved unique genetic mechanisms that enable it to generate an almost unlimited number of different light and heavy chains in a remarkably economical way by joining separate gene segments together before they are transcribed. For each type of Ig chain—κ light chains, λ light chains, and heavy chain—there is a separate pool of gene segments from which a single peptide chain is eventually synthesized. Each pool is on a different chromosome and usually contains a large number of gene segments encoding the V region of an Ig chain and a smaller number of gene segments encoding the C region. During B cell development a complete coding sequence for each of the two Ig chains to be synthesized is assembled by site-specific genetic recombination, bringing together the entire coding sequences for a V region and the coding sequence for a C region. In addition, the V region of a light chain is encoded by a DNA sequence assembled from two gene segments—a V gene segment and short joining or J gene segment. The V region of a heavy chain is encoded by a DNA sequence assembled from three gene segments—a V gene segment, a J gene segment and a diversity or D segment.

The large number of inherited V, J and D gene segments available for encoding Ig chains makes a substantial contribution on its own to antibody diversity, but the combinatorial joining of these segments greatly increases this contribution. Further, imprecise joining of gene segments and somatic mutations introduced during the V-D-J segment joining at the pre-B cell stage greatly increases the diversity of the V regions.

After immunization against an antigen, a mammal goes through a process known as affinity maturation to produce antibodies with higher affinity toward the antigen. Such antigen-driven somatic hypermutation fine-tunes antibody responses to a given antigen, presumably due to the accumulation of point mutations specifically in both heavy-and light-chain V region coding sequences and a selected expansion of high-affinity antibody-bearing B cell clones.

Great efforts have been made to mimic such a natural maturation of antibodies against various antigens, especially antigens associated with diseases such as autoimmune diseases, cancer, AIDS and asthma. In particular, phage display technology has been used extensively to generate large libraries of antibody fragments by exploiting the capability of bacteriophage to express and display biologically functional protein molecule on its surface. Combinatorial libraries of antibodies have been generated in bacteriophage lambda expression systems which may be screened as bacteriophage plaques or as colonies of lysogens (Huse et al. (1989) Science 246: 1275; Caton and Koprowski (1990) Proc. Natl. Acad. Sci. (U.S.A.) 87: 6450; Mullinax et al (1990) Proc. Natl. Acad. Sci. (U.S.A.) 87: 8095; Persson et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 2432). Various embodiments of bacteriophage antibody display libraries and lambda phage expression libraries have been described (Kang et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 4363; Clackson et al. (1991) Nature 352: 624; McCafferty et al. (1990) Nature 348: 552; Burton et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 10134; Hoogenboom et al. (1991) Nucleic Acids Res. 19: 4133; Chang et al. (1991) J. Immunol. 147: 3610; Breitling et al. (1991) Gene 104: 147; Marks et al. (1991) J. Mol. Biol. 222: 581; Barbas et al. (1992)

Proc. Natl. Acad. Sci. (U.S.A.) 89: 4457; Hawkins and Winter (1992) J. Immunol. 22: 867; Marks et al. (1992) Biotechnology 10: 779; Marks et al. (1992) J. Biol. Chem. 267: 16007; Lowman et al (1991) Biochemistry 30: 10832; Lerner et al. (1992) Science 258: 1313). Also see review by Rader, C. and Barbas, C. F. (1997) "Phage display of combinatorial antibody libraries" Curr. Opin. Biotechnol. 8:503–508.

Various scFv libraries displayed on bacteriophage coat proteins have been described. Marks et al. (1992) Biotechnology 10: 779; Winter G and Milstein C (1991) Nature 349: 293; Clackson et al. (1991) op.cit.; Marks et al. (1991) J. Mol. Biol. 222: 581; Chaudhary et al. (1990) Proc. Natl. Acad. Sci. (USA) 87: 1066; Chiswell et al. (1992) TIBTECH 10: 80; and Huston et al. (1988) Proc. Natl. Acad. Sci. (USA) 85: 5879.

Generally, a phage library is created by inserting a library of a random oligonucleotide or a cDNA library encoding antibody fragment such as $V_L$ and $V_H$ into gene 3 of M13 or fd phage. Each inserted gene is expressed at the N-terminal of the gene 3 product, a minor coat protein of the phage. As a result, peptide libraries that contain diverse peptides can be constructed. The phage library is then affinity screened against immobilized target molecule of interest, such as an antigen, and specifically bound phages are recovered and amplified by infection into *Escherichia coli* host cells. Typically, the target molecule of interest such as a receptor (e.g., polypeptide, carbohydrate, glycoprotein, nucleic acid) is immobilized by covalent linkage to a chromatography resin to enrich for reactive phage by affinity chromatography) and/or labeled for screen plaques or colony lifts. This procedure is called biopanning. Finally, amplified phages can be sequenced for deduction of the specific peptide sequences. During the inherent nature of phage display, the antibodies displayed on the surface of the phage may not adopt its native conformation under such in vitro selection conditions as in a mammalian system. In addition, bacteria do not readily process, assemble, or express/secrete functional antibodies.

Transgenic animals such as mice have been used to generate fully human antibodies by using the XENOMOUSE™ technology developed by companies such as Abgenix, Inc., Fremont, Calif. and Medarex, Inc. Annandale, N.J. Strains of mice are engineered by suppressing mouse antibody gene expression and functionally replacing it with human antibody gene expression. This technology utilizes the natural power of the mouse immune system in surveillance and affinity maturation to produce a broad repertoire of high affinity antibodies. However, the breeding of such strains of transgenic mice and selection of high affinity antibodies can take a long period of time. Further, the antigen against which the pool of the human antibody is selected has to be recognized by the mouse as a foreign antigen in order to mount immune response; antibodies against a target antigen that does not have immunogenicity in a mouse may not be able selected by using this technology. In addition, there may be a regulatory issue regarding the use of transgenic animals, such as transgenic goats (developed by Genzyme Transgenics, Framingham, Mass.) and chickens (developed by Geneworks, Inc., Ann Arbor, Mich.), to produce antibody, as well as safety issues concerning containment of transgenic animals infected with recombinant viral vectors.

Antibodies and antibody fragments have also been produced in transgenic plants. Plants, such as corn plants (developed by Integrated Protein Technologies, St. Louis, Mo.), are transformed with vectors carrying antibody genes, which results in stable integration of these foreign genes into the plant genome. In comparison, most microorganisms transformed with plasmids can lose the plasmids during a prolonged fermentation. Transgenic plant may be used as a cheaper means to produce antibody in large scales. However, due to the long growth circles of plants screening for antibody with high binding affinity toward a target antigen may not be efficient and feasible for high throughput screening in plants.

## SUMMARY OF THE INVENTION

The present invention provides compositions, methods, and kits for efficiently generating and screening protein complexes for their ability to bind to other proteins or oligonucleotide sequences. One feature of the present invention is the production of two or more polypeptides which self-assemble to form a protein complex in vivo. The in vivo formed protein complex is then tested in the same in vivo system for the complex's ability to bind to either a protein or a nucleotide sequence (DNA or RNA). The ability to express polypeptides, form protein complexes of those polypeptides, and screen the protein complexes all in the same intracellular system enables the present invention to screen large populations of protein complexes for binding with high throughput.

In one aspect of the present invention, compositions are provided. These compositions may be used for screening affinity-binding pairs between a tester protein complex and a target molecule in vitro or in vivo. The target molecule may be a protein, peptide, DNA, RNA, or small molecules.

In one embodiment, a library of yeast expression vectors is provided which express the protein complex to be screened. The yeast expression vectors forming the library comprise a first nucleotide sequence encoding a first polypeptide subunit; and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequences each independently varying within the library of expression vectors.

According to this embodiment, the first polypeptide subunit and the second polypeptide subunit can be expressed as separate proteins or peptides. This may be accomplished by expressing the first and second polypeptide subunits from separate promoters, or by expressing the polypeptide subunits bicistronically from the same promoter via an internal ribosomal entry site (IRES) or via a splicing donor-acceptor mechanism.

Also according to the embodiment, the yeast expression vector may be a $2\mu$ plasmid or a yc-type (centromeric) vector, preferably a yeast-bacterial shuttle vector which contains a bacterial origin of replication.

Also according to the embodiment, the first polypeptide subunit and/or the second polypeptide can be expressed as a fusion protein with a cell wall/membrane protein, such as the yeast agglutinin cell wall protein. Such a fusion allows transportation of the protein complex (e.g. antibody) formed between the first and second subunits to the cell wall/membrane, thus effectively mimicking the cell surface display of antibodies by B cells in the immune system for affinity maturation in vivo.

Alternatively, the first polypeptide subunit or the second polypeptide can be expressed as a fusion protein with nucleus protein, such as the nucleus transportation domain of a transcription factor. Such a fusion allows transportation of the protein complex (e.g. antibody) formed between the first and second subunits to the nucleus where interaction of the antibody with nuclear target(s) occurs.

In another embodiment, a library of expression vectors is provided. The expression vectors forming in the library comprise: a transcription sequence encoding an activation domain or a DNA binding domain of a transcription activator; a first nucleotide sequence encoding a first polypeptide subunit; and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequence each independently varying within the library of expression vectors.

The activation domain or the DNA binding domain of the transcription activator and the first polypeptide subunit are expressed as a single fusion protein. The second polypeptide subunit is expressed as a separate protein or peptide from the first polypeptide.

According to this embodiment, the expression vector may be a bacterial, phage, yeast, mammalian and viral expression vector, preferably a yeast expression vector, and more preferably a $2\mu$ plasmid yeast expression vector.

Also according to this embodiment, the transcription activator sequence may be located 5' relative to the first nucleotide sequence. Alternatively, the transcription activator sequence may be located 3' relative to the first nucleotide sequence.

In yet another embodiment, a library of transformed yeast cells is provided. The library of yeast cells comprises a library of yeast expression vectors. The expression vectors in the library of transformed yeast cells comprise: a transcription sequence encoding an activation domain or a DNA binding domain of a transcription activator; a first nucleotide sequence encoding a first polypeptide subunit; and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequence each independently varying within the library of expression vectors. The activation domain or the DNA binding domain of the transcription activator and the first polypeptide subunit are expressed as a single fusion protein. The second polypeptide subunit is expressed as a separate protein or peptide from the first polypeptide.

According to this embodiment, the yeast cells may be diploid yeast cells. Alternatively, the yeast cells may be haploids such as the a and α strain of yeast haploid cells.

In another aspect of the present invention, methods are provided for generating a library of yeast expression vectors that may be used for screening protein-protein or protein-DNA binding pairs.

In one embodiment, the method comprises: transforming into yeast cells a library of insert nucleotide sequences that are linear and double-stranded, and a library of linearized yeast expression vectors, each having a 5'- and 3'-terminus sequence at the site of linearization.

The linearized yeast expression vectors of the vector library comprise a first polynucleotide sequence encoding a first polypeptide subunit which varies within the vector library. The insert sequences of the insert library comprise a second nucleotide sequence encoding a second polypeptide subunit which varies within the insert library. Each of the insert sequences also comprises a 5'- and 3'-flanking sequence at the respective ends of the insert sequence. The 5'- and 3'-flanking sequences of the insert sequence are sufficiently homologous to the 5'- and 3'-terminus sequences of the linearized yeast expression vector, respectively, to enable homologous recombination to occur.

Homologous recombination occurring between the vector and the insert sequence results in inclusion of the insert sequence into the vector in the transformed yeast cells. Since the first and second nucleotide sequences vary independently within the insert library (having a complexity of $10^x$) and vector library (having a complexity of $10^y$), respectively, the complexity of the library formed as a result of homologous recombination should theoretically be $10^{x+y}$.

In this embodiment, the first polypeptide subunit and the second polypeptide subunit are expressed as separate proteins or peptides. This may be accomplished by expressing the first and second polypeptide subunits from separate promoters on the vector, or by expressing the polypeptide subunits bicistronically from the same promoter on the vector via an internal ribosomal entry site (IRES) or via a splicing donor-acceptor mechanism.

According to the embodiment, the 5'- and 3'-flanking sequences of the insert sequence is preferably between about 30–120 bp in length, more preferably between about 40–90 bp in length, and most preferably between about 45–55 bp in length.

According to the embodiment, the vector library comprising the second nucleotide sequences may be constructed by directional cloning of a library of the second nucleotide sequence inserts into a yeast expression vector in bacteria. Alternatively, the vector library may be constructed by inserting a library of the second nucleotide sequence inserts into a yeast expression vector via homologous recombination in yeast. Homologous recombination in yeast is preferred due to its higher transformation efficiency.

In yet another aspect of the present invention, methods are provided for selecting tester protein complexes capable of binding to a target peptide, protein, or DNA.

In an embodiment where the target molecule is a target peptide or protein, the method comprises:

expressing a library of tester protein complexes in yeast cells, each tester protein complex being formed between a first polypeptide subunit whose sequence varies within the library, and a second polypeptide subunit whose sequence varies within the library independently of the first polypeptide; expressing one or more target fusion proteins in the yeast cells expressing the tester proteins, each of the target fusion proteins comprising a target peptide or protein; and

selecting those yeast cells in which a reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target fusion protein.

According to this embodiment, expression of the reporter gene may be activated by a functional transcription activator being formed by the binding of the tester protein complex to the target peptide or protein as in a yeast two-hybrid system.

In a variation of the embodiment employing the yeast two-hybrid system, the tester protein forms a portion of a fusion protein with either a DNA binding domain or an activation domain of a transcriptional activator. The target protein meanwhile forms a portion of a fusion protein comprising the DNA binding domain or the activation domain of the transcriptional activator which is not present in the fusion protein comprising the tester protein. If the tester protein is able to bind to the target protein, a functional transcriptional activator is formed.

According to this variation, the step of expressing the library of tester protein complexes may include transforming a library of tester expression vectors into the yeast cells which contain a reporter construct comprising the reporter gene whose expression is under transcriptional control of a transcription activator comprising an activation domain and a DNA binding domain.

Each of the tester expression vectors comprises a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator, a first nucleotide sequence encoding the first polypeptide subunit, and a second nucleotide sequence encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The domain encoded by the first transcription sequence and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

Optionally, the step of expressing the target fusion proteins includes transforming a target expression vector into the yeast cells simultaneously or sequentially with the library of tester expression vectors. The target expression vector comprises a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide.

In another variation of the embodiment involving the yeast two-hybrid system, the steps of expressing the library of tester protein complexes and expressing the target fusion protein includes causing mating between first and second populations of haploid yeast cells of opposite mating types.

The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vector comprises a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator, a first nucleotide sequence encoding the first polypeptide subunit, and a second nucleotide sequence encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The domain encoded by the first transcription sequence and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

The second population of haploid yeast cells comprises a target expression vector. The target expression vector comprises a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide.

Either the first or second population of haploid yeast cells comprises a reporter construct comprising the reporter gene whose expression is under transcriptional control of the transcription activator.

In this variation, the haploid yeast cells of opposite mating types may preferably be α and a type strains of yeast. The mating between the first and second populations of haploid yeast cells of α and a type strains may be conducted in a rich nutritional culture medium.

Optionally, a plurality of target fusion proteins may be expressed and screened against the library of tester proteins at the same time. According to this variation, the population of haploid yeast cells comprising the expression vector encoding a target protein comprises a plurality of expression vectors encoding a plurality of target proteins. Each target protein forms a portion of a fusion protein which also comprises either an activation domain or a DNA binding domain.

According to this variation, members of the library of tester expression vectors may be arrayed as individual yeast clones in one or more multiple-well plates.

Also according to this variation, the plurality of the target expression vectors may be arrayed as individual yeast clones in one or more multiple-well plates.

Also according to this variation, mating may be based on clonal mating in which each yeast clone containing a members of the tester expression vectors is mated individually with each of the plurality of target expression vectors.

Also according to this variation, the plurality of the target expression vectors may be a library of expression vectors containing a collection of human EST clones or a collection of domain structures.

According to any of the above-described methods for selecting protein-protein binding pairs, the target fusion protein comprises an antigen associated with a disease state such as a tumor-surface antigen. Optionally, the target fusion protein may comprise a human growth factor receptor such as epidermal growth factors, transferrin, insulin-like growth factor, transforming growth factors, interleukin-1, and interleukin-2.

In another embodiment, a method is provided for screening protein-DNA binding pairs in a yeast one-hybrid system. The method comprises: expressing a library of tester protein complexes in yeast cells which contain a reporter construct comprising a reporter gene whose expression is under a transcriptional control of a target DNA sequence; and selecting the yeast cells in which the reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target DNA sequence.

In a variation of the embodiment, the step of expressing the library of tester protein complexes includes transforming into the yeast cells a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vectors comprises a transcription sequence encoding an activation domain of a transcription activator, a first nucleotide sequence encoding the first polypeptide subunit, and a second nucleotide sequence encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The transcriptional activation domain and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

In another variation of the embodiment, the step of expressing a library of tester protein complexes in yeast cells includes causing mating between a first and second populations of haploid yeast cells of opposite mating types. The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester protein complexes described above. The second population of haploid yeast cells comprises the reporter construct.

According to the variation, the haploid yeast cells of opposite mating types may preferably be α and a type strains of yeast. The mating between the first and second populations of haploid yeast cells of α and a type strains is preferably conducted in a rich nutritional culture medium.

According to any of the above-described methods for selecting protein-DNA binding pairs, the target DNA sequence in the reporter construct is preferably positioned in 2–6 tandem repeats 5' relative to the reporter gene.

The target DNA sequence in the reporter construct is preferably between about 15–75 bp in length and more preferably between about 25–55 bp in length.

In yet another embodiment, a method is provided for screening protein-protein binding pairs in a yeast one-hybrid system. The method comprises: expressing a library of tester protein complexes in yeast cells which contain a reporter construct comprising a reporter gene whose expression is under a transcriptional control of a specific DNA binding site; expressing a target protein in the yeast cells expressing the tester protein complexes, where the target protein binds to the specific DNA binding site; and selecting the yeast cells in which the reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target protein.

In a variation of the embodiment, the step of expressing the library of tester protein complexes includes transforming into the yeast cells a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vectors comprises a transcription sequence encoding an activation domain of a transcription activator, a first nucleotide sequence encoding the first polypeptide subunit, and a second nucleotide sequence encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The transcriptional activation domain and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

In another variation of the embodiment, the steps of expressing the library of tester protein complexes and expressing the target fusion protein includes causing mating between a first and second populations of haploid yeast cells of opposite mating types. The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester protein complexes described above. The second population of haploid yeast cells comprises a target expression vector comprising a target sequence encoding the target protein. Either the first or second population of haploid yeast cells comprises the reporter construct.

In any of the above-described methods for selecting tester proteins capable of binding to a target peptide, protein, or DNA, the method may further comprise isolating the tester expression vectors from the selected yeast cells; and mutagenizing the first and second nucleotide sequences in the isolated tester expression vectors to form a library of mutagenized expression vectors.

Examples of mutagenesis methods include, but are not limited to, error-prone PCR mutagenesis, site-directed mutagenesis, DNA shuffling and combinations thereof. The library of mutagenized expression vectors may be screened against the same or different target peptide, protein or DNA by following similar procedures used for screening the tester expression vectors.

In yet another aspect of the present invention, methods are provided for producing a library of assembled antibodies. Examples of the assembled antibodies include, but are not limited to, a double-chain protein complex (dcFv) formed between the variable regions of the light chain ($V_L$) and heavy chain ($V_H$), the Fab (fragment antigen-binding) fragments, and a fully assembled antibody having both the variable and constant regions of the light chain and heavy chain.

In an embodiment, the method comprises: expressing in cells a library of expression vectors. Each of the expression vectors comprises a first nucleotide sequence encoding a first polypeptide subunit comprising an antibody heavy chain variable region, a second nucleotide sequence encoding a second polypeptide subunit comprising an antibody light chain variable region. The first and second polypeptide subunits are expressed as separate proteins and self assembled to form a dcFv, Fab, or a full antibody upon interacting with each other. Also, the first and second nucleotide sequences each independently varies within the library of expression vectors to generate a library of assembled antibodies with a diversity of at least $10^7$.

According to the embodiment, the diversity of the library of assembled antibodies is preferably between $10^6$–$10^{16}$, more preferably between $10^8$–$10^{16}$, and most preferably between $10^{10}$–$10^{16}$.

The cells may be prokaryotic or eukaryotic cells, such as bacteria, yeast, insect, plant and mammalian cells. In a preferred embodiment, the cells where the library of antibodies are expressed are yeast cells.

In yet another aspect of the present invention, a kit is provided for selecting tester proteins capable of binding to a target peptide, protein, or DNA.

In an embodiment, a kit is provided which comprises: a library of tester expression vectors and a yeast cell line. Each of the tester expression vectors comprises a first transcription sequence encoding either an activation domain or a DNA binding domain of a transcription activator, a first nucleotide sequence encoding a first polypeptide subunit, and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequences each independently varying within the library of expression vectors. The first and second polypeptide subunits are expressed as separate proteins and form a protein complex upon interacting with each other. A reporter construct may be contained in the yeast cell line. The reporter construct comprises a reporter gene whose expression is under a transcriptional control of a specific DNA binding site.

Optionally, the kit may further comprise a target expression vector which comprises a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide.

In another embodiment, the kit comprises: first and second populations of haploid yeast cells of opposite mating types. The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vectors comprises a first transcription sequence encoding either an activation domain or a DNA binding domain of a transcription activator, a first nucleotide sequence encoding a first polypeptide subunit, and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequences each independently varying within the library of expression vectors. The first and second polypeptide subunits are expressed as separate proteins and form a protein complex upon interacting with each other. The second population of haploid yeast cells comprises a target expression vector. The target expression vector encodes either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide. Either the first or second population of haploid yeast cells comprises a reporter construct comprising a reporter gene whose expression is under transcriptional control of the transcription activator.

Optionally, the second population of haploid yeast cells comprises a plurality of target expression vectors. Each of the target expression vectors encodes either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide. Either the first or second population of haploid yeast cells comprises a reporter construct comprising a reporter gene whose expression is under transcriptional control of the transcription activator.

According to any of the above-described compositions, methods and kits, the diversity of the first and/or the second polypeptide subunit encoded by the first and second nucleotide sequences within the library of expression vectors is preferably between $10^3$–$10^8$, more preferably between $10^4$–$10^8$, and most preferably between $10^5$–$10^8$.

Also according to any of the above-described compositions, methods and kits, the diversity of the protein complexes encoded by the library of expression vectors may be preferably at least $10^6$–$10^{18}$, more preferably at least $10^9$–$10^{18}$, and most preferably at least $10^{10}$–$10^{18}$.

Also according to any of the above-described compositions, methods and kits, the diversities of the first and second polypeptide subunits may be each independently derived from libraries of precursor sequences that are not specifically designed for the target peptide, protein or DNA.

Also according to any of the above-described compositions, methods and kits, the diversities of the first and second polypeptide subunits optionally are not derived from one or more proteins that are known to bind to the target peptide, protein or DNA.

Also according to any of the above-described compositions, methods and kits, the diversities of the first and second polypeptide subunits optionally are not generated by mutagenizing one or more proteins that are known to bind to the target peptide, protein or DNA.

Also according to any of the above-described compositions, methods and kits, the first and the second polypeptide subunits may be subunits of a multimeric protein whose sequence varies within a library of multimeric proteins. Examples of multimeric proteins include, but are not limited to, growth factor receptors, T cell receptors, cytokine receptors, tyrosine kinase-associated receptors, and MHC proteins.

Also according to any of the above-described compositions, methods and kits, the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody heavy-chain variable region ($V_H$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**). The second nucleotide sequence comprises a coding sequence of an antibody light-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$).

Alternatively, the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody light-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$). The second nucleotide sequence comprises a coding sequence of an antibody heavy-chain variable region ($V_H$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**).

The source of the coding sequences of the antibody light-chain and heavy-chain variable and constant regions is preferably from human, non-human primate, or rodent. Optionally, the source of the coding sequences of the antibody light-chain and heavy-chain variable and constant regions may be from one or more non-immunized animals. Preferably, the source of the coding sequences of the antibody light-chain and heavy-chain variable and constant regions may be from human fetal spleen, lymph nodes or peripheral blood cells.

Also according to any of the above-described compositions, methods and kits, the first and second polypeptide subunits may each further comprise a plurality of cysteine residues, preferably 2–8 Cys residues, at or adjacent the N- or C-terminus of the polypeptide. It is believed that by adding more cysteine subunits near the termini of the subunits, the intermolecular interactions between the two subunits should be enhanced through formation of Cys-Cys disulfide bonds, thus further stabilizing the assembly of the protein complex formed by the two subunits.

Alternatively, the first and second polypeptide subunits may each further comprise a "zipper" domain at or adjacent the N- or C-terminus of the polypeptide. As used herein, a "zipper domain" refers to a protein or peptide structural motif that can interact with another "zipper domain" with a different sequence to form a hetero-polymer such as a heterodimer. It is believed that by adding a zipper domain near the termini of the subunits, the intermolecular interactions between the two subunits should be enhanced through non-covalent interactions (e.g. hydrophobic interactions), thus further stabilizing the assembly of the protein complex formed by the two subunits.

In addition, the first or the second polypeptide subunit may further comprise a "bundle" domain at or adjacent the C-terminus of the polypeptide. As used herein, a "bundle domain" refers to a protein or peptide structural motif that can interact with itself to form a homo-polymer such as a homopentamer. The bundle domains bring the protein complex together by polymerization through non-covalent interactions such as coiled-coil interactions. It is believed that polymerization of the protein complex should enhance the avidity of the protein complexes to their binding target through multivalent binding. For example, avidity of antibody of the present invention may be dramatically increased by fusing a bundle domain (e.g. the coiled-coil domain of the cartilage oligomeric matrix protein) to the C-terminus of the heavy chain via a semi-rigid linker.

Also, the first or second polypeptide subunit may further comprise a signaling domain for screening the library of the protein complexes based non-conventional two-hybrid methods such as the SRS (Sos recruitment system) and RRS (Ras Recruitment System). Examples of such signaling domain includes but are not limited to a Ras guanyl nucleotide exchange factor (e.g. human SOS factor), a membrane targeting signal such as a myristoylation sequence and farnesylation sequence, mammalian Ras lacking the carboxy-terminal domain (the CAAX box), and a ubiquitin sequence.

Also according to any of the above-described compositions, methods and kits, each of the expression vectors may further comprise a sequence encoding an affinity tag. Examples of affinity tags include, but are not limited to, polyhistidine tags, polyarginine tags, glutathione-S-transferase, maltose binding protein, staphylococcal protein A tag, and EE-epitope tags.

Also according to any of the above-described compositions, methods and kits, the transcription activator may be any transcription activator having separable DNA-binding and transcriptional activation domains. Examples of transcription activators include, but are not limited to, GAL4, GCN4, and ADR1 transcription activators.

Also according to any of the above-described compositions, methods and kits, the reporter protein encoded by the reporter gene may be any reporter genes whose expression shows a distinct genotype or phenotype in a cell. Examples of such a reporter protein include, but are not limited to, β-galactosidase, α-galactosidase, luciferase, β-glucuronidase, chloramphenicol acetyl transferase, secreted embryonic alkaline phosphatase, green fluorescent protein, enhanced blue fluorescent protein, enhanced yellow fluorescent protein, and enhanced cyan fluorescent protein.

## BRIEF DESCRIPTION OF FIGURES

FIG. **1**A illustrates a flow chart of a process that may be used in the present invention to screen for high affinity antibodies in a yeast two-hybrid system.

FIG. **1**B illustrates a flow chart of a process that may be used in the present invention to screen for high affinity antibodies displayed on the surface of yeast cells.

FIG. **2** illustrates an embodiment of a method for generating a library of expression vectors by sequentially inserting **V1** and **V2** fragments into a linearized expression vector via homologous recombination.

FIG. **3** illustrates an embodiment of a method for generating a library of expression vectors by inserting **V1** fragment into an expression vector through directional cloning in bacteria and by inserting **V2** segment into the linearized expression vector via homologous recombination in yeast.

FIG. **4** illustrates an embodiment of a method or selecting protein-protein binding pair in a two-hybrid system where the expression vectors carrying the AD and BD domains are co-transformed or sequentially transformed into yeast.

FIG. **5** illustrates an embodiment of the method for selecting protein-protein binding pairs in a two-hybrid system where the expression vectors carrying the AD and BD domains are introduced into diploid yeast cells via mating between two haploid yeast strains of opposite mating types.

FIG. **6** illustrates an embodiment of a method for selecting protein-DNA binding pair in a one-hybrid system where the expression vector carrying the AD domain is transformed into yeast.

FIG. **7** illustrates an embodiment of the method for selecting protein-protein binding pairs in a one-hybrid system where the expression vector carrying the AD domain is transformed into yeast.

FIG. **8** illustrates an embodiment of a high throughput method for selecting protein-protein binding pairs in a two-hybrid system where the library of the tester expression vectors and the library of expression vector carrying the target expression vectors are each arrayed in multi-well plates.

FIG. **9** illustrates an embodiment of a method used for mutagenesis and further screening of the clones selected from a primary screening of the tester protein complexes carried by the expression vector of the present invention.

FIG. **10**A illustrates secondary structures of double-chain variable fragments (dcFv), antibody fragments (Fab), and a fully-assembled antibody (Ab).

FIG. **10**B illustrates secondary structures of dcFv, Fab, and Ab with zipper domains attached to the heavy chain and light regions.

FIG. **10**C illustrates secondary structures of clusters of dcFv, Fab, and Ab with bundle domains attached to the heavy chain region.

FIG. **10**D illustrates secondary structures of clusters of dcFv, Fab, or Ab with bundle domains attached to the heavy chain region via a linker.

FIG. **11** illustrates examples of functional expression systems for antibody selected by using the method of the present invention.

FIG. **12**A illustrates the plasmid map of pACT2.

FIG. **12**B illustrates the plasmid map of pBridge.

FIG. **12**C depicts a method of modifying pACT2 in order to introduce another expression vector derived from pBridge into the plasmid to produce a yeast expression vector having double expression cassette (designated pACT2-DC).

FIG. **12**D illustrates the plasmid map of pYD1.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel compositions, kits and efficient methods for preparing extremely diverse libraries of tester protein complexes, and selecting from these libraries proteins with high affinity and specificity toward a target protein, peptide or DNA in vivo. One feature of the present invention is the production of two or more polypeptide in vivo which self-assemble to form a protein complex in vivo. The in vivo formed protein complex is then tested in the same in vivo system for the complex's ability to bind to either a protein or a nucleotide sequence (DNA or RNA). The ability to express polypeptides, form protein complexes of those polypeptides, and screen the protein complexes all in the same intracellular system enables the present invention to screen large populations of protein complexes for binding with high throughput.

In one particular embodiment, highly diverse libraries of human antibodies can be produced and screened against virtually any target antigen by using the compositions, kits and methods of the present invention.

The present invention provides a general method for screening these diverse libraries of tester protein complexes against a single or a plurality of target proteins or peptides.

The method comprise: expressing a library of tester protein complexes in yeast cells, each tester protein complexes being formed between a first polypeptide subunit whose sequence varies within the library, and a second polypeptide subunit whose sequence varies within the library independently of the first polypeptide; expressing one or more target fusion proteins in the yeast cells expressing the tester proteins, each of the target fusion proteins comprising a target peptide or protein; and selecting those yeast cells in which a reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target fusion protein. The library of tester protein complexes may be any multimeric proteins wherein the first and second polypeptide subunit are subunits of a multimeric protein whose sequence varies within the library of tester protein complexes.

The first and second polypeptide subunits are expressed as separate proteins by various mechanisms, such as expression from separate promoters and by expressing bicistronically from the same promoter via an internal ribosomal entry site (IRES, Paz et al. (1999) J. Biol. Chem. 274:21741–21745) or via a splicing donor-acceptor mechanism. The first and second subunits form a tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds). Since the sequences of the first and second polypeptide subunits (with a complexity of $10^x$ and $10^y$, respectively) vary independently within the library of the tester protein complexes, the complexity of the library of the protein complexes formed as a result of binding between the first and second polypeptide subunits should be $10^{x+y}$ theoretically.

In a preferred embodiment, the library of tester protein complexes is a library of antibodies where the first and second polypeptide subunits comprise antibody heavy chain and light chain sequences, respectively. Alternatively, the library of tester protein complexes is a library of antibodies where the first and second polypeptide subunits comprise antibody light chain and heavy chain sequences, respectively. The first polypeptide subunit may comprise an antibody heavy-chain variable region ($V_H$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$1, $C_H$2, and $C_H$3). The second nucleotide sequence may comprise an antibody light-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$). These light chain and heavy fragments are assembled in yeast cells to form a double-chain protein complex (dcFv) between $V_L$ and $V_H$, a Fab (fragment antigen-binding) fragments between ($V_L$+$C_L$) and ($V_H$+$C_H$1), and a fully assembled antibody formed between ($V_L$+$C_L$) and ($V_H$+$C_H$1+$C_H$2+$C_H$3).

The source of the coding sequences of the antibody light chain and heavy chain may be from humans, non-human primates, or rodents. For example, the source of the antibody coding sequences may be cDNA libraries derived from human spleen, peripheral white blood cells, fetal liver, and bone marrow.

From these libraries of antibodies, antibodies with high affinity and specificity are selected by screening against the libraries single or a plurality of target antigens and antibodies, in particular, in yeast. Compared to conventional approaches of generating monoclonal antibody by hybridoma technology and the recently developed XENOMOUSE® technology, the present invention provides a more efficient and economical way to screen for fully human antibodies in a much shorter period of time. More importantly, the production and screening of the antibody libraries can be readily adopted for high throughput screening in vivo.

The library of the tester protein complexes may be produced in vivo or in vitro by using any methods known in the art. The present invention provides a novel method for generating and screening libraries of expression vectors encoding these tester proteins against a single or a plurality of target molecules in vivo. These methods are developed by exploiting the intrinsic property of yeast—homologous recombination at an extremely high level of efficiency.

FIG. 1A shows a flow chart delineating a preferred embodiment of the above method of the present invention for generating and screening highly diverse libraries of human antibodies or antibody fragments in yeast. As illustrated in FIG. 1A, a highly complex library of human antibody is constructed in yeast cells. In particular, cDNA libraries of the heavy chain and light chain are transferred into a yeast expression vector by direct homologous recombination between the sequences encoding the heavy chain or the light chain and the yeast expression vector containing homologous recombination sites. The resulting expression vector is called Ab expression vector. This primary antibody library may reach a diversity preferably between $10^8$–$10^{14}$, more preferably between $10^{10}$–$10^{12}$, and most preferably between $10^{12}$–$10^{14}$.

These highly complex primary antibody libraries can be used in a wide variety of applications. In particular, this library is used for screening of fully human antibody against a wide variety of targets, such as a defined antigen or a library of antigens associated with diseases.

The screening for antibody-antigen interaction may be conveniently carried out in yeast by using a yeast two-hybrid method. For example, a library of Ab expression vectors are introduced into yeast cells. Expression of the antibody library in the yeast cells produces a library of assembled antibody (the tester protein complexes) with either the heavy chain or the light chain fused with an activation domain (AD) of a transcription activator. The yeast cells are also modified to express a recombinant fusion protein comprising a DNA-binding domain (BD) of the transcription activator and a target antigen. The yeast cells are also modified to express a reporter gene whose expression is under the control of a specific DNA binding site. Upon binding of the antibody from the library to the target antigen, the AD is brought into close proximity of BD, thereby causing transcriptional activation of a reporter gene downstream from a specific DNA binding site to which the BD binds. It is noted that the library of Ab expression vectors may contain the BD domain while the modified yeast cells express a fusion protein comprising the AD domain and the target antigen.

These Ab expression vectors may be introduced to yeast cells by co-transformation of diploid yeast cells or by direct mating between two strains of haploid yeast cells. For example, the Ab expression vectors containing libraries of $V_H$ and $V_L$ and an expression vector containing the target antigen can be used to co-transform diploid yeast cells in a form of yeast plasmid or bacteria-yeast shuttle plasmid. Alternatively, two strains haploid yeast cells (e.g. α- and a-type strains of yeast), each containing the Ab expression vector and the target antigen expression vector, respectively, are mated to produce a diploid yeast cell containing both expression vectors. Preferably, the haploid yeast strain containing the target antigen expression vector also contains the reporter gene positioned downstream of the specific DNA binding site.

The yeast clones containing antibodies with binding affinity to the target antigen are selected based on phenotypes of the cells or other selectable markers. The plasmids encoding these primary antibody leads can be isolated and further characterized.

Alternatively, the first polypeptide subunit and/or the second polypeptide can be expressed as a fusion protein with a cell wall/membrane protein, such as the yeast agglutinin Aga2p cell wall protein. Such a fusion allows transportation of the protein complex (e.g. antibody) formed between the first and second subunits to the cell wall/membrane, thus effectively mimicking the cell surface display of antibodies by B cells in the immune system for affinity maturation in vivo.

FIG. 1B depicts a general scheme for this alternative method of selection of antibodies displayed on the surface of yeast cells. As illustrated in FIG. 1B, the primary antibody library contains antibody variants having the heavy chain region fused to the C-terminus of a yeast agglutinin protein such as the yeast Aga2 subunit of a-agglutinin. Shusta et al. (1999) "Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency" J. Mol. Biol. 292:949–956.

Transportation of the antibody by the yeast cell wall protein allows the antibody library to be displayed on the surface of transformed yeast cells. One or more target molecules such as fluorescence-labeled antigen(s)s are added to the cells. The cells displaying antibodies that bind to the antigen(s) can be conveniently selected by using fluorescence-activated cell sorting (FACS) or by using magnetic beads to isolate these cells.

After the selection of the primary library of human antibodies by using a yeast two-hybrid method or a yeast cell surface display method, the sequences encoding $V_H$ and $V_L$ of the primary antibody leads are mutagenized in vitro to produce a secondary antibody library. The $V_H$ and $V_L$ sequences can be randomly mutagenized by "poison" PCR (or error-prone PCR), by DNA shuffling, or by any other way of random or site-directed mutagenesis (or cassette mutagenesis). After mutagenesis in the regions of $V_H$ and $V_L$, the complexity of the secondary antibody library may reach $10^4$ or more. Overall, the combined diversity or complexity of the total antibody libraries generated by using the methods of the present invention, including the primary and the secondary antibody libraries, may reach $10^{18}$ or more. The secondary antibody library are further screened for antibodies that bind the target antigen at high affinity by using the yeast-2-hybrid method as described above or other methods of screening in vivo or in vitro.

An advantage of the present invention is that the overall process of generating, selecting and optimizing large, diverse libraries of antibodies mimics the process of natural antibody diversification and maturation in a mammal. In the natural process of antibody affinity maturation, the affinity of the antibodies against their antigen(s) is progressively increased with the passage of time after immunization, largely due to the accumulation of point mutations specifically in the coding sequences of both the heavy- and light-chain variable regions.

According to the present invention, extensive diversification is achieved by recombination and mutagenesis of the $V_H$ and $V_L$ chain libraries derived from a wide variety of sources including natural and artificial or synthetic sources. The homologous combination of $V_H$ and $V_L$ in vivo to form the primary library of single-chain antibodies mimics the natural process of antibody gene assembly from different pools of gene segments encoding $V_H$ and $V_L$ of the antibodies. Since the method is preferably practiced with yeast cells, the highly efficient homologous recombination in yeast is particularly useful to facilitate such assembly of $V_H$ and $V_L$ in vivo.

The fast proliferation rate of yeast cells and ease of handling makes a process of "molecular evolution" dramatically shorter than the natural process of antibody affinity maturation in a mammal. Therefore, antibody repertoires with extremely high diversity can be produced and screened directly in yeast cells at a much lower cost and higher efficiency than prior processes such as the painstaking, stepwise "humanization" of monoclonal murine antibodies isolated by using the conventional hybridoma technology (a "protein redesign") or the recently-developed XENOMOUSE™ technology.

According to the "protein redesign" approach, murine monoclonal antibodies of desired antigen specificity are modified or "humanized" in vitro in an attempt to reshape the murine antibody to resemble more closely its human counterpart while retaining the original antigen-binding specificity. Riechmann et al. (1988) Nature 332:323–327. This humanization demands extensive, systematic genetic engineering of the murine antibody, which could take months, if not years. Additionally, extensive modification of the backbone of the murine monoclonal antibody may result in reduced specificity and affinity.

In comparison, by using the method of the present invention, fully human antibodies with high affinity to a specified antigen or antigens can be screened and isolated directly from yeast cells without going through site-by-site modification of the antibody, and without sacrifice of specificity and affinity of the selected antibodies.

The XENOMOUSE™ technology has been used to generate fully human antibodies with high affinity by creating strains of transgenic mice that produce human antibodies while suppressing the endogenous murine Ig heavy- and light-chain loci. However, the breeding of such strains of transgenic mice and selection of high affinity antibodies can take a long period of time. The antigen against which the pool of the human antibody is selected has to be recognized by the mouse as a foreign antigen in order to mount immune response; antibodies against a target antigen that does not have immunogenicity in a mouse may not be able to be selected by using this technology.

In contrast, by using the method of the present invention, libraries of antibody can not only be generated at a great diversity and complexity in yeast cells more efficiently and economically, but also be screened against virtually any protein or peptide target regardless of its immunogenicity. According to the present invention, any protein/peptide target can be expressed as a fusion protein with a DNA-binding domain (or an activation domain) of a transcription activator and selected against the library of antibody in a yeast-2-hybrid system. Moreover, multiple protein targets or a library of antigens may be arrayed in multiple-well plates and screened against the library of antibodies in a high throughput and automated manner.

Also compared to other approaches using transgenic goats and chickens to produce antibodies, the method of the present invention can be used to screen and produce fully human antibodies in large amounts without involving serious regulatory issues regarding the use of transgenic animals, as well as safety issues concerning containment of transgenic animals infected with recombinant viral vectors.

By using the method of the present invention, many requisite steps in the traditional construction of cDNA libraries can be eliminated. For example, the time-consuming and labor-intensive steps of ligation and recloning of cDNA libraries into expression vectors can be eliminated by direct recombination or "gap-filling" in yeast through general homologous recombination and/or site-specific recombination. Throughout the whole process of antibody library construction, the DNA fragments encoding antibody heavy chain and light chain are directly incorporated into a linearized yeast expression vector via homologous recombination without the recourse to extensive recloning.

Compared with the approach of using phage display to screen for high affinity antibodies in vitro, the method of the present invention provides efficient ways of screening for high affinity antibodies in eukaryotic cells in vivo. By using phage display technology, human Ig heavy chain and light chain variable regions are cloned, combinatorially reasserted, expressed and displayed as antigen-binding human Fab or scFv fragements on the surface of filamentous phage. Winter et al. (1994) Ann. Rev. Immunol. 433–455; and Rader et al. (1997) Current Opinion in Biotechnol. 8:503–508. The phage-displayed human antigen-binding fragments are then screened for their ability to bind an immobilized target antigen in vitro, a process called biopanning. When high affinity human antibodies are desired, the phage display approach can be problematic, presumably due to non-native conformation of antibody display on the surface and/or extensive selection or panning required for selection under in vitro conditions which bear little resemblance to the physiological condition of a human body. In contrast, by using the method of the present invention antibodies are selected based on their binding affinity to the target antigen in vivo. The antibodies are expressed in the cell, go through protein folding, and binds to its target antigen under a natural environment. Thus, the antibodies selected by using the method of the present invention should be more functionally relevant than those selected by panning in vitro.

### 1. Libraries of the Expression Vectors of the Present Invention

The present invention provides a library of expression vectors. In one embodiment, a library of yeast expression vectors is provided. The yeast expression vectors forming in the library comprise a first nucleotide sequence V**1** encoding a first polypeptide subunit; and a second nucleotide sequence V**2** encoding a second polypeptide subunit, the first and second nucleotide sequence each independently varying within the library of expression vectors.

According to the embodiment, the first polypeptide subunit and the second polypeptide subunit can be expressed as separate proteins or peptides. This may be accomplished by expressing the first and second polypeptide subunits from separate promoters, or by expressing bicistronically from the same promoter via an internal ribosomal entry site (IRES) or via a splicing donor-acceptor mechanism.

According to the embodiment, the yeast expression vector may be a $2\mu$ plasmid vector or a yc-type (centromeric) yeast vector, preferably a yeast-bacterial shuttle vector which contains a bacterial origin of replication.

Also according to the embodiment, V**1** in the library of expression vectors comprises a coding sequence of an antibody heavy-chain variable region ($V_H$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+CH, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**). V**2** comprises a coding sequence of an antibody light-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$).

Alternatively, V**1** in the library of expression vectors comprises a coding sequence of an antibody heavy-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$). V**2** comprises a coding sequence of an antibody heavy-chain variable region ($V_L$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**).

When V**1** and V**2** are expressed by the yeast expression vector in yeast cells, such as cells from the *Saccharomyces cerevisiae* strains, the protein subunits comprising the V**1** and V**2** polypeptide segments respectively interact with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds) to form a double-chain protein complex.

Optionally, the first and second polypeptide subunits may each further comprise a plurality of cysteine residues, preferably 2–8 Cys residues. The additional cysteines residues may be located at or adjacent the N- or C-terminus of the first and second polypeptide subunits. As illustrated in FIG. **10**A, the additional cysteines residues is preferably located near the C-terminus of the heavy chain and light chain regions of a dcFv, Fab and a fully assembled antibody.

It is believed that by adding more cysteine subunits near the termini of the subunits, the intermolecular interactions between the two subunits should be enhanced through formation of Cys-Cys disulfide bonds, thus further stabilizing the assembly of the protein complex formed by the two subunits.

Alternatively, the first and second polypeptide subunits may each further comprise a "zipper" domain at or adjacent the N- or C-terminus of the polypeptide. As illustrated in FIG. **10**B, the zipper domain is preferably located at the C-terminus of the heavy chain and light chain regions of a dcFv, Fab and a fully assembled antibody.

A zipper domain is a protein or peptide structural motif that interacts with each other through non-covalent interactions such as coiled-coil interactions and brings other proteins fused with the zipper domains into close proximity. Examples of zipper domains include, but are not limited to, leucine zippers (or helix-loop-helix, also called bHLHzip motif formed between the nuclear oncoproteins Fos and Jun (Kouzarides and Tiff (1989) "Behind the Fos and Jun leucine zipper° Cancer Cells 1: 71–76); leucine zippers formed between proto-oncoproteins Myc and Max (Luscher and Larsson (1999) "The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation" Oncogene 18:2955–2966); zipper motifs from adhesion proteins such as N-terminal domain of neural cadherin (Weis (1995) "Cadherin structure: a revealing zipper" 3:425–427); zipper-like structural motifs from collagen triple helices or cartilage oligomeric matrix proteins (Engel and Prockop "The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper" Annu. Rev. Biophys. Biophys. Chem. 20:137–152; and Terskikh et al. (1997) "Peptabody": a new type of high avidity binding protein" Proc. Natl. Acad. Sci. USA 94:1663–1668).

The zipper domain may be fused to the N- or C-terminus of the polypeptide subunits, preferably at the C-terminus of the subunits. For example, the leucine zipper domain derived from the oncoprotein Jun can be expressed as a fusion protein with an antibody heavy chain whereas the leucine zipper domain derived from the oncoprotein Fos can be expressed as another fusion protein with an antibody light chain. Since the Jun and Fos leucine zipper domains can bind to each other with high affinity, the antibody heavy chain and light chain fused with Jun and Fos zipper, respectively, can be brought into close proximity and form a heterodimer upon binding between these two zipper domains.

It is believed that by adding a zipper domain near the termini of the subunits, the intermolecular interactions between the two subunits should be enhanced through non-covalent interactions (e.g. hydrophobic interactions), thus further stabilizing the assembly of the protein complex formed by the two subunits. Moreover, fusing a zipper domain derived from nuclear protein such as Jun and Fos to the subunits may facilitate efficient transportation of the subunits to the nucleus where the protein complex formed between the two subunits performs desired functions such as transcriptional activation of a reporter gene.

In addition, the first or the second polypeptide subunit may further comprise a "bundle" domain at or adjacent the C-terminus of the polypeptide. As used herein, a "bundle domain" refers to a protein or peptide structural motif that can interact with itself to form a homo-polymer such as a homopentalmer. As illustrated in FIG. **10**C, the bundle domains bring the protein complex together by polymerization through non-covalent interactions such as coiled-coil interactions. It is believed that polymerization of the protein complex should enhance the avidity of the protein complexes to their binding target through multivalent binding.

For example, the coiled-coil assembly domain of the cartilage oligomeric matrix protein (COMP) may serve as a bundle domain. The N-terminal fragment of rat COMP comprises residue 20–83. This fragment can form pentamers simillar to the assembly domain of the native protein. The fragment adopts a predominantly alpha-helical structure. Efimov et al. (1994) "The thrombospondin-like chains of cartilage oligomeric matrix protein are assembled by a five-stranded alpha-helical bundle between residues 20 and 83" FEBS Lett. 341:54–58.

The coiled-coil domain of the nudE gene of the filamentous fungus *Aspergillu nidulans* or the gene encoding the nuclear distribution protein RO11 of *Neurospora crassa* may also serve a bundle domain. The product of the nudE gene, NUDE, is a homologue of the RO11 protein. The N-terminal coiled-coil domain of the NUDE protein is highly conserved; and a similar coiled-coil domain is present in several putative human proteins and in the mitotic phosphoprotein 43 (MP43) of *X. laevis*. Efimov and Morris (2000) "The LIS1-related NUDF protein of *Aspergillu nidulans* interacts with the coiled-coil domain of the NUDE/RO11 protein" J. Cell Biol. 150:681–688.

In addition, the coiled-coil segments or fribritin encoded by bacteriophage T4 may also serve as a bundle domain. The bacteriophage T4 late gene wac (Whisker's antigen control) encodes a fibrous protein which forms a collar/whiskers complex. Analysis of the 486 amino acid sequence of fibritin reveals three structural components: a 408 amino acid region that contains 12 putative coiled-coil segments with a canonical heptad (a-b-c-d-e-f-g)n substructure where the "a" and "d" positions are preferentially occupied by apolar residues, and the N and C-terminal domains (47 and 29 amino acid residues, respectively). The alpha-helical segments are separated by short "linker" regions, variable in length, that have a high proportion of glycine and proline residues. Co-assembly of full-length fibritin and the N-terminal deletion mutant, as well as analytical centrifugation, indicates that the protein is a parallel triple-standard alpha-helical coiled-coil. The last 18 C-terminal residues of fibritin are required for correct trimerisation of gpwac monomers in vivo. Efimov et al. (1994) "Fibritin encoded by bacteriophage T4 gene wac has a parallel triple-stranded alpha-helical coiled-coiled structure" J. Mol. Biol. 242:470–486.

The bundle domain may be fused to the C-terminus of the first or second polypeptide subunit. Optionally, a semi-rigid linker may be used to link the bundle domain to the subunit. As illustrated in FIG. **10**D, this linker serves a hinge that allows a controlled conformational flexibility of the cluster of protein complexes formed between the first and second polypeptide subunits. For example, the 24 amino acid hinge region derived from camel IgG, $(PQ)_2PK(PQ)_4PKPQPK(PE)_2$ [SEQ ID NO: 79] may be used as such a semi-rigid linker. This linker serves a hinge that allows a controlled conformational flexibility of the cluster of protein complexes formed between the first and second polypeptide subunits, which provides the space necessary for multivalent binding. Further, cysteine residues may be introduced to the bundle domain, preferably near the N-terminus, to allow the formation of additional disulfide bonds between the bundle domains.

According to this design of the present invention, avidity of the protein complex formed between a heavy chain and light chain region of antibody (i.e. an antibody) may be dramatically increased by fusing a bundle domain (e.g. COMP) to the C-terminus of the heavy chain. Polymerization of the bundle domains should bring multiple antibodies together and thus enhance the avidity interactions between the antibodies with their targets due to multivalent binding. This process mimics the natural assembly of multiple IgM produced during the primary immune response. The low affinity of IgM is compensated by its pentameric structure resulting a high avidity toward repetitive antigenic determinants present on the surface of bacteria or viruses. Roitt (1991) Essential Immunology (Oxford/Blackwell, London), 7th Ed., pp. 65–84.

In another embodiment, a library of expression vectors is provided. The expression vector in the library comprises: a transcription sequence encoding an activation domain AD or a DNA binding domain BD of a transcription activator; a first nucleotide sequence **V1** encoding a first polypeptide subunit; and a second nucleotide sequence **V2** encoding a second polypeptide subunit. The activation domain or the DNA binding domain of the transcription activator and the first polypeptide subunit are expressed as a single fusion protein. The second polypeptide subunit is expressed as a separate protein or peptide from the first polypeptide. In addition, **V1** and **V2** each independently varies within the library of expression vectors.

According to the embodiment, the expression vector may be any gene-transferring vector as long as it is able to introduce the library of expression vectors to a desired location within a host cell, such as by transformation, transfection and transduction of the expression vector into a host cell. The expression vector may be a bacterial, phage, yeast, mammalian or a viral expression vector, and preferably a yeast expression vector.

Also according to the embodiment, the transcription activator sequence may be located 5' relative to the first nucleotide sequence. Alternatively, the transcription activator sequence may be located 3' relative to the first nucleotide sequence.

In a variation of the embodiment, **V1** is a coding sequence of an antibody heavy-chain variable region ($V_H$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**). **V2** is a coding sequence of an antibody light-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$).

Alternatively, **V1** is a coding sequence of an antibody light-chain variable region ($V_L$) or an antibody light-chain including both the variable and constant region ($V_L$+$C_L$). **V2** is a coding sequence of an antibody heavy-chain variable region ($V_H$) or an antibody heavy-chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**).

Optionally, AD is an activation domain of yeast GAL 4 transcription activator; and BD is a DNA binding domain of yeast GAL 4 transcription activator.

When **V1** and **V2** are expressed by the expression vector in host cells, such as cells from the *Saccharomyces cerevisiae* strains, the fusion protein comprising the AD and **V1**-encoded polypeptide subunit, and **V2**-encoded polypeptide subunit interact with each other and form a protein complex with one or more conformations. The conformation (s) adopted by the protein complex of the AD/**V1** fusion and **V2**-encoded polypeptide subunit may have suitable binding site(s) for a specific target protein. For example, the protein complex may be dsFv, Fab or an full antibody that binds to its specific target antigen. The AD domain of the fusion protein should be able to activate transcription of gene(s) once the AD and BD domains are reconstituted to form an active transcription activator in vitro or in vivo by a two-hybrid method.

According to any of the libraries described above, the diversity of the first and/or the second polypeptide subunit encoded by **V1** and **V2** within the library of expression vectors may be preferably between $10^3$–$10^8$, more preferably between $10^4$–$10^8$, and most preferably between $10^5$–$10^8$.

According to any of the libraries described above, the diversity of the first and/or the second polypeptide subunit encoded by **V1** and **V2** within the library of expression vectors may be preferably at least $10^3$, more preferably at least $10^4$, and most preferably at least $10^5$.

Also according to any of the libraries described above, the diversity of the fusion proteins encoded by the library of expression vectors is preferably between $10^6$–$10^{18}$, more preferably between $10^9$–$10^{18}$ and most preferably between $10^{10}$–$10^{18}$.

Also according to any of the libraries described above, the diversities of the first and second polypeptide subunits need not be derived from mutagenizing one or more proteins that are known to bind to a target peptide or protein. For example, the first and second polypeptide subunits need not be derived from mutagenizing a single antibody (e.g. the antibody Herceptin®) which is known to bind to a target peptide or protein (Her-2 receptor). This reflects a novel ability of the present invention to identify new protein-protein binding pairs from a random pool of sequences instead of having to know in advance a protein that binds to a target and then form a library of mutants from that known binding protein.

The elements of the expression vector in the library are described in detail below.

1) The Backbone of the Expression Vector

The expression vector of the present invention may be based on any type of vector as long as the vector that can transform, transfect or transduce a host cell. The expression vector contains a library of the **V1** sequences and a library of **V2** sequences, and preferably contains a sequence encoding an activation domain (AD) of a transcriptional activator. The acceptor vector may be plasmids, phages or viral vectors as long as it is able to replicate in vitro, or in a host cell, or to convey the library of the **V1** and **V2** sequences to a desired location within a host cell. Examples of host cells include, but are not limited to, bacterial (e.g. *E. coli, Bacillus subtilis*, etc.), yeast, animal, plant, and insect cells.

In a preferred embodiment, the expression vector is based on a yeast plasmid, especially one from *Saccharomyces cerevisiae*. After transformation of yeast cells, the exogenous DNA encoding the **V1** and **V2** sequences are uptaken by the cells and subsequently expressed by the transformed cells.

More preferably, the expression vector may be a yeast-bacteria shuttle vector which can be propagated in either *Escherichia coli* or yeast Struhl, et al. (1979) Proc. Natl. Acad. Sci. 76:1035–1039. The inclusion of *E. coli* plasmid DNA sequences, such as pBR322, facilitates the quantitative preparation of vector DNA in *E. coli*, and thus the efficient transformation of yeast.

The types of yeast plasmid vector that may serve as the shuttle may be a replicating vector or an integrating vector. A replicating vector is yeast vector that is capable of mediating its own maintenance, independent of the chromosomal DNA of yeast, by virtue of the presence of a functional origin of DNA replication. An integrating vector relies upon recombination with the chromosomal DNA to facilitate replication and thus the continued maintenance of the recombinant DNA in the host cell. A replicating vector may be a 2μ-based plasmid vector in which the origin of DNA replication is derived from the endogenous 2μ plasmid of yeast. Alternatively, the replicating vector may be an autonomously replicating (ARS) vector, in which the "apparent" origin of replication is derived from the chromosomal DNA of yeast. Optionally, the replicating vector may be a centromeric (CEN) plasmid which carries in addition to one of the above origins of DNA replication a sequence of yeast chromosomal DNA known to harbor a centromere.

The vectors may be transformed into yeast cells in a closed circular form or in a linear form. Transformation of yeast by integrating vectors, although with inheritable stability, may not be efficient when the vector is in in a close circular form (e.g. 1–10 transformants per ug of DNA). Linearized vectors, with free ends located in DNA sequences homologous with yeast chromosomal DNA, transforms yeast with higher efficiency (100–1000 fold) and the transforming DNA is generally found integrated in sequences homologous to the site of cleavage. Thus, by cleaving the vector DNA with a suitable restriction endonuclease, it is possible to increase the efficiency of transformation and target the site of chromosomal integration. Integrative transformation may be applicable to the genetic modification of brewing yeast, providing that the efficiency of transformation is sufficiently high and the target DNA sequence for integration is within a region that does not disrupt genes essential to the metabolism of the host cell.

ARS plasmids, which have a high copy number (approximately 20–50 copies per cell) (Hyman et al., 1982), tend to be the most unstable, and are lost at a frequency greater than 10% per generation. However, the stability of ARS plasmids can be enhanced by the attachment of a centromere; centromeric plasmids are present at 1 or 2 copies per cell and are lost at only approximately 1% per generation.

The expression vector of the present invention is preferably based on the 2μ plasmid. The 2μ plasmid is known to be nuclear in cellular location, but is inherited in a non-Mendelian fashion. Cells that lost the 2μ plasmid have been shown to arise from haploid yeast populations having an average copy number of 50 copies of the 2μ plasmid per cell at a rate of between 0.001% and 0.01% of the cells per generation. Futcher & Cox (1983) J. Bacteriol. 154:612. Analysis of different strains of *S. cerevisiae* has shown that the plasmid is present in most strains of yeast including brewing yeast. The 2μ plasmid is ubiquitous and possesses a high degree of inheritable stability in nature.

The 2μ plasmid harbors a unique bidirectional origin of DNA replication which is an essential component of all 2μ-based vectors. The plasmid contains four genes, REP1, REP2, REP3 and FLP which are required for the stable maintenance of high plasmid copy number per cell Jaysram et al. (1983) Cell 34:95. The REP1 and REP2 genes encode transacting proteins which are believed to function in concert by interacting with the REP3 locus to ensure the stable partitioning of the plasmid at cell division. In this respect, the REP3 gene behaves as a cis acting locus which effects the stable segregation of the plasmid, and is phenotypically analogous to a chromosomal centromere. An important feature of the 2μ plasmid is the presence of two inverted DNA sequence repeats (each 559 base-pairs in length) which separate the circular molecule into two unique regions. Intramolecular recombination between the inverted repeat sequences results in the inversion of one unique region relative to the other and the production in vivo of a mixed population of two structural isomers of the plasmid, designated A and B. Recombination between the two inverted repeats is mediated by the protein product of a gene called the FLP gene, and the FLP protein is capable of mediating high frequency recombination within the inverted repeat region. This site specific recombination event is believed to provide a mechanism which ensures the amplification of plasmid copy number. Murray et al. (1987) EMBO J. 6:4205.

The expression vector may also contain an *Escherichia coli* origin of replication and *E. coli* antibiotic resistance genes for propagation and antibiotic selection in bacteria. Many *E. coli* origins are known, including ColE1, pMB1 and pBR322, The ColE origin of replication is preferably used in this invention. Many *E. coli* drug resistance genes are known, including the ampicillin resistance gene, the chloramphenoicol resistance gene and the tetracycline resistance gene. In one particular embodiment, the ampicillin resistance gene is used in the vector.

The transformants that carry the **V1** and **V2** sequences may be selected by using various selection schemes. The selection is typically achieved by incorporating within the vector DNA a gene with a discernible phenotype. In the case of vectors used to transform laboratory yeast, prototrophic genes, such as LEU2, URA3 or TRP1, are usually used to complement auxotrophic lesions in the host. However, in order to transform brewing yeast and other industrial yeasts, which are frequently polyploid and do not display auxotrophic requirements, it is necessary to utilize a selection system based upon a dominant selectable gene. In this respect replicating transformants carrying $2\mu$-based plasmid vectors may be selected based on expression of marker genes which mediate resistance to: antibiotics such as G418, hygromycin B and chloramphenicol, or otherwise toxic materials such as the herbicide sulfometuron methyl, compactin and copper.

2) The **V1** and **V2** Variable Sequences

The first and the second polypeptide subunits encoded by **V1** and **V2**, respectively, may be subunits of any multimeric protein. The sequence of the multimeric protein varies within a library or a collection of multimeric proteins. Example of the multimeric proteins include, but are not limited to antibodies, growth factor receptors, T cell receptors, cytokine receptors, tyrosine kinase-associated receptors, and MHC proteins.

In preferred embodiment, the multimeric proteins are a library of antibodies, and more preferably human antibodies. For example, the first polypeptide subunit encoded by the library of expression vectors may be a human antibody heavy chain variable region ($V_H$) or a full heavy chain including both the variable and constant regions ($V_H$+$C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**). The second polypeptide subunit encoded by by the library of expression vectors may be a human antibody light-chain variable region ($V_L$) or a light chain including both the variable and constant region ($V_L$+$C_L$).

DNA sequences encoding human antibody heavy chain and light chain may be polynucleotide segments of at least 30 contiguous base pairs substantially encoding genes of the immunoglobulin superfamily. A. F. Williams and A. N. Barclay (1989) "The Immunoglobulin Gene Superfamily", in Immunoglobulin Genes, T. Honjo, F. W. Alt, and T. H. Rabbitts, eds., Academic Press: San Diego, Calif., pp.361–387. The antibody genes are most frequently encoded by human, non-human primate, avian, porcine, bovine, ovine, goat, or rodent heavy chain and light chain gene sequences.

The library of DNA sequences encoding human antibody heavy chain and light chain may be derived from a variety of sources. For example, mRNA encoding the human antibody libraries may be extracted from cells or organs from immunized or non-immunized animals or humans. Preferably, organs such as human fetal spleen and lymph nodes may be used. Peripheral blood cells from non-immunized humans may also be used. The blood samples may be from an individual donor, from multiple donors, or from combined blood sources.

The human antibody coding sequences may be derived and amplified by using sets of oligonucleotide primers to amplify the cDNA of human heavy and light chains by polymerase chain reaction (PCR). Orlandi et al. (1989) Proc. Natl. Acad. Sci. USA 86: 3833–3837. For example, blood sample may be from healthy volunteers and B-lymphocyte in the blood can be isolated. RNA can be prepared by following standard procedures. Cathala et al. (1983) DNA 3:329. The cDNA can be made from the isolated RNA by using reverse transcriptase.

Alternatively, the antibody coding sequences may be derived from an artificially rearranged immunoglobulin gene or genes. For example, immunoglobulin genes may be rearranged by joining of germ line V segments in vitro to J segments, and, in the case of $V_H$ domains, D segments. The joining of the V, J and D segments may be facilitated by using PCR primers which have a region of random or specific sequence to introduce artificial sequence or diversity into the products.

Optionally, the variable sequences **V1** and **V2** of the library of expression vectors may also be derived from multimeric proteins other than antibodies. **V1** and **V2** may be different subunits of a non-antibody multimeric protein, such as membrance proteins and cell surfaces receptor proteins, e.g. insulin receptor, MHC proteins (e.g. class I MHC and class II MHC protein), CD3 receptor, T cell receptors, cytokine receptors such as interleukin-2 (IL-2) receptor which is made of $\alpha$, $\gamma$, and $\gamma$ subunits, tyrosine-kinase-associated receptors such as Src, Yes, Fgr, Lck, Lyn, Hck, and Blk. The tyrosine-kinase-associated receptors contain SH2 and SH3 domains which are held there partly by their interactions with transmembrane receptor proteins and partly by covalently attached lipid chains. For example, **V1** and **V2** sequences may be mutagenized sequences of the SH2 and SH3 domains of a tyrosine-kinase-associated receptor such as Src, respectively, which are incorporated into the expression of vector of the present invention and screened against various ligands for this receptor.

A reflection of the power and versatility of the methods of the present invention is that the **V1** and **V2** sequences need not be based in any way on a protein sequence known to bind to the target. Instead, **V1** and **V2** may be from any source and may have a diversity that is entirely independent from the target, or one or more lead proteins known to bind to the target.

3) The Target Proteins and Peptides

The target fusion protein may comprise any target protein or peptide that may be expressed or otherwise present in a host cell. The target protein may be a member of library of proteins or peptides, such as a collection of human ESTs, a total library of human ESTs, a collection of domain structures (e.g. Zn-finger protein domains), or a totally random peptide library. For example, the target protein or peptide may be a disease-associated antigen, such as tumor surface antigen such as B-cell idiotypes, CD20 on malignant B cells, CD33 on leukemic blasts, and HER2/neu on breast cancer. Antibody selected against these antigens can be used in a wide variety of therapeutic and diagnostic applications, such as treatment of cancer by direct administration of the antibody itself or the antibody conjugated with a radioisotope or cytotoxic drug, and in a combination therapy involving coadministration of the antibody with a chemotherapeutic agent, or in conjunction with radiation therapy.

Alternatively, the target protein may be a growth factor receptor. Examples of the growth factor include, but are not limited to, epidermal growth factors (EGFs), transferrin, insulin-like growth factor, transforming growth factors (TGFs), interleukin-1, and interleukin-2. For example, high expression of EGF receptors have been found in a wide variety of human epithelial primary tumors. TGF-α have been found to mediate an autocrine stimulation pathway in cancer cells. Several murine monoclonal antibody have been demonstrated to be able to bind EGF receptors, block the binding of ligand to EGF receptors, and inhibit proliferation of a variety of human cancer cell lines in culture and in xenograft medels. Mendelsohn and Baselga (1995) Antibodies to growth factors and receptors, in Biologic Therapy of Cancer, $2^{nd}$ Ed., J B Lippincott, Philadelphia, pp607–623. Thus, fully human antibodies selected against these growth factors by using the method of the present invention can be used to treat a variety of cancer.

The target protein may also be cell surface protein or receptor associated with coronary artery disease such as platelet glycoprotein lib/IIIa receptor, autoimmune diseases such as CD4, CAMPATH-1 and lipid A region of the gram-negative bacterial lipopolysaccharide. Humanized antibodies against CD4 has been tested in clinical trials in the treatment of patients with mycosis fungoides, generalized postular psoriasis, severe psorisis, and rheumatoid arthritis. Antibodies against lipid A region of the gram-negative bacterial lipopolysaccharide have been tested clinically in the treatment of septic shock. Antibodies against CAMPATH-1 has also been tested clinically in the treatment of against refractory rheumatoid arthritis. Thus, fully human antibodies selected against these growth factors by using the method of the present invention can be used to treat a variety of autoimmune diseases. Vaswani et al. (1998) "Humanized antibodies as potential therapeutic drugs" Annals of Allergy, Asthma and Immunology 81:105–115.

The target protein or peptide may also be proteins or peptides associated with human allergic diseases, such as those inflammatory mediator protein, e.g. Interleukin-1 (IL-1), tumor necrosis factor (TNF), leukotriene receptor and 5-lipoxygenase, and adhesion molecules such as V-CAMNLA-4. In addition, IgE may also serve as the target antigen because IgE plays pivotal role in type I immediate hypersensitive allergic reactions such as asthma. Studies have shown that the level of total serum IgE tends to correlate with severity of diseases, especially in asthma. Burrows et al. (1989) "Association of asthma with serum IgE levels and skin-test reactivity to allergens" New Engl. L. Med. 320:271–277. Thus, fully human antibodies selected against IgE by using the method of the present invention may be used to reduce the level of IgE or block the binding of IgE to mast cells and basophils in the treatment of allergic diseases without having substantial impact on normal immune functions.

The target protein may also be a viral surface or core protein which may serve as an antigen to trigger immune response of the host. Examples of these viral proteins include, but are not limited to, glycoproteins (or surface antigens, e.g., GP120 and GP41) and capsid proteins (or structural proteins, e.g., P24 protein); surface antigens or core proteins of hepatitis A, B, C, D or E virus (e.g. small hepatitis B surface antigen (SHBsAg) of hepatitis B virus and the core proteins of hepatitis C virus, NS3, NS4 and NS5 antigens); glycoprotein (G-protein) or the fusion protein (F-protein) of respiratory syncytial virus (RSV); surface and core proteins of herpes simplex virus HSV-1 and HSV-2 (e.g., glycoprotein D from HSV-2).

The target protein may also be a mutated tumor suppressor gene that have lost its tumor-suppressing function and may render the cells more susceptible to cancer. Tumor suppressor genes are genes that function to inhibit the cell growth and division cycles, thus preventing the development of neoplasia. Mutions in tumor suppressor genes cause the cell to ignore one or more of the components of the network of inhibitory signals, overcoming the cell cycle check points and resulting in a higher rate of controlled cell growth—cancer. Examples of the tumor suppressor genes include, but are not limited to, DPC-4, NF-1, NF-2, RB, p53, WT1, BRCA1 and BRCA2.

DPC-4 is involved in pancreatic cancer and participates in a cytoplasmic pathway that inhibits cell division. NF-1 codes for a protein that inhibits Ras, a cytoplasmic inhibitory protein. NF-1 is involved in neurofibroma and pheochromocytomas of the nervous system and myeloid leukemia. NF-2 encodes a nuclear protein that is involved in meningioma, schwanoma, and ependymoma of the nervous system. RB codes for the pRB protein, a nuclear protein that is a major inhibitor of cell cycle. RB is involved in retinoblastoma as well as bone, bladder, small cell lung and breast cancer. P53 codes for p53 protein that regulates cell division and can induce apoptosis. Mutation and/or inaction of p53 is found in a wide ranges of cancers. WTI is involved in Wilms tumor of the kidneys. BRCA 1 is involved in breast and ovarian cancer, and BRCA2 is involved in breast cancer. Thus, fully human antibodies selected against a mutated tumor suppressor gene product by using the method of the present invention can be used to block the interactions of the gene product with other proteins or biochemicals in the pathways of tumor onset and development.

## 2. Construction of the Library of Expression Vectors of the Present Invention

The library of expression vectors described above can be constructed using a variety of recombinant DNA techniques. The present invention provides novel and efficient methods of constructing these libraries of expression vectors with extreme diversity of **V1** and **V2** in vivo and in vitro.

The methods of the present invention are provided by exploiting the inherent ability of yeast cells to facilitate homologous recombination at an extremely high efficiency. The mechanism of homologous recombination in yeast and its applications is briefly described below.

Yeast *Saccharomyces cerevisiae* has an inherited genetic machinery to carry out efficient homologous recombination in the cell. This mechanism is believed to benefit the yeast cells for chromosome repair purpose and traditionally also called gap repair or gap filling. By this mechanism of efficient gap filling, mutations can be introduced into specific loci of the yeast genome. For example, a vector carrying the mutant gene contains two sequence segments that are homologous to the 5' and 3' open reading frame (ORF) sequences of the gene that is intended to be interrupted or mutated. The plasmid also contains a positive selection marker such as a nutritional enzyme allele, such as ura3, or an antibiotic resistant marker such as Geneticine (g418) that are flanked by the two homologous segments. This plasmid is linearized and transformed into the yeast cells. Through homologous recombination between the plasmid and the yeast genome at the two homologous recombination sites, a reciprocal exchange of the DNA content occurs between the wild type gene in the yeast genome and the mutant gene (including the selection marker gene) that are flanked by the two homologous sequence segments. By selecting for the positive nutritional marker, surviving yeast cells will loose the original wild type gene and will adopt the mutant gene. Pearson B M, Hernando Y, and Schweizer M, (1998) Yeast 14: 391–399. This mechanism has also been used to make systematic mutations in all 6,000 yeast genes or ORFs for functional genomics studies. Because the exchange is reciprocal, similar approach has been used successfully for cloning yeast genomic fragments into plasmid vector. Iwasaki T, Shirahige K, Yoshikawa H, and Ogasawara N, Gene 1991, 109 (1): 81–87.

By using homologous recombination in yeast, gene fragments or synthetic oligonucleotides can also be cloned into a plasmid vector without a ligation step. In this application, a targeted gene fragment is usually obtained by PCR amplification (or by using the conventional restriction digestion out of an original cloning vector). Two short fragment sequences that are homologous to the plasmid vector are added to the 5' and 3' of the target gene fragment in the PCR amplification. This can be achieved by using a pair of PCR primers that incorporate the added sequences. The plasmid vector typically includes a positive selection marker such as nutritional enzyme allele such as ura3, or an antibiotic resistant marker such as geneticin (g418). The plasmid vector is linearized by a unique restriction cut in between the sequence homologies that are shared with the PCR-amplified target, thereby creating an artificial gap at the cleavage site. The linearized plasmid vector and the target gene fragment flanked by sequences homologous to the plasmid vector are co-transformed into a yeast host strain. The yeast recognizes the two stretches of sequence homologies between the vector and target fragment, and facilitates a reciprocal exchange of DNA contents through homologous recombination at the gap. As the consequence, the target fragment is automatically inserted into the vector without ligation in vitro.

There are a few factors that may influence the efficiency of homologous recombination in yeast. The efficiency of the gap repair is correlated with the length of the homologous sequences flanking both the linearized vector and the targeted gene. Preferably, a minimum of 30 base pairs may be required for the length of the homologous sequence, and 80 base pairs may give a near-optimized result. Hua, S. B. et al. (1997) “Minimum length of sequence homology required for in vitro cloning by homologous recombination in yeast” Plasmid 38:91–96. In addition, the reciprocal exchange between the vector and gene fragment is strictly sequence-dependent, i.e. not causing frame shift in this type of cloning. Therefore, such a unique characteristic of the gap-repair cloning assures insertion of gene fragments with both high efficiency and precision. The high efficiency makes it possible to clone two or three targeted gene fragments simultaneously into the same vector in one transformation attempt. Raymond K., Pownder T. A., and Sexson S. L., (1999) Biotechniques 26: 134–141. The nature of precision sequence conservation through homologous recombination makes it possible to clone targeted genes in question into expression or fusion vectors for direct function examinations. So far many functional or diagnostic applications have been reported using homologous recombination. El-Deiry W. W., et al., Nature Genetics1: 4549, 1992 (for p53), and Ishioka C., et al., PNAS, 94: 2449–2453, 1997 (for BRCA1 and APC).

A library of gene fragments may also be constructed in yeast by using homologous recombination. For example, a human brain cDNA library can be constructed as a two-hybrid fusion library in vector pJG4-5. Guidotti E., and Zervos A. S. (1999) “In vivo construction of cDNA library for use in the yeast two-hybrid systems” Yeast 15:715–720.

It has been reported that a total of 6,000 pairs of PCR primers were used for amplification of 6,000 known yeast ORFs for a study of total yeast genomic protein interaction. Hudson, J. Jr, et al. (1997) Genome Res. 7:1169–1173. Uetz et al. conducted a comprehensive analysis-of protein-protein interactions in Saccharomyces cerevisiae. Uetz et al. (2000) Nature 403:623–627. The protein-protein interaction map of the budding yeast was studied by using a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Ito et al. (2000) Proc. Natl. Acad. Sci. USA. 97:1143–1147. The genomic protein linkage map of Vaccinia virus was studied by McCraith S., Holtzman T., Moss B., and Fields, S. (2000) Proc. Natl. Acad. Sci. USA 97: 4879–4884.

According to the present invention, the **V1** and **V2** sequences are introduced into an expression vector by homologous recombination performed directly in yeast cells.

1) Cloning of **V1** and **V2** in Separate Fragments into an Expression Vector Through Two Independent Events of Homologous Recombination in Yeast

In one embodiment for the method for generating the library of expression vectors, the **V1** and **V2** sequences may be cloned into an expression vector in vivo in two separate fragments through two independent events of homologous recombination in yeast.

The method comprises:

a) transforming into yeast cells i) a linearized yeast expression vector having a 5'- and 3'-terminus sequence at a first site of linearization; and ii) a library of first insert nucleotide sequences that are linear, double stranded, each of the first insert sequences comprising a first nucleotide sequence **V1** encoding a first polypeptide subunit, a 5'- and 3'-flanking sequence at the ends of the first insert sequence which are sufficiently homologous to the 5'- and 3'-terminus sequences of the vector at the first site of linearization, respectively, to enable homologous recombination to occur;

b) having homologous recombination occur between the vector and the irst insert sequence in the transformed yeast cells, such that the first insert equence is included in the vector;

c) isolating from the transformed yeast cells the vectors that contain the library of the first insert sequences;

d) linearizing the vectors containing the library of the first insert sequences to generate a 5'- and 3'-terminus sequence at a second site of linearization;

e) transforming into yeast cells
  i) the linearized yeast expression vectors in step d), and
  ii) a library of second insert nucleotide sequences that are linear, double stranded, each of the second insert sequences comprising a second nucleotide sequence **V2** encoding a second polypeptide subunit, a 5'- and 3'-flanking sequence at the ends of the second insert sequence which are sufficiently homologous to the 5'- and 3'-terminus sequences of the vector at the second site of linearization, respectively, to enable homologous recombination to occur; and

f) having homologous recombination occur between the linearized yeast expression vector at the second linearization site and the second insert sequences in the transformed yeast cells, such that the second insert sequence is included in the vector.

According to the embodiment, the first polypeptide subunit and the second polypeptide subunit are expressed as separate proteins or peptides. This may be accomplished by expressing the first and second polypeptide subunits from separate promoters, or by expressing bicistronically from the same promoter via an internal ribosomal entry site (IRES) or via a splicing donor-acceptor mechanism.

According to the embodiment, the 5'- or 3'-flanking sequence of the insert nucleotide sequence is preferably between about 30–120 bp in length, more preferably between about 40–90 bp in length, and most preferably between about 60–80 bp in length.

FIG. **2** illustrates an embodiment of this method according to the present invention. The coding sequences for **V1** (e.g., $V_H+C_H1$) and **V2** (e.g., $V_L+C_L$) are carried by separate PCR fragments and cloned into an expression vector sequentially following two independent events of homologous recombination in yeast.

As illustrated in FIG. **2**, the **V1** fragment has a 5' flanking sequence and a 3' flanking sequence that are homologous to the 5' and 3' terminus of a linearized expression vector, respectively. When the **V1** fragment and the linearized expression vector are introduced into a host cell, for example, transformed into a yeast cell, the "gap" (the first linearization site) created by linearization of the expression vector is filled by the **V1** fragment insert through recombination of the homologous sequences at the 5' and 3' terminus of these two linear double-stranded DNA. Through this event of homologous recombination, a library of circular vectors carrying the variable sequence **V1** is generated.

This library of circular vectors is then cleaved at a second linearization site, for example, a site downstream of **V1**. The **V2** fragment has a 5' flanking sequence and a 3' flanking sequence that are homologous to the 5' and 3' terminus of the linearized expression vector at the second linearization site. The **V2** fragment and the linearized expression vector are transformed into a yeast cell. Through a second event of homologous recombination, the **V2** fragment is inserted into the linearized expression vector at the second linearization site. As a result, a library of circular vectors carrying the variable sequences **V1** and **V2** is generated.

Each flanking sequence added to the **V1** and **V2** coding sequence may be preferably between about Each flanking sequence added to the 5' and 3'-terminus of **V2** sequence is preferably between about 30–120 bp in length, more preferably between about 40–90 bp in length, and most preferably between about 45–55 bp in length.

When the **V1** and **V2** coding sequences are inserted into an expression vector containing an AD domain, it is preferred that the reading frames of the **V1** or **V2** fragments are conserved with upstream AD reading frame.

Depending on the cloning expression vector used, additional features such as affinity tags and unique restriction enzyme recognition sites may be added to the expression for the convenience of detection and purification of the inserted **V1** and **V2** sequences. Examples of affinity tags include, but are not limited to, a polyhistidine tract, polyarginine, glutathione-S-transferase (GST), maltose binding protein (MBP), a portion of staphylococcal protein A (SPA), and various immunoaffinity tags (e.g. protein A) and epitope tags such as those recognized by the EE (Glu-Glu) antipeptide antibodies.

In a preferred embodiment, the **V1** and **V2** sequences may be the coding sequences for a heavy chain and a light-chain, respectively, which are derived from a human antibody repertoire. To generate the **V1** and **V2** coding sequences from the human antibody repertoire, a complex human antibody cDNA gene pool may generated by using the methods known in the art. Sambrook, J., et al. (1989) Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; and Ausubel, F. M. et al. (1995) Current Protocols in Molecular Biology" John Wiley & Sons, NY.

Total RNA may be isolated from sources such as the white cells (mainly B cells) contained in peripheral blood supplied by un-immunized humans, or from human fetal spleen and lymph nodes. First strand cDNA synthesis may be synthesized performed by using methods known in the art, such as those described by Marks et al. Marks et al. (1991) Eur. J. Immunol. 21:985–991.

Specifically, a mixture of heavy and light chain cDNA primer sets designed to anneal to the constant regions may be used for priming the synthesis of cDNA of heavy chain and light chains (both kappa Vκ and lambda Vλ) antibody genes. Examples of how to generate the cDNA library of human antibody genes are illustrated in Example 1.

The coding sequences of human heavy and light chain genes may be mplified from the antibody cDNA library generated above by using PCR primer sets used in combination to prime the heavy chain variable region ($V_H$), the full heavy chain including both the variable and constant regions ($V_H+C_H$, $C_H$ including $C_H$**1**, $C_H$**2**, and $C_H$**3**), the light chain variable region ($V_L$, including Vλ and Vκ) and the full light chain including both the variable and constant region ($V_L+C_L$). The each of the PCR primers may include both an antibody partial sequence and a 5' or 3' flanking sequence for facilitating homologous recombination between the antibody fragments and a cloning expression vector. Examples of these primers are listed in Table 2.

2) Cloning of **V1** (or **V2**) into an Expression Vector in Bacteria Followed by Cloning **V2** (or **V1**) into the Vector via Homologous Recombination in Yeast

In another embodiment of the method for generating the library of expression vectors, the **V1** (or **V2**) sequences are cloned into an yeast-bacteria shuttle vector such as a modified vector derived from pACT2 (supplied by Clontech, Palo Alto, Calif.) in bacteria. The **V2** (or **V1**) sequences are then inserted into the library of expression vector comprising **V1** (or **V2**) via homologous recombination in yeast.

In one embodiment, the method comprises: transforming into yeast cells a library of insert nucleotide sequences that are linear and double-stranded, and a library of linearized yeast expression vectors, each having a 5'- and 3'-terminus sequence at the site of linearization.

The linearized yeast expression vectors of the vector library comprise a first polynucleotide sequence V1 encoding a first polypeptide subunit and varying within the vector library. The insert sequences of the insert library comprise a second nucleotide sequence V2 encoding a second polypeptide subunit and varying within the insert library. Each of the insert sequences also comprises a 5'- and 3'-flanking sequence at the ends of the insert sequence. The 5'- and 3'-flanking sequence of the insert sequence are sufficiently homologous to the 5'- and 3'-terminus sequences of the linearized yeast expression vector, respectively, to enable homologous recombination to occur.

In this embodiment, the first polypeptide subunit and the second polypeptide subunit are expressed as a single fusion protein. Also, the first and second nucleotide sequences each independently varies within the library of expression vectors.

According to the embodiment, the 5'- or 3'-flanking sequence of the insert nucleotide sequence is preferably between about 30–120 bp in length, more preferably between about 40–90 bp in length, and most preferably. between about 60–80 bp in length.

FIG. **3** illustrates an embodiment of this method according to the present invention. The coding sequences for **V1** (e.g., antibody heavy chain or light chain) are amplified by PCR to generate separate fragments which are directionally cloned into an expression vector in bacteria, resulting a library of expression vectors. The **V2** inserts are then cloned into this library of expression vectors through homologous recombination in yeast. The detailed procedures are described in Example 1.

As illustrated in FIG. **3**, the **V1** fragment has a restriction site its 5' terminus that matches with a restriction site at the 5' terminus of a linearized expression vector, and a restriction site its 3' terminus that matches with a restriction site at the 3' terminus of the linearized expression vector. By using a method of directional cloning, the **V1** fragments are ligated into the expression vectors to generate a library of vectors encoding **V1**. The resulting library of closed circular vectors are transformed into and propagated in bacteria. The **V1**-encoding vector library is dthen cleaved at a second linearization site, for example, a site downstream of **V1**. The **V2** fragment has a 5' flanking sequence and a 3' flanking sequence that are homologous to the 5' and 3' terminus of the linearized expression vector at the second linearization site. The **V2** fragment and the linearized expression vector are transformed into a yeast cell. Through homologous recombination in yeast, the **V2** fragment is inserted into the linearized expression vector at the second linearization site. As a result, a library of circular vectors carrying the variable sequences **V1** and **V2** is generated.

Each flanking sequence added to the 5' and 3'-terminus of **V2** sequence is preferably between about 30–120 bp in length, more preferably between about 40–90 bp in length, and most preferably between about 45–55 bp in length.

By using similar methods as described above, the variable sequences **V1** and **V2** can be inserted into an expression vector containing an activation domain (AD) or a DNA-binding domain (BD) of a transcription activator. The AD or BD domain may be positioned upstream or downstream of **V1** (or **V2**). It is preferred that the reading frames of the **V1** (or **V2**) fragments are conserved with the AD or BD reading frame.

The expression vector containing an AD (or BD) domain may be any vector engineered to carry the coding sequence of the AD domain. The expression vector is preferably a yeast vector such as pGAD10 (Feiloter et al. (1994) "Construction of an improved host strain for two hybrid screening" Nucleic Acids Res. 22: 1502–1503), pACT2 (Harper et al (1993) "The p21 Cdk-interacting protein Cip1 is a protein inhibitor of G1 cyclin-dependent kinase" Cell 75:805–816), and pGADT7 ("Matchmaker Gal4 two hybrid system 3 and libraries user manual" (1999), Clontech PT3247-1, supplied by Clontech, Palo Alto, Calif.).

The expression vector containing an AD (or BD) domain may also include another expression unit which is capable of expressing the second polypeptide subunit encoded by **V2**.

Expression of **V1** and/or **V2** may be separately under the transcriptional control of a constitutive promoter or an inducible promoter. One example of such an expression vector is available from Clontech, pBridge® (catalog No. 6184-1). The expression vector, pBridge®, contains one expression unit that controls expression of a Gal 4 BD domain and another expression unit that includes an inducible promoter Pmat25. Tirode, E. et al. (1997) J. Biol. Chem. 272:22995–22999.

The linearized vector DNA may be mixed with equal or excess amount of the **V1** or **V2** inserts. The linearized vector DNA and the inserts are co-transformed into host cells, such as competent yeast cells. Recombinant clones may be selected based on survival of cells in a nutritional selection medium or based on other phenotypic markers. Either the linearized vector or the insert alone may be used as a control for determining the efficiency of recombination and transformation.

Other homologous recombination systems may be used to generate the library of expression vectors of the present invention. For example, the recombination between the library of **V1** or **V2** sequences and the recipient expression vector may be facilitated by site-specific recombination.

The site-specific recombination employs a site-specific recombinase, an enzyme which catalyzes the exchange of DNA segments at specific recombination sites. Site-specific recombinases present in some viruses and bacteria, and have been characterized to have both endonuclease and ligase properties. These recombinases, along with associated proteins in some cases, recognize specific sequences of bases in DNA and exchange the DNA segments flanking those segments. Landy, A. (1993) Current Opinion in Biotechnology 3:699–707.

A typical site-specific recombinase is CRE recombinase. CRE is a 38-kDa product of the cre (cyclization recombination) gene of bacteriophage P1 and is a site-specific DNA recombinase of the Int family. Sternberg, N. et al. (1986) J. Mol. Biol. 187: 197–212. CRE recognizes a 34-bp site on the P1 genome called loxP (locus of X-over of P1) and efficiently catalyzes reciprocal conservative DNA recombination between pairs of loxP sites. The loxP site [SEQ ID NO: 1] consists of two 13-bp inverted repeats flanking an 8-bp nonpalindromic core region. CRE-mediated recombination between two directly repeated loxP sites results in excision of DNA between them as a covalently closed circle. Cre-mediated recombination between pairs of loxP sites in inverted orientation will result in inversion of the intervening DNA rather than excision. Breaking and joining of DNA is confined to discrete positions within the core region and proceeds on strand at a time by way of transient phophotyrosine DNA-protein linkage with the enzyme.

The CRE recombinase also recognizes a number of variant or mutant lox sites relative to the loxP sequence. Examples of these Cre recombination sites include, but are not limited to, the loxB, loxL and loxR sites which are found in the *E. coli* chromosome. Hoess et al. (1986) Nucleic Acid Res. 14:2287–2300. Other variant lox sites include, but are not limited to, loxB, loxL, loxR, loxP3, loxP23, loxΔ86, loxΔ117, loxP511 [SEQ ID NO:2], and loxC2 [SEQ ID NO:3]. Table 1 lists examples of lox sites that may be used in the present invention, including wild-type loxP sites LoxP WT [SEQ ID NO: 1] and loxP2 [SEQ ID NO: 5], and other loxP variants with mutations in the 13-bp inverted repeats region and/or the 8-bp nonpalindromic core region (underlined), loxP511 [SEQ ID NO: 2], loxC2 [SEQ ID NO: 3], loxP1 [SEQ ID NO: 4], loxP3 [SEQ ID NO: 6], loxP4 [SEQ ID NO: 7], loxP5 [SEQ ID NO: 8], loxP6 [SEQ ID NO: 9], loxP7 [SEQ ID NO: 10], loxP8 [SEQ ID NO: 11], loxP9 [SEQ ID NO: 12], and loxP10 [SEQ ID NO: 13].

Examples of the non-CRE recombinases include, but are not limited to, site-specific recombinases include: att sites recognized by the Int recombinase of bacteriophage λ (e.g. att1, att2, att3, attp, attB, attL, and attR), the FRT sites recognized by FLP recombinase of the 2 pi plasmid of *Saccharomyces cerevisiae*, the recombination sites recognized by the resolvase family, and the recombination site recognized by transposase of *Bacillus thruingiensis.*

Subsequent analysis may also be carried out to determine the efficiency of homologous recombination that results in correct insertion of the **V1** and **V2** sequences into the expression vector. For example, PCR amplification of the V1 or V2 inserts directly from the selected yeast clone may reveal how many clones are recombinant. Libraries with minimum of 90% recombinant clones are preferred. The same PCR amplification of selected clones may also reveal the insert size. Although a small fraction of the library may contain double or triple inserts, the majority (>90%) is preferably to have a single insert with the expected size.

To verify sequence diversity of the inserts in the selected clones, PCR amplification product with the correct size of insert may be fingerprinted with frequent digesting restriction enzymes. From a gel electrophoresis pattern, it may be determined whether the clones analyzed are of the same identity or of the distinct or diversified identity. The PCR products may also be sequenced directly to reveal the identity of inserts and the fidelity of the cloning procedure and to prove the independence and diversity of the clones.

In an embodiment where the V1 and V2 sequences are the coding sequences for a heavy chain and a light chain derived from a human antibody repertoire, respectively, monoclonal antibody may be generated from hybridoma cell lines as controls by following the same procedures described above. Examples of hybridoma cell lines include, but are not limit to, anti-GFP antibody producing cell line (Clontech), anti-p53 antibodies producing cell lines (NeoMarker), and other hybridoma cell lines available from ATCC (Atlanta). The hybridoma cell line is subjected to the same procedures described above, i.e., RNA isolation, cDNA synthesis, PCR amplification, and homologous recombination into yeast. Other antibody libraries may also be generated from mouse fetal liver and fetal spleen using the same principle.

The mouse antibody library generated can provide a direct control for existing individual mouse monoclonal antibody with its cognate antigen. Most studies for antigen-antibody interaction have been performed with mouse antibodies. The mouse antibody library should serve as an excellent control in the selection of human antibody library against a target antigen by yeast two-hybrid method described below.

### 3. Selection of Affinity Binding Pairs between the Library of Fusion Proteins of the Present Invention and Target Proteins

The present invention also provides methods for screening protein-protein or protein-peptide binding pairs in a yeast two-hybrid system.

The two-hybrid system is a selection scheme designed to screen for polypeptide sequences which bind to a predetermined polypeptide sequence present in a fusion protein. Chien et al. (1991) Proc. Natl. Acad. Sci. (USA) 88: 9578). This approach identifies protein-protein interactions in vivo through reconstitution of a transcriptional activator. Fields and Song (1989) Nature 340: 245), the yeast Gal 4 transcription protein. The method is based on the properties of the yeast Gal 4 protein, which consists of separable domains responsible for DNA-binding and transcriptional activation. Polynucleotides encoding two hybrid proteins, one consisting of the yeast Gal 4 DNA-binding domain (BD) fused to a polypeptide sequence of a known protein and the other consisting of the Gal4 activation domain (AD) fused to a polypeptide sequence of a second protein, are constructed and introduced into a yeast host cell. Intermolecular binding between the two fusion proteins reconstitutes the Gal4 DNA-binding domain with the Gal4 activation domain, which leads to the transcriptional activation of a reporter gene (e.g., lacZ, HIS3) which is operably linked to a Gal4 binding site.

Typically, the two-hybrid method is used to identify novel polypeptide sequences which interact with a known protein. Silver and Hunt (1993) Mol. Biol. Rep. 17: 155; Durfee et al. (1993) Genes Devel. 7; 555; Yang et al. (1992) Science 257: 680; Luban et al. (1993) Cell 73: 1067; Hardy et al. (1992) Genes Devel. 6; 801; Bartel et al. (1993) Biotechniques 14: 920; and Vojtek et al. (1993) Cell 74: 205. The two-hybrid system was used to detect interactions between three specific single-chain variable fragments (scFv) and a specific antigen. De Jaeger et al. (2000) FEBS Left. 467:316–320. The two-hybrid system was also used to screen against cell surface proteins or receptors such as receptors of hematopoietic super family in yeast. Ozenberger, B. A., and Young, K. H. (1995) "Functional interaction of ligands and receptors of hematopoietic superfamily in yeast" Mol Endocrinol. 9:1321–1329.

Variations of the two-hybrid method have been used to identify mutations of a known protein that affect its binding to a second known protein Li and Fields (1993) FASEB J. 7: 957; Lalo et al. (1993) Proc. Natl. Acad. Sci. (USA) 90: 5524; Jackson et al. (1993) Mol. Cell. Biol. 13; 2899; and Madura et al. (1993) J. Biol. Chem. 268: 12046.

Two-hybrid systems have also been used to identify interacting structural domains of two known proteins or domains responsible for oligomerization of a single protein. Bardwell et al. (1993) Med. Microbiol. 8: 1177; Chakraborty et al. (1992) J. Biol. Chem. 267: 17498; Staudinger et al. (1993) J. Biol. Chem. 268: 4608; and Milne GT; Weaver DT (1993) Genes Devel. 7; 1755; Iwabuchi et al. (1993) Oncogene 8; 1693; Bogerd et al. (1993) J. Virol. 67: 5030).

Variations of two-hybrid systems have been used to study the in vivo activity of a proteolytic enzyme. Dasmahapatra et al. (1992) Proc. Natl. Acad. Sci. (USA) 89: 4159. Alternatively, an *E. coli*/BCCP interactive screening system was used to identify interacting protein sequences (i.e., protein sequences which heterodimerize or form higher order heteromultimers). Germino et al. (1993) Proc. Natl. Acad. Sci. (U.S.A.) 90: 933; and Guarente L (1993) Proc. Natl. Acad. Sci. (U.S.A.) 90: 1639.

Typically, selection of binding protein using a two-hybrid method relies upon a positive association between two Gal4 fusion proteins, thereby reconstituting a functional Gal4 transcriptional activator which then induces transcription of a reporter gene operably linked to a Gal4 binding site. Transcription of the reporter gene produces a positive readout, typically manifested either (1) as an enzyme activity (e.g., β-galactosidase) that can be identified by a calorimetric enzyme assay or (2) as enhanced cell growth on a defined medium (e.g., HIS3 and Ade 2). Thus, the method is suited for identifying a positive interaction of polypeptide sequences, such as antibody-antigen interactions.

False positives clones that indicate activation of the reporter gene irrespective of the specific interaction between the two hybrid proteins, may arise in the two-hybrid screening. Various procedures have developed to reduce and eliminate the false positive clones from the final positives. For example, 1) prescreening the clones that contains the target vector and shows positive in the absence of the two-hybrid partner (Bartel, P. L., et al. (1993) "Elimination of false positives that arise in using the two-hybrid system" BioTechniques 14:920–924); 2) by using multiple reporters such as His3, β-galactosidase, and Ade2 (James, P. et al. (1996) "Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast" Genetics 144:1425–1436); 3) by using multiple reporters each of which is under different GAL 4 -responsive promoters such as those in yeast strain Y190 where each of the His 3 and β-Gal reporters is under the control of a different promoter

Gal 1 or Gal 10, but both response to Gal 4 signaling (Durfee, T., et al (1993) “The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit” Genes Devel. 7:555–569); and 4) by post-screening assays such as testing isolates with target consisting of GAL 4-BD alone.

In addition, the false positive clones may also be eliminated by using unrelated targets to confirm specificity. This is a standard control procedure in the two-hybrid system which can be performed after the library isolate is confirmed by the above-described 1)–4) procedures. Typically, the library clones are confirmed by co-transforming the initially isolated library clones back into the yeast reporter strain with one or more control targets unrelated to the target used in the original screening. Selection is conducted to eliminate those library clones that show positive activation of the reporter gene and thus indicate non-specfic interactions with multiple, related proteins.

The present invention provides efficient methods for screening the polypeptide encoded by **V1** and **V2** in the library of expression vectors for their affinity binding to one or more target proteins.

According to the present invention, the method comprises:

expressing a library of tester protein complexes in yeast cells, each tester protein complexes being formed between a first polypeptide subunit whose sequence varies within the library, and a second polypeptide subunit whose sequence varies within the library independently of the first polypeptide; expressing one or more target fusion proteins in the yeast cells expressing the tester proteins, each of the target fusion proteins comprising a target peptide or protein; and

selecting those yeast cells in which a reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target fusion protein.

According to the method, the diversity of the first or the second polypeptide subunit is preferably between $10^3$–$10^8$, more preferably between $10^4$–$10^8$, and most preferably between $10^5$–$10^8$.

Also according to the method, the diversity of the protein complexes encoded by the library of expression vectors is preferably between $10^6$–$10^{18}$, more preferably between $10^9$–$10^{18}$ and most preferably between $10^{10}$–$10^{18}$.

A feature of the present invention is that the first and second polypeptide subunits may be selected entirely independent of the target peptide or protein and need not be based on in any way upon one or more proteins known to the bind to the target. As a result, the diversities of the first and second polypeptide subunits may be each independently derived from libraries of precursor sequences that are not specifically designed for the target peptide or protein. For example, the libraries of precursor sequences need not be derived from a small group (e.g. 2–20) of genes with predetermined sequences and encoding proteins that are known to the bind the target peptide or protein.

The diversities of the first and second polypeptide subunits also need not be derived from one or more proteins that are known to bind to the target peptide or protein. For example, the one or more proteins need not be derived from a small group (e.g. 2–20) of proteins with predetermined sequences that are known to bind to the target peptide or protein.

The diversities of the first and second polypeptide subunits also need not be generated by mutagenizing one or more proteins that are known to bind to the target peptide or protein. For example, the first and second polypeptide subunits need not be generated by mutagenizing a small group (e.g. 2–20) of proteins with predetermined sequences and known to bind to the target peptide or protein.

In a variation of the embodiment, a single target fusion protein is expressed and screened against the library of tester proteins. According to the variation, the step of expressing the library of tester protein complexes may include transforming a library of tester expression vectors into the yeast cells which contain a reporter construct comprising the reporter gene whose expression is under transcriptional control of a transcription activator comprising an activation domain and a DNA binding domain.

Each of the tester expression vectors comprises a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator, a first nucleotide sequence encoding the first polypeptide subunit, and a second nucleotide sequence encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The domain encoded by the first transcription sequence and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

Optionally, the step of expressing the target protein complexes includes transforming a target expression vector into the yeast cells simultaneously or sequentially with the library of tester expression vectors. The target expression vector comprises a second transcription sequence encoding either the activation domain AD or the DNA binding domain BD of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide.

FIG. **4** illustrates a flow diagram of a preferred embodiment of the above described method. As illustrated in FIG. **4**, the sequence library containing **V1** fused with an AD domain upstream and **V2** is carried by a library of expression vectors, the AD-**V1**/**V2** vectors. The coding sequence of the target protein (labeled as “Target”) is contained in another expression vector and fused with a BD domain, forming the BD-Target vector.

The AD-**V1**/**V2** vector and the BD-Target vector may be co-transformed into a yeast cell by using method known in the art. Gietz, D. et al. (1992) “Improved method for high efficiency transformation of intact yeast cells” Nucleic Acids Res. 20:1425. The construct carrying the specific DNA binding site and the reporter gene (labeled as “Reporter”) may be stably integrated into the genome of the host cell or transiently transformed into the host cell. Upon expression of the sequences in the expression vectors, the library of protein complexes comprising AD-**V1** fusion and **V2**, labeled as the AD-**V1**/**V2** protein complexes, undergo protein folding in the host cell and adopt various conformations. Some of the AD-**V1**/**V2** protein complexes may bind to the Target protein expressed by the BD-Target vector in the host cell, thereby bringing the AD and BD domains to a close proximity in the promoter region (i.e., the specific DNA binding site) of the reporter construct and thus reconstituting a functional transcription activator composed of the AD and BD domains. As a result, the AD activates the transcription of the reporter gene downstream from the specific DNA binding site, resulting in expression of the reporter gene, such as the lacZ reporter gene. Clones showing the phenotype of the reporter gene expression are selected, and the AD-V1/V2 vectors are isolated. The coding sequences for V1 and V2 are identified and characterized.

Alternatively, the steps of expressing the library of tester protein complexes and expressing the target fusion protein includes causing mating between first and second populations of haploid yeast cells of opposite mating types.

The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester protein complexes. Each of the tester expression vector comprises a first transcription sequence encoding either the activation domain AD or the DNA binding domain BD of the transcription activator, a first nucleotide sequence V1 encoding the first polypeptide subunit, and a second nucleotide sequence V2 encoding the second polypeptide subunit.

The second population of haploid yeast cells comprises a target expression vector. The target expression vector comprises a second transcription sequence encoding either the activation domain AD or the DNA binding domain BD of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide. Either the first or second population of haploid yeast cells comprises a reporter construct comprising the reporter gene whose expression is under transcriptional control of the transcription activator.

In this method, the haploid yeast cells of opposite mating types may preferably be α and a type strains of yeast. The mating between the first and second populations of haploid yeast cells of α and a-type strains may be conducted in a rich nutritional culture medium.

FIG. **5** illustrates a flow diagram of a preferred embodiment of the above described method. As illustrated in FIG. **5**, the sequence library containing V1 fused with an AD domain upstream and V2 is carried by a library of expression vectors, the AD-V1/V2 vectors. The library of the AD-V1/V2 vectors are transformed into haploid yeast cells such as the a type strain of yeast.

The coding sequence of the target protein (labeled as "Target") is contained in another expression vector and fused with a BD domain, forming the BD-Target vector. The BD-Target vector is transformed into haploid cells of opposite mating type of the haploid cells containing the AD-V1/V2 vectors, such as the α type strain of yeast. The construct carrying the specific DNA binding site and the reporter gene (labeled as "Reporter") may be transformed into the haploid cells of either the type a or type α strain of yeast.

The haploid cells of the type a and type α strains of yeast are mated under suitable conditions such as low speed of shaking in liquid culture, physical contact in solid medium culture, and rich medium such as YPD. Bendixen, C. et al. (1994) "A yeast mating-selection scheme for detection of protein-protein interactions", Nucleic Acids Res. 22: 1778–1779. Finley,Jr., R. L. & Brent, R. (1994) "Interaction mating reveals lineary and ternery connections between Drosophila cell cycle regulators", Proc. Natl. Acad. Sci. USA, 91:12980–12984. As a result, the AD-V1/V2, the BD-Target expression vectors and the Reporter construct are taken into the parental diploid cells of the a and type α strain of haploid yeast cells.

Upon expression of the sequences in the expression vectors in the parental diploid cells, the library of protein complexs formed between AD-V1 fusion and V2, labeled as the AD-V1/V2 protein complexes, undergo protein folding in the host cell and adopt various conformations. Some of the AD-V1/V2 protein complexes may bind to the Target protein expressed by the BD-Target vector in the parental diploid cell, thereby bringing the AD and BD domains to a close proximity in the promoter region (i.e., the specific DNA binding site) of the reporter construct and thus reconstituting a functional transcription activator composed of the AD and BD domains. As a result, the AD activates the transcription of the reporter gene downstream from the specific DNA binding site, resulting in expression of the reporter gene, such as the lacZ reporter gene. Clones showing the phenotype of the reporter gene expression are selected, and the AD-V1/V2 vectors are isolated. The coding sequences for V1 and V2 are identified and characterized.

A wide variety of reporter genes may be used in the present invention. Examples of proteins encoded by reporter genes include, but are not limited to, easily assayed enzymes such as β-galactosidase, α-galactosidase, luciferase, β-glucuronidase, chloramphenicol acetyl transferase (CAT), secreted embryonic alkaline phosphatase (SEAP), fluorescent proteins such as green fluorescent protein (GFP), enhanced blue fluorescent protein (EBFP), enhanced yellow fluorescent protein (EYFP) and enhanced cyan fluorescent protein (ECFP); and proteins for which immunoassays are readily available such as hormones and cytokines. The expression of these reporter genes can also be monitored by measuring levels of mRNA transcribed from these genes.

When the screening of the V1 and V2 library is conducted in yeast cells, certain reporter(s) are of nutritional reporter which allows the yeast to grow on the specific selection medium plate. This is a very powerful screening process, as has been shown by many published papers. Examples of the nutritional reporter include, but are not limited to, His3, Ade2, Leu2, Ura3, Trp1 and Lys2. The His3 reporter is described in Bartel, P. L. et al. (1993) "Using the two-hybrid system to detect protein-protein interactions", in Cellular interactions in Development: A practical approach, ed. Hastley, D. A., Oxford Press, pages 153–179. The Ade2 reporter is described in Jarves, P. et al. (1996) "Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast" Genetics 144:1425–1436.

For example, a library of antibody expression vectors may be transformed into haploid cells of the α mating type of yeast strain. The antibody expression vector may contain an antibody light chain fused with an AD domain of GAL 4 transcription activator and an antibody heavy chain expressed from a separate expression cassette in the vector. A BD domain of GAL 4 transcription activator is fused with the sequence encoding the target protein to be selected against the antibody library in a plasmid. This plasmid is transformed into haploid cells of the a mating type of yeast strain.

Equal volume of AD-Antibody library-containing yeast stain (α-type) and the BD-target-containing yeast strain (a-type) are inoculated into selection liquid medium and incubated separately first. These two cultures are then mixed and allowed to grow in rich medium such as 1×YPD and 2×YPD. Under the rich nutritional culture condition, the two haploid yeast strains will mate and form diploid cells. At the end of this mating process, these yeast cells are plated into selection plates. A multiple-marker selection scheme may be used to select yeast clones that show positive interaction between the antibodies in the library and the target. For example, a scheme of SD/-Leu-Trp-His-Ade may be used. The first two selections (Leu-Trp) are for markers (Leu and Trp) expressed from the AD-Antibody library and the BD-Target vector, respectively. Through this dual-marker selection, diploid cells retaining both BD and AD vectors in the same yeast cells are selected. The latter two markers, His-Ade, are used to screen for those clones that express the reporter gene from parental strain, presumably due to affinity binding between the antibodies in the library and the target.

After the screening by co-transformation, or by mating screening as described above, the putative interaction between the gene probe and the library clone isolates can be further tested and confirmed in vitro or in vivo.

In vitro binding assays may be used to confirm the positive interaction between the tested protein expressed by the clone isolate and the target protein or peptide. For example, the in vitro binding assay may be a "pull-down" method, such as using GST (glutathione S-transferase)-fused gene probe as matrix-binding protein, and with in vitro expressed library clone isolate that are labeled with a radioactive or non-radioactive group. While the probe is bound to the matrix through GST affinity substrate (glutathione-agarose), the library clone isolate will also bind to the matrix through its affinity with the gene probe. The in vitro binding assay may also be a co-immuno-precipitation (Co-IP) method using two affinity tag antibodies. In this assay, both the target gene probe and the library clone isolate are in vitro expressed fused with peptide tags, such as HA (haemaglutinin A) or Myc tags. The gene probe is first immuno-precipitated with an antibody against the affinity peptide tag (such as HA) that the target gene probe is fused with. Then the second antibody against a different affinity tag (such as Myc) that is fused with the library clone isolate is used for reprobing the precipitate.

In vivo assays may also be used to confirm the positive interaction between the tested protein expressed by the clone isolate and the target protein or peptide. For example, a mammalian two-hybrid system may serve as a reliable verification system for the yeast two-hybrid library screening. In this system, the target gene probe and the library clone are fused with Gal 4 DNA-binding domain or a mammalian activation domain (such as VP-16) respectively. These two fusion proteins under control of a strong and constitutive mammalian promoter (such as CMV promoter) are introduced into mammalian cells by transfection along with a reporter responsive to Gal 4. The reporter can be CAT gene (chloramphenical acetate transferase) or other commonly used reporters. After 2–3 days of transfection, CAT assay or other standard assays will be performed to measure the strength of the reporter which is correlated with the strength of interaction between the gene probe and the library clone isolate.

The present invention also provides a kit for selecting selecting tester proteins capable of binding to a target peptide or protein.

In an embodiment, the kit comprises: a library of tester expression vectors and a yeast cell line. Each of the tester expression vectors comprises a first transcription sequence encoding either an activation domain or a DNA binding domain of a transcription activator, a first nucleotide sequence encoding a first polypeptide subunit, and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequences each independently varying within the library of expression vectors. The first and second polypeptide subunits are expressed as separate proteins and form a protein complex upon interacting with each other. A reporter construct may be contained in the yeast cell line. The reporter construct comprises a reporter gene whose expression is under a transcriptional control of a specific DNA binding site.

Optionally, the kit may further comprise a target expression vector which comprises a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide.

In another embodiment, the kit comprises: a first and second populations of haploid yeast cells of opposite mating types. The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vector comprises a first transcription sequence encoding either an activation domain or a DNA binding domain of a transcription activator, a first nucleotide sequence encoding a first polypeptide subunit, and a second nucleotide sequence encoding a second polypeptide subunit, the first and second nucleotide sequences each independently varying within the library of expression vectors. The first and second polypeptide subunits are expressed as separate proteins and form a protein complex upon interacting with each other. The second population of haploid yeast cells comprises a target expression vector. The target expression vector encodes either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide. Either the first or second population of haploid yeast cells comprises a reporter construct comprising a reporter gene whose expression is under transcriptional control of the transcription activator.

Optionally, the second population of haploid yeast cells comprises a plurality of target expression vectors. Each of the target expression vectors encodes either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and a target sequence encoding the target protein or peptide. Either the first or second population of haploid yeast cells comprises a reporter construct comprising a reporter gene whose expression is under transcriptional control of the transcription activator.

According to the present invention, other yeast two-hybrid systems may be employed, including but not limited to SOS-RAS system (SRS), Ras recruitment system (RRS), and ubiquitin split system. Brachmann and Boeke (1997) "Tag games in yeast: the two-hybrid system and beyond" Current Opinion Biotech. 8:561–568. In these non-conventional yeast two-hybrid systems, the first or second polypeptide subunit may further comprise a signaling domain for screening the library of the protein complexes based these non-conventional two-hybrid methods. Examples of such signaling domain includes but are not limited to a Ras guanyl nucleotide exchange factor (e.g. human SOS factor), a membrane targeting signal such as a myristoylation sequence and farnesylation sequence, mammalian Ras lacking the carboxy-terminal domain (the CAAX box), and a ubiquitin sequence.

SRS and RRS systems are alternative two-hybrid systems for studying protein-protein interaction in cytoplasm. Both systems use a yeast strain with temperature-sensitive mutation in the cdc25 gene, the yeast homologue of human Sos (hSos). This protein, a guanyl nucleotide exchange factor, binds and activates Ras, that triggers the Ras signaling pathway. The mutation in the cdc25 protein is temperature sensitive; the cells can grow at 25° C. but not at 37° C. In the SRS system, this cdc25 mutation is complemented by the hSos gene product to allow growth at 37° C., providing that the hSos protein is localized to the membrane via a protein-protein interaction (Aronheim et al. 1997, Mol. Cel. Biol. 17:3094–3102). In the RRS system, the mutation is complemented by a mammalian activated Ras with its CAAX box at its carboxy terminus upon recruitment to the plasma membrane via protein-protein interaction (Broder et al, 1998, Current Biol. 8:1121–1124).

For example, the library of expression vectors encoding human antibody library can be constructed for the selection based on the SRS system. A vector, pMyr (Stratagene, CA), is modified by replacing the f1 origin region of pMyr expression cassette with MET25 promoter and PGK terminator from pBridge-1 (described in EXAMPLE) through homologous recombination, resulting in pMyr-DC. The light chain sequence is cloned into the MCS site downstream from myristoylation signal sequence using ligation-based approach. The heavy chain us cloned into the MCS downstream from the MET25 promoter by homologous recombination. The library is made in the mutant cdc25H α strain (Stratagene, CA). The myristoylation signal anchors the antibody fusion proteins to the plasma membrane. DNA encoding the target protein is cloned into the MCS of pSos vector, which is available from Stratagene. Such construct expresses a fusion protein of hSos and the target protein.

The antibody library can be screened by co-transformation of the pSos with the target sequence into the cdc25H α strain. The transformed yeast cells are incubated under the restrictive temperature of 37° C. on the yeast medium plate with galactose and low concentration of methionine, since the antibody expressions are under the controls of GAL1 and MET25 promoters, respectively.

The antibody library can also be screened by yeast mating. The pSos vector with bait sequence is first transformed into cdc25H a strain (available from Stratagene). The transformed a strain is then mated with the α strain containing the antibody library, followed by incubation of the mated yeast cells incubated under the restrictive temperature of 37° C. on the yeast medium plate with galactose and low concentration of methionine.

Alternatively, the antibody library can be made in the modified pSos. The target protein is cloned into the pMyr. Library screening can be performed similarly either by co-transformation or by mating.

### 4. Selection of Affinity Binding Pairs between the Library of Protein Complexes of the Present Invention and Target Nucleic Acids

As described above, the libraries of V1 and V2 sequences of the present invention can be used for selecting protein-protein or protein-peptide binding pairs against single or arrayed multiple protein/peptide targets in a two-hybrid screening system. As described in the following, these libraries can also be used for selecting protein-DNA or protein-RNA binding pairs in a one-hybrid system or three-hybrid system, respectively.

The general scheme for screening protein-DNA binding pair using an one-hybrid system is described in Li and Herskowitz (1993) Science 262:1870–1874. Typically, this method is used to identify genes encoding proteins that recognize a specific DNA sequence. A library of random protein segments tagged with a transcriptional activation domain (AD) is screened for proteins that can activate a reporter gene containing the specific DNA sequence in its promoter region. By using this strategy, an essential protein that interacts in vivo with the yeast origin of DNA replication was identified. In a three-hybrid system, the target nucleic acid is RNA or RNA-associated proteins. SanGupta, et al. (1996) Proc. Natl. Acad. Sci. USA 93:8496–8501.

The present invention provides a method is provided for screening protein-DNA binding pairs in a yeast one-hybrid system.

In an embodiment, the method comprises: expressing a library of tester protein complexes in yeast cells which contain a reporter construct comprising a reporter gene whose expression is under a transcriptional control of a target DNA sequence; and selecting the yeast cells in which the reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target DNA sequence.

In a variation of the embodiment, the step of expressing the library of tester protein complexes includes transforming into the yeast cells a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vector comprises a transcription sequence encoding an activation domain of a transcription activator, a first nucleotide sequence **V1** encoding the first polypeptide subunit, and a second nucleotide sequence **V2** encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The transcriptional activation domain AD and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

In another variation of the embodiment, the step of expressing a library of tester protein complexes in yeast cells includes causing mating between a first and second populations of haploid yeast cells of opposite mating types. The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester protein complexes described above. The second population of haploid yeast cells comprises the reporter construct.

According to the variation, the haploid yeast cells of opposite mating types may preferably be α and a type strains of yeast. The mating between the first and second populations of haploid yeast cells of α and a type strains may preferably conducted in a rich nutritional culture medium.

According to any of the above-described methods for selecting protein-DNA binding pairs, the target DNA sequence in the reporter construct may preferably be positioned in 2–6 tandem repeats 5' relative to the reporter gene.

The target DNA sequence in the reporter construct may be preferably between about 15–75 bp in length and more preferably between about 25–55 bp in length.

FIG. **6** illustrates a flow diagram of a preferred embodiment of the above-described method. As illustrated in FIG. **6**, the tester sequence library containing V1 fused with an AD domain upstream and V2 is carried by a library of expression vectors, the AD-V1/V2 vector. The target DNA sequence (labeled "Target DNA") is positioned in the promoter region of a reporter gene (labeled "Reporter").

The AD-V1/V2 vector is transformed into a yeast cell by using methods known in the art. Gietz, D. et al. (1992) "Improved method for high efficiency transformation of intact yeast cells" Nucleic Acids Res. 20:1425. The construct carrying the target DNA sequence and the reporter gene may be stably integrated into the genome of the host cell or transiently transformed into the host cell.

As illustrated in FIG. **6**, upon expression of the tester sequences in the expression vectors, the library of tester protein complexes formed between AD-V1 fusion and V2, labeled as the AD-V1/V2 fusion protein complexes, undergo protein folding in the host cell and adopt various conformations. Some of the AD-V1/V2 protein complexes may bind to the target DNA sequence in the promoter region of the reporter gene, thereby bringing the AD domain to a close proximity in the promoter region. As a result, the AD activates the transcription of the reporter gene downstream from the target DNA sequence, resulting in expression of the reporter gene, such as the lacZ reporter gene. Clones showing the phenotype of the reporter gene expression are selected, and the AD-V1/V2 vectors are isolated. The coding sequences for V1 and V2 are identified and characterized.

Alternatively, the AD-V1/V2 vector and the reporter construct may be introduced a diploid yeast cell by mating between two haploid yeast strains. For example, the AD-V1/V2 vector may be transformed into a haploid yeast strain such as the α strain; and the reporter construct may be transformed into another haploid yeast strain such as the a strain. Upon mating between these two haploid strains, diploid cells are formed to merge the genetic materials carried by the two haploid cells. As a result, the AD-V1/V2 vector and the reporter construct are introduced into a diploid cell which is then screened for positive interactions between the tester protein and the target DNA in the cell.

The target DNA sequence may be a regulatory element, or a putative chromosome remodeling protein complex opening site, preferably in a short stretch of DNA sequence (20–80 bp). The target DNA sequence may be cloned into a yeast one-hybrid system reporter vector, e.g., pHIS (Clontech, Palo Alto, Calif.; Luo et al. (1996) "Cloning and analysis of DNA-binding proteins by yeast one-hybrid and one-two-hybrid system" Biotechniques 20:564–568). To increase the sensitivity, the target sequence may be cloned as in a few tandem repeats (e.g., 4–5 copies) into the reporter vector. The recombinant reporter vector may be integrated into the yeast reporter strain by a transformation with linearized vector and selection for rescuing the integration marker. The integration should be at a single chromosome location and usually at high efficiency.

The tester sequence library containing V1 and V2 may encode an antibody library that can be used to screen against a target DNA antigen. The antibody expression library may be introduced into yeast by transformation or by mating with the yeast strain of the opposite mating type and harboring the reporter construct. The transformation and mating procedures are described in detail in Example 3. Pre-screening of self-activating clones may be necessary for eliminating the false positive clones. The procedures are similar to the two-hybrid library pre-screening described in Section 3.

The library clones isolated from such a one-hybrid system screening may indicate that antibody(s) expressed from these clones are capable of binding to the DNA target. Such antibody may be have significant applications in DNA vaccine and diagnostics of diseases.

The one-hybrid system of the present invention may also be modified to screen for novel co-factors that bind to a known DNA-binding factor. The library of protein complexes formed between AD-V1 fusion and V2 subunit may be screened for affinity binding toward a specific factor that binds to a DNA sequence in the promoter region of a reporter gene.

In yet another embodiment, a method is provided for screening protein-protein binding pairs in a yeast one-hybrid system. The method comprises: expressing a library of tester protein complexes in yeast cells which contain a reporter construct comprising a reporter gene whose expression is under a transcriptional control of a specific DNA binding site; expressing a target protein in the yeast cells expressing the tester protein complexes, where the target protein binds to the specific DNA binding site; and selecting the yeast cells in which the reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target protein.

In a variation of the embodiment, the step of expressing the library of tester protein complexes includes transforming into the yeast cells a library of tester expression vectors for the library of tester fusion proteins. Each of the tester expression vector comprises a transcription sequence encoding an activation domain of a transcription activator, a first nucleotide sequence V1 encoding the first polypeptide subunit, and a second nucleotide sequence V2 encoding the second polypeptide subunit, the first and second nucleotide sequences varying independently within the library of tester expression vectors. The transcriptional activation domain AD and the first polypeptide subunit are expressed as a fusion protein. The first and second polypeptide subunits are expressed as separate proteins, and form the tester protein complex upon binding with each other through non-covalent interactions (e.g. hydrophobic interactions) or covalent interactions (e.g. disulfide bonds).

In another variation of the embodiment, the steps of expressing the library of tester protein complexes and expressing the target fusion protein includes causing mating between a first and second populations of haploid yeast cells of opposite mating types. The first population of haploid yeast cells comprises a library of tester expression vectors for the library of tester protein complexes described above. The second population of haploid yeast cells comprises a target expression vector comprising a target sequence encoding the target protein. Either the first or second population of haploid yeast cells comprises the reporter construct.

FIG. **7** illustrates a flow diagram of a preferred embodiment of the above-described method. As illustrated in FIG. **8**, the tester sequence library containing V1 fused with an AD domain upstream (AD-V1 fusion) and V2 is carried by a library of expression vectors, the AD-V1/V2 vector. The AD-V1/V2 vectors are introduced into host cells, for example, by transformation. The target protein (labeled "Target") that is known to bind to a specific DNA sequence may be expressed by an expression vector in the host cells or otherwise present in the cells. The specific DNA sequence (labeled "*DNA") is positioned in the promoter region of a reporter gene (labeled "Reporter"). The construct carrying the specific DNA sequence and the reporter gene may be stably integrated into the genome of the host cell or transiently transformed into the host cell.

As illustrated in FIG. **7**, upon expression of the tester sequences in the expression vectors, the library of tester protein complexes formed between AD-V1 fusion and V2, labeled as the AD-V1/V2 protein complexes, undergo protein folding in the host cell and adopt various conformations. Some of the AD-V1/V2 fusion proteins may bind to the target protein that binds to the specific DNA sequence in the promoter region of the reporter gene, thereby bringing the AD domain to a close proximity in the promoter region. As a result, the AD activates the transcription of the reporter gene downstream from the target DNA sequence, resulting in expression of the reporter gene, such as the lacZ reporter gene. Clones showing the phenotype of the reporter gene expression are selected, and the AD-V1/V2 vectors are isolated. The coding sequences for V1 and V2 are identified and characterized.

The specific target protein may be any protein that has been characterized to be a DNA-binding fact by using various assays such as in vitro gel shifting assays, or through conventional one-hybrid screening. The target protein (without being fused to an AD domain) may be expressed in the yeast one-hybrid reporter strain. The level of target protein expression is then adjusted to such an extent that no measurable activation is observed. The yeast strain may also contain the reporter construct that is integrated into the yeast genome.

The tester sequence library containing **V1** and **V2** may encode a library of antibody that can be used to screen against a target protein that a DNA-binding factor. The library clones isolated from such a modified one-hybrid system screening may indicate that antibody(s) expressed from these clones are capable of binding to the protein target. Such antibody may be have significant applications in therapeutics and diagnostics of diseases.

### 5. High Throughput Selection of Affinity Binding Pairs between the Library of Protein Complexes of the Present Invention and a Library of Target Proteins

The present invention also provides a method for high throughput screening of the above-described libraries of protein complexes encoded by **V1** and **V2**. The library of expression vectors, for example, the AD-antibody yeast expression vector library, may be screen for the binding of the antibodies to multiple target proteins expressed by a yeast clone library (BD-Target library), each clone carrying a BD-Target vector for each target protein to be selected against. The BD-Target clone library may be arrayed in multiple-well plates, such as 96- and 384-well plates, and then screened against the antibody library in an automated and high throughput manner.

For example, a collection of EST clones (or a total library of EST) from human, mouse or other organisms may be screened against the antibody library generated by using the methods of the present invention. Such a collection of EST clones may be ordered from a public resource in a library format with individually clones arrayed in 96-well or 384-well plates. Lennon, G. et al. (1996) "The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression" Genomics 33:151–152. The EST inserts from the original collection (usually in bacterial cloning and sequencing vectors) may be PCR amplified with extended homologous sequences at both ends following similar procedures used in the generation of the antibody library. Through the same homologous recombination procedure as used in the generation of the antibody library, the EST inserts are inserted into an expression vector containing a BD domain of a transcription activator in yeast cells.

Optionally, a collection of certain domain structures, such as zinc finger and helix-loop-helix protein domains, may be inserted into the BD-containing expression vector in yeast cell via homologous recombination. The yeast clones containing the vector with BD fused to each domain structure may be arrayed in multiple-well plates and screened against the antibody library for affinity binding between the antibody and each domain structure. The domain structure may be 18–20 amino acids at length and its sequence may not be totally random. Such a collection of domain structures may be generated by using synthetic oligonucleotides with characteristic conserved and random/degenerate residues to cover most of the rational domain structures.

Also optionally, the coding sequences of a random peptide library may be inserted into the BD-containing expression vector in yeast cell via homologous recombination. The yeast clones containing the vector with BD fused to each random peptide may be arrayed in multiple-well plates and screened against the antibody library for affinity binding between the antibody and each random peptide target. The random peptide may be 16–20 amino acid at length. Such a library of random peptide can generated by random oligonucleotide synthesis or by partially random oligonucleotide synthesis biased toward a sequence encoding a specific target.

Alternatively, a library of short peptides may also be may be inserted into the BD-containing expression vector in yeast cell via homologous recombination. Accordingly, the antibody library may be fused with the BD domain in the expression vector and screened against this library of short peptide. Through this selection, peptide ligands may be selected for each antibody. Structural and functional analysis of the selected peptides should aid in the rational design of antigens and structural improvement of specific target antigens.

FIG. **8** depicts a general scheme of high throughput screening of a library of V/**V2** protein complexes against a library of target proteins in yeast via mating of two strains of yeast haploid cells.

As illustrated in FIG. **8**, the each member of the library of target proteins or peptides is fused with the BD domain of an expression vector contained in yeast a-type of host strain.

The yeast clones of the library of target proteins may be arrayed as a clone library. This may be achieved by depositing each clone containing the BD-Target fusion into a well of a 96- or 384-well plate. Optionally, prior to using this library of BD-Target clones, the BD-Target library may be preselected to filter out any self-activating clones. This selection may be accomplished by allowing the yeast clones that contain the BD-Target fusion to grow in a selection medium used for two-hybrid selection at a later stage, such as the medium SD/-Trp-His. The clones are checked for self-activation of the reporter gene in the absence of the AD domain.

Alternatively, the BD-Target library may be preselected in a selection medium with β- or α-galactosidase substrate. Any positive clones will produce a colored reaction catalyzed the galactosidase expressed from a LacZ reporter gene and can be easily detected by naked eyes or by an instrument. Such clones are self-activating clones that express the reporter gene in the absence of the AD domain. The clones may be excluded from the library of BD-Target clones.

Still referring to FIG. **8**, the BD-target clones of a-strain of yeast may be inoculated into a plate which is pre-seeded with an arrayed library of **V1**/**V2** library of α-strain of yeast haploid cells. The two haploid yeast strains mate in the rich medium and form diploid. The parental clones are screened for expression of the reporter gene which indicates positive interactions between a **V1**/**V2** protein complex and a target protein expressed by the clones in the same well. The scoring of the positive clones may be conveniently carried out by machine-aided automatic screening using β- or α-galactosidase substrate. Aho, S. et al. (1997) "A novel reporter gene MEL1 for the yeast two-hybrid system" Anal. Biochem. 253:270–272.

Compared to the screening of a single target protein against a library of **V1**/**V2** protein complexes, the method illustrated in FIG. **8** is based on clonal mating, i.e., mating between an individual target protein against an individual **V1**/**V2** protein complex. The advantage of such clonal mating is that the efficiency of mating and selection may be enhanced through clonal mating when large numbers of target proteins and **V1**/**V2** protein complexes such as antibodies are involved.

The methods described can be used for large scale screening of libraries of biomolecules, such as fully human antibody repertoires, against a wide variety target molecules or ligands. The screening process may be automated for high throughput screening of the biomolecules. For example, such screening process allows for efficient isolation and collection of antibodies against any EST (human, mouse, or any other organisms), or any known structural/functional protein domains (Zinc finger, helix-loop-helix, etc.), or totally random peptides with various lengths.

In contrast, by using conventional methods for screening antibody in vivo, such as the hybridoma and "XENOMOUSE" technologies, such a large-scale and comprehensive antibody collection may have been impractical due to technical limitations associated with using animal as the host for the libraries of antibodies and target molecules.

By using the method of the present invention, the antibody repertoires can be screened for affinity interaction between an antibody in the library and a target antigen individually in vivo by clonal mating without losing track of individual clones. The screening should be more efficient than the procedure performed on mice, owing the to fast proliferation rate and ease of handling of yeast cells.

The method of the present invention should provide vary useful tools for profiling functions of genes, in particular, functional proteomics, efficiently and economically. With the completion of human genome sequencing, the demands are tremendous for efficient large-scale screening for functional proteins aimed at large numbers of target molecules. The high affinity and functional antibodies, as well as other multimeric proteins, that are selected by using the methods of the present invention should find a wide variety applications in prevention, diagnosis, therapeutic treatment of diseases and in other biomedical or industrial uses.

## 6. Mutagenesis of the Fusion Protein Leads Positively Selected Against Target Protein(s)

As described above, protein leads, such as dsFv, Fab or antibody leads, can be identified through selection of the primary library carrying V1 and V2 against one or more target proteins. The coding sequences of these protein leads may be mutagenized in vitro or in vivo to generated a secondary library more diverse than these leads. The mutagenized leads can be selected against the target protein (s) again in vivo following similar procedures described for the selection of the primary library carrying V1 and V2. Such mutagenesis and selection of primary antibody leads effectively mimics the affinity maturation process naturally occurring in a mammal that produces antibody with progressive increase in the affinity to the immunizing antigen.

The coding sequences of the fusion protein leads may be mutagenized by using a wide variety of methods. Examples of methods of mutagenesis include, but are not limited to site-directed mutagenesis, error-prone PCR mutagenesis, cassette mutagenesis, random PCR mutagenesis, DNA shuffling, and chain shuffling.

Site-directed mutagenesis or point mutagenesis may be used to gradually change the V1 and V2 sequences in specific regions. This is generally accomplished by using oligonucleotide-directed mutagenesis. For example, a short sequence of an antibody lead may be replaced with a synthetically mutagenized oligonucleotide in either the heavy chain or light chain region or both. The method may not be efficient for mutagenizing large numbers of V1 and V2 sequences, but may be used for fine toning of a particular lead to achieve higher affinity toward a specific target protein.

Cassette mutagenesis may also be used to mutagenize the V1 and V2 sequences in specific regions. In a typical cassette mutagenesis, a sequence block, or a region, of a single template is replaced by a completely or partially randomized sequence. However, the maximum information content that can be obtained may be statistically limited by the number of random sequences of the oligonucleotides. Similar to point mutagenesis, this method may also be used for fine toning of a particular lead to achieve higher affinity toward a specific target protein.

Error-prone PCR, or "poison" PCR, may be used to the V1 and V2 sequences by following protocols described in Caldwell and Joyce (1992) PCR Methods and Applications 2:28–33. Leung, D. W. et al. (1989) Technique 1:11–15. Shafikhani, S. et al. (1997) Biotechniques 23:304–306. Stemmer, W. P. et al. (1994) Proc. Natl. Acad. Sci. USA 91:10747–10751.

FIG. **9** illustrates an example of the method of the present invention for affinity maturation of antibody leads selected from the primary antibody library. As illustrated in FIG. **9**, the coding sequences of the antibody leads selected from clones containing the primary library are mutagenized by using a poison PCR method. Since the coding sequences of the antibody library are contained in the expression vectors isolated from the selected clones, one or more pairs of PCR primers may be used to specifically amplify the $V_H$ and $V_L$ region out of the vector. The PCR fragments containing the $V_H$ and $V_L$ sequences are mutagenized by the poison PCR under conditions that favors incorporation of mutations into the product.

Such conditions for poison PCR may include a) high concentrations of $Mn^{2+}$ (e.g. 0.4–0.6 mM) that efficiently induces malfunction of Taq DNA polymerase; and b) disproportionally high concentration of one nucleotide substrate (e.g., dGTP) in the PCR reaction that causes incorrect incorporation of this high concentration substrate into the template and produce mutations. Additionally, other factors such as, the number of PCR cycles, the species of DNA polymerase used, and the length of the template, may affect the rate of mis-incorporation of "wrong" nucleotides into the PCR product. Commercially available kits may be utilized for the mutagenesis of the selected antibody library, such as the "Diversity PCR random mutagenesis kit" (catalog No. K1830-1, Clontech, Palo Alto, Calif.).

The PCR primer pairs used in mutagenesis PCR may preferably include regions matched with the homologous recombination sites in the expression vectors. This design allows re-introduction of the PCR products after mutagenesis back into the yeast host strain again via homologous recombination. This also allows the modified $V_H$ or $V_L$ region to be fused with the AD domain directly in the expression vector in the yeast.

Still referring to FIG. **9**, the mutagenized scFv fragments are inserted into the expression vector containing an AD domain via homologous recombination in haploid cells of α type yeast strain. Similarly to the selection of antibody clones from the primary antibody library, the AD-antibody containing haploid cells are mated with haploid cells of opposite mating type (e.g. a type) that contains the BD-Target vector and the reporter gene construct. The parental diploid cells are selected based on expression of the reporter gene and other selection criteria as described in detail in Section 3.

Other PCR-based mutagenesis method can also be used, alone or in conjunction with the poison PCR described above. For example, the PCR amplified $V_H$ and $V_L$ segments may be digested with DNase to create nicks in the double DNA strand. These nicks can be expanded into gaps by other exonucleases such as Bal 31. The gaps may be then be filled by random sequences by using DNA Klenow polymerase at low concentration of regular substrates dGTP, dATP, dTTP, and dCTP with one substrate (e.g., dGTP) at a disproportionately high concentration. This fill-in reaction should produce high frequency mutations in the filled gap regions. These method of DNase I digestion may be used in conjunction with poison PCR to create highest frequency of mutations in the desired $V_H$ and $V_L$ segments.

The PCR amplified $V_H$ and $V_L$ segments or antibody heavy chain and light chain segments may be mutagenized in vitro by using DNA shuffling techniques described by Stemmer (1994) Nature 370:389–391; and Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747–10751. The $V_H$, $V_L$ or antibody segments from the primary antibody leads are digested with DNase I into random fragments which are then reassembled to their original size by homologous recombination in vitro by using PCR methods. As a result, the diversity of the library of primary antibody leads are increased as the numbers of cycles of molecular evolution increase in vitro.

The $V_H$, $V_L$ or antibody segments amplified from the primary antibody leads may also be mutagenized in vivo by exploiting the inherent ability of mution in pre-B cells. The Ig gene in pre-B cells is specifically susceptible to a high-rate of mutation in the development of pre-B cells. The Ig promoter and enhancer facilitate such high rate mutations in a pre-B cell environment while the pre-B cells proliferate. Accordingly, $V_H$ and $V_L$ gene segments may be cloned into a mammalian expression vector that contains human Ig enhancer and promoter. This construct may be introduced into a pre-B cell line, such as 38B9, which allows the mutation of the $V_H$ and $V_L$ gene segments naturally in the pre-B cells. Liu, X., and Van Ness, B. (1999) Mol. Immunol. 36:461–469. The mutagenized $V_H$ and $V_L$ segments can be amplified from the cultured pre-B cell line and re-introduced back into the AD-containing yeast strain via, for example, homologous recombination.

The secondary antibody library produced by mutagenesis in vitro (e.g. PCR) or in vivo, i.e., by passing through a mammalian pre-B cell line may be cloned into an expression vector and screened against the same target protein as in the first round of screening using the primary antibody library. For example, the expression vectors containing the secondary antibody library may be transformed into haploid cells of $\alpha$ type yeast strain. These $\alpha$ cells are mated with haploid cells a type yeast strain containing the BD-target expression vector and the reporter gene construct. The positive interaction of antibodies from the secondary antibody library is screened by following similar procedures as described for the selection of the primary antibody leads in yeast.

Alternatively, since the secondary antibody library may be relatively low in complexity (e.g.,$10^4$–$10^5$ independent clones) as compared to the primary libraries (e.g.,$10^7$–$10^4$), the screening of the secondary antibody library may be performed without mating between two yeast strains. Instead, the linearized expression vectors containing the AD domain and the mutagenized $V_H$ and $V_L$ segments may be directly co-transformed into yeast cells containing the BD-target expression vector and the reporter gene construct. Via homologous recombination in yeast, the secondary antibody library are expressed by the recombined AD-antibody vector and screened against the target protein expressed by the BD-target vector by following similar procedures as described for the selection of the primary antibody leads in yeast.

### 7. Functional Expression and Purification of Selected Antibody

The library of proten complexes encoded by V1 and V2 that are generated and selected in the screening against the target protein(s) may be expressed in hosts after the V1 and V2 sequences are operably linked to an expression control DNA sequence, including naturally-associated or heterologous promoters, in an expression vector. By operably linking the V1 and V2 sequences to an expression control sequence, the V1 and V2 coding sequences are positioned to ensure the transcription and translation of these inserted sequences. The expression vector may be replicable in the host organism as episomes or as an integral part of the host chromosomal DNA. The expression vector may also contain selection markers such as antibiotic resistance genes (e.g. neomycin and tetracycline resistance genes) to permit detection of those cells transformed with the expression vector.

Preferably, the expression vector may be a eukaryotic vector capable of transforming or transfecting eukaryotic host cells. Once the expression vector has been incorporated into the appropriate host cells, the host cells are maintained under conditions suitable for high level expression of protein complexes encoded by V1 and V2, such as dcFv, Fab and antibody. The polypeptides expressed are collected and purified depending on the expression system used.

The dcFv, Fab, or fully assembled antibodies selected by using the methods of the present invention may be expressed in various scales in any host system. FIG. **11** illustrates examples of host systems: bacteria (e.g. *E. coli*), yeast (e.g. *S. cerevisiae*), and mammalian cells (COS). The bacteria expression vector may preferably contain the bacterial phage T7 promoter and express either the heavy chain and/or light chain region of the selected antibody. The yeast expression vector may contain a constitutive promoter (e.g. ADGI promoter) or an inducible promoter such as (e.g. GCN4 and Gal 1 promoters). All three types of antibody, dcFv, Fab, and full antibody, may be expressed in a yeast expression system.

The expression vector may be a mammalian express vector that can be used to express the protein complexes encoded by V1 and V2 in mammalian cell culture transiently or stably. Examples of mammalian cell lines that may be suitable of secreting immunoglobulins include, but are not limited to, various COS cell lines, HeLa cells, myeloma cell lines, CHO cell lines, transformed B-cells and hybridomas.

Typically, a mammalian expression vector includes certain expression control sequences, such as an origin of replication, a promoter, an enhancer, as well as necessary processing signals, such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcriptional terminator sequences. Examples of promoters include, but are not limited to, insulin promoter, human cytomegalovirus (CMV) promoter and its early promoter, simian virus SV40 promoter, Rous sarcoma virus LTR promoter/enhancer, the chicken cytoplasmic β-actin promoter, promoters derived from immunoglobulin genes, bovine papilloma virus and adenovirus.

One or more enhancer sequence may be included in the expression vector to increase the transcription efficiency. Enhancers are cis-acting sequences of between 10 to 300 bp that increase transcription by a promoter. Enhancers can effectively increase transcription when positioned either 5' or 3' to the transcription unit. They may also be effective if located within an intron or within the coding sequence itself. Examples of enhancers include, but are not limited to, SV40 enhancers, cytomegalovirus enhancers, polyoma enhancers, the mouse immunoglobulin heavy chain enhancer. and adenovirus enhancers. The mammalian expression vector may also typically include a selectable marker gene. Examples of suitable markers include, but are not limited to, the dihydrofolate reductase gene (DHFR), the thymidine kinase gene (TK), or prokaryotic genes conferring antibiotic resistance. The DHFR and TK genes prefer the use of mutant cell lines that lack the ability to grow without the addition of thymidine to the growth medium. Transformed cells can then be identified by their ability to grow on non-supplemented media. Examples of prokaryotic drug resistance genes useful as markers include genes conferring resistance to G418, mycophenolic acid and hygromycin.

The expression vectors containing the **V1** and **V2** sequences can then be transferred into the host cell by methods known in the art, depending on the type of host cells. Examples of transfection techniques include, but are not limited to, calcium phosphate transfection, calcium chloride transfection, lipofection, electroporation, and microinjection.

The **V1** and **V2** sequences may also be inserted into a viral vector such as adenoviral vector that can replicate in its host cell and produce the polypeptide encoded by **V1** and **V2** in large amounts.

In particular, as illustrated in FIG. **11**, the dcFv, Fab, or fully assembled antibody may be expressed in mammalian cells by using a method described by Persic et al. (1997) Gene, 187:9–18. The mammalian expression vector that is described by Persic and contains EF-$\alpha$ promoter and SV40 replication origin is preferably utilized. The SV40 origin allows a high level of transient expression in cells containing large T antigen such as COS cell line. The expression vector may also include secretion signal and different antibiotic markers (e.g. neo and hygro) for integration selection.

Once expressed, polypeptides encoded by **V1** and **V2** may be isolated and purified by using standard procedures of the art, including ammonium sulfate precipitation, fraction column chromatography, and gel electrophoresis. Once purified, partially or to homogeneity as desired, the polypeptides may then be used therapeutically or in developing, performing assay procedures, immunofluorescent stainings, and in other biomedical and industrial applications. In particular, the antibodies generated by the method of the present invention may be used for diagnosis and therapy for the treatment of various diseases such as cancer, autoimmune diseases, or viral infections.

In a preferred embodiment, the human antibodies that are generated and screened by using the methods of the present invention may be expressed directly in yeast. According to this embodiment, the heavy chain and light chain regions from the selected expression vectors may be PCR amplified with primers that simultaneously add appropriate homologous recombination sequences to the PCR products. These PCR segments of heavy chain and light chain may then be introduced into a yeast strain together with a linearized expression vector containing desirable promoters, expression tags and other transcriptional or translational signals.

For example, the PCR segments of heavy chain and light chain regions may be homologously recombined with a yeast expression vector that already contains a desirable promoter in the upstream and stop codons and transcription termination signal in the downstream. The promoter may be a constitutive expression promoter such as ADH1, or an inducible expression promoter, such as Gal 1, or GCN4 (A. Mimran, I. Marbach, and D. Engelberg, (2000) Biotechniques 28:552–560). The latter inducible promoter may be preferred because the induction can be easily achieved by adding 3-AT into the medium.

The yeast expression vector to be used for expression of the antibody may be of any standard strain with nutritional selection markers, such as His 3, Ade 2, Leu 2, Ura 3, Trp 1 and Lys 2. The marker used for the expression of the selected antibody may preferably be different from the AD vector used in the selection of antibody in the two-hybrid system. This may help to avoid potential carryover problem associated with multiple yeast expression vectors.

For expressing the dcFv antibody in a secreted form in yeast, the expression vector may include a secretion signal in the 5' end of the $V_H$ and $V_L$ segments, such as an alpha factor signal and a 5-pho secretion signal. Certain commercially available vectors that contain a desirable secretion signal may also be used (e.g., pYEX-S1, catalog #6200-1, Clontech, Palo Alto, Calif.).

The dcFv antibody fragments generated may be analyzed and characterized for their affinity and specificity by using methods known in the art, such as ELISA, western, and immune staining. Those dcFv antibody fragments with reasonably good affinity (with dissociation constant preferably above $10^{-6}$ M) and specificity can be used as building blocks in Fab expression vectors, or can be further assembled with the constant region for full length antibody expression. These fully assembled human antibodies may also be expressed in yeast in a secreted form.

FIG. **10**A illustrates the secondary structures of the dcFv, Fab and a fully assembled antibody. The $V_H$ sequence encoding the selected dcFv protein may be linked with the constant regions of a full antibody, $C_H\mathbf{1}$, $C_H\mathbf{2}$ and $C_H\mathbf{3}$. Similarly, the $V_L$ sequence may be linked with the constant region $C_L$. The assembly of two units of $V_H$–$C_H\mathbf{1}$–$C_H\mathbf{2}$–$C_H\mathbf{3}$ and $V_L$–$C_L$ leads to formation of a fully functional antibody. The present invention provides a method for producing fully functional antibody in yeast. Fully functional antibody retaining the rest of the constant regions may have a higher affinity (or avidity) than a dcFv or a Fab. The full antibody should also have a higher stability, thus allowing more efficient purification of antibody protein in large scale.

The method is provided by exploiting the ability of yeast cells to uptake and maintain multiple copies of plasmids of the same replication origin. According to the method, different vectors may be used to express the heavy chain and light chain separately, and yet allows for the assembly of a fully functional antibody in yeast. This approach has been successfully used in a two-hybrid system design where the BD and AD vectors are identical in backbone structure except the selection markers are distinct. This approach has been used in a two-hybrid system design for expressing both BD and AD fusion proteins in the yeast. The BD and AD vectors are identical in their backbone structures except the selection markers are distinct. Both vectors can be maintained in yeast in high copy numbers. Chien, C. T., et al. (1991) "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest" Proc. Natl. Acad. Sci. USA 88:9578–9582.

In the present invention, the heavy chain gene and light chain genes are placed in two different vectors. Under a suitable condition, the $V_H$–$C_H\mathbf{1}$–$C_H\mathbf{2}$–$C_H\mathbf{3}$ and $V_L$–$C_L$ sequences are expressed and assembled in yeast, resulting in a fully functional antibody protein with two heavy chains and two light chains. This fully functional antibody may be secreted into the medium and purified directly from the supernatant.

The dcFv with a constant region, Fab, or fully assembled antibody can be purified using methods known in the art. Conventional techniques include, but are not limited to, precipitation with ammonium sulfate and/or caprylic acid, ion exchange chromatography (e.g. DEAE), and gel filtration chromatography. Delves (1997) "Antibody Production: Essential Techniques", New York, John Wiley & Sons, pages 90–113. Affinity-based approaches using affinity matrix based on Protein A, Protein G or Protein L may be more efficiency and results in antibody with high purity. Protein A and protein G are bacterial cell wall proteins that bind specifically and tightly to a domain of the Fc portion of certain immunoglobulins with differential binding affinity to different subclasses of IgG. For example, Protein G has higher affinities for mouse IgG1 and human IgG3 than does Protein A. The affinity of Protein A of IgG1 can be enhanced by a number of different methods, including the use of binding buffers with increased pH or salt concentration. Protein L binds antibodies predominantly through kappa light chain interactions without interfering with the antigen-binding site. Chateau et al. (1993) "On the interaction between Protein L and immunoglobulins of various mammalian species" Scandinavian J. Immunol., 37:399–405. Protein L has been shown to bind strongly to human kappa light chain subclasses I, III and IV and to mouse kappa chain subclasses I. Protein L can be used to purify relevant kappa chain-bearing antibodies of all classes (IgG, IgM, IgA, IgD, and IgE) from a wide variety of species, including human, mouse, rat, and rabbit. Protein L can also be used for the affinity purification of scFv and Fab antibody fragments containing suitable kappa light chains. Protein L-based reagents is commercially available from Actigen, Inc., Cambridgem, England. Actigen can provide a line of recombinant Protein products, including agarose conjugates for affinity purification and immobilized forms of recombinant Protein L and A fusion protein which contains four protein A antibody-binding domains and four protein L kappa-binding domains.

Other affinity matrix may also be used, including those that exploit peptidomimetic ligands, anti-immunoglobulins, mannan binding protein, and the relevant antigen. Peptidomimetic ligands resemble peptides but they do not correspond to natural peptides. Many of Peptidomimetic ligands contain unnatural or chemically modified amino acids. For example, peptidomimetic ligands designed for the affinity purification of antibodies of the IGA and IgE classes are commercially available from Tecnogen, Piana di Monte Verna, Italy. Mannan binding protein (MBP) is a mannose- and N-acetylglucosamine-specific lectin found in mammalian sera. This lectin binds IgM. The MBP-agarose support for the purification IgM is commercially available from Pierce.

Immunomagnetic methods that combine an affinity reagent (e.g. protein A or an anti-immunoglobulin) with the ease of separation conferred by paramagnetic beads may be used for purifying the antibody produced. Magnetic beads coated with Protein or relevant secondary antibody may be commercially available from Dynal, Inc., N.Y.; Bangs Laboratories, Fishers, Ind.; and Cortex Biochem Inc., San Leandro, Calif.

Direct expression and purification of the selected antibody in yeast is advantageous in various aspects. As a eukaryotic organism, yeast is more of an ideal system for expressing human proteins than bacteria or other lower organisms. It is more likely that yeast will make the dcFv, Fab, or fully assembled antibody in a correct conformation (folded correctly), and will add post-translation modifications such as correct disulfide bond(s) and glycosylations.

Yeast has been explored for expressing many human proteins in the past. Many human proteins have been successfully produced from the yeast, such as human serum albumin (Kang, H. A. et al. (2000) Appl. Microbiol. Biotechnol. 53:578–582) and human telomerase protein and RNA complex (Bachand, F., et al. (2000) RNA 6:778–784).

Yeast has fully characterized secretion pathways. The genetics and biochemistry of many if not all genes that regulate the pathways have been identified. Knowledge of these pathways should aid in the design of expression vectors and procedures for isolation and purification of antibody expressed in the yeast.

Moreover, yeast has very few secreted proteases. This should keep the secreted recombinant protein quite stable. In addition, since yeast does not secrete many other and/or toxic proteins, the supernatant should be relatively uncontaminated. Therefore, purification of recombinant protein from yeast supernatant should be simple, efficient and economical.

Additionally, simple and reliable methods have been developed for isolating proteins from yeast cells. Cid, V. J. et al. (1998) "A mutation in the Rho&GAP-encoding gene BEM2 of Saccharomyces cerevisiae affects morphogenesis and cell wall functionality" Microbiol. 144:25–36. Although yeast has a relatively thick cell wall that is not present in either bacterial or mammalian cells, the yeast cells can still keep the yeast strain growing with the yeast cell wall striped from the cells. By growing the yeast strain in yeast cells without the cell wall, secretion and purification of recombinant human antibody may be made more feasible and efficient.

By using yeast as host system for expression, a streamlined process can be established to produce recombinant antibodies in fully assembled and purified form. This may save tremendous time and efforts as compared to using any other systems such as humanization of antibody in vitro and production of fully human antibody in transgenic animals.

In summary, the compositions, kits and methods provided by the present invention should be very useful for selecting proteins such as human antibodies with high affinity and specificity against a wide variety of targets including, but not limited to, soluble proteins (e.g. growth factors, cytokines and chemokines), membrane-bound proteins (e.g. cell surface receptors), and viral antigens. The whole process of library construction, functional screening and expression of highly diverse repertoire of human antibodies can be streamlined, and efficiently and economically performed in yeast in a high throughput and automated manner. The selected proteins can have a wide variety of applications. For example, they can be used in therapeutics and diagnosis of diseases including, but not limited to, autoimmune diseases, cancer, transplant rejection, infectious diseases and inflammation.

## EXAMPLE

### Example 1

### Construction of Expression Vectors Containing Human Antibody Library Using Homologous Recombination in Vivo

The following illustrates examples of how to use general homologous recombination as an efficient way of constructing recombinant human antibody library. The coding sequence of each member of the antibody library includes heavy-chain and light chain regions derived from a library of human antibody repertoire. The light chain region of the antibody is fused with a two-hybrid system activation domain (AD) to form a two-hybrid expression vector in the yeast. In an alternative design, the light chain region of the antibody is fused with Aga2 subunit of yeast a-agglutinin to form a surface dislay expression vector in the yeast. The heavy chain region of the antibody is expressed separately from the light chain region by a different promoter.

1) Isolation of Human Antibody cDNA Gene Pool

A complex human antibody cDNA gene pool is generated by using the method described in Sambrook, J., et al. (1989) Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; and Ausubel, F. M. et al. (1995) Current Protocols in Molecular Biology" John Wiley & Sons, NY.

Briefly, total RNA is isolated from the white cells (mainly B cells) contained in peripheral blood supplied by un-immunized humans. Blood sample at 500 ml, which contains approximately $10^8$ B-lymphocytes, are obtained from healthy donors from Stanford Hospital Blood Center. The white blood cells are separated on Ficoll and RNA is isolated by a modified method. Sambrook, J., et al. (1989), supra; and Zhu, L. et al. (1997) "Yeast Gal 4 activation domain fusion expression libraries" in "The Yeast Two-Hybrid System", S. Fields and P. Bartel, Ed., Oxford University Press, pages 73–98.

If starting from tissue, RNA is first isolated using standard procedures. Ramirez, F. et al. (1975) "Changes in globin messenger rNA content during erythroid cell differentiation" J. Biol. Chem. 250:6054–6058; and Sambrook, J., et al. (1989), supra. First strand cDNA synthesis is performed using the method of Marks et al. in which a set of heavy and light chain cDNA primers are designed to anneal to the constant regions for priming the synthesis of cDNA of heavy chain and light chains (both kappa and lambda) antibody genes in separate tubes. Marks et al. (1991) Eur. J. Immunol. 21:985–991.

Alternatively, human spleen, leukocyte, fetal liver, or bone marrow cDNA can be purchased directly from commercial sources, such as Clontech, Palo Alto, Calif.

2) PCR Amplification of Heavy and Light Chain Genes

The coding sequences of human heavy and light chain genes are amplified from the cDNA library generated above by using a method described by Sblattero and Bradbury (1998) Immunotechnology 3:271–278. This method allows almost 100% coverage of all human $V_H$, Vλ and Vκ genes from the known Ig gene database. Specifically, cDNA pool from human spleen is used (human spleen Marathon-Ready cDNA, Cat.#7412-1, Clontech, Palo Alto, Calif.). Alternatively, cDNA pool from human leukocytes can also be used (human leukocyte Marathon-Ready cDNA, catalog #7406-1, Clontech, Palo Alto, Calif.).

The genes encoding human antibody heavy chain and light chain regions are amplified separately by PCR using sets of mixed 5' and 3' primers for each class of variable region fragment (Fv), fragment antigen binding region (Fab) and full length heavy chain region (Ab). Primers used for PCR amplification of these regions of the heavy chain and light chain are listed in Table 2 and named as follows:

| Heavy Chain Primers for Directional Cloning: | | |
|---|---|---|
| Fv, | 5' primers: | Sequences VH5'1–7 [SEQ ID NO: 14–20] |
| | 3' primers: | Sequences VH3'1–6 [SEQ ID NO: 21–26] |
| Fab, | 5' primers: | Sequences FabH5'1–7 [SEQ ID NO: 14–20] |
| | 3' primers: | Sequences FabH3'1 [SEQ ID NO: 27] |
| Full length, | 5' primers: | Sequences AbH5'1–7 [SEQ ID NO: 14–20] |
| | 3' primers: | Sequences AbH3'1 [SEQ ID NO: 28] |
| Light Chain Primers for Cloning into a Site Downstream of GAL-4 AD: | | |
| λ chain | | |
| Fv, | 5' primers: | Sequences Vλ5'1–9 [SEQ ID NO: 29–37] |
| | 3' primers: | Sequences Vλ3'1–2 [SEQ ID NO: 38–39] |
| Full length, | 5' primers: | Sequences Abλ5'1–9 [SEQ ID NO: 29–37] |
| | 3' primers: | Sequences Abλ3'1–2 [SEQ ID NO: 40–41] |
| κ chain | | |
| Fv, | 5' primers: | Sequences Vκ5'1–4 [SEQ ID NO: 42–45] |
| | 3' primers: | Sequences Vκ3'1–4 [SEQ ID NO: 46–49] |
| **Full length,** | **5' primers:** | **Sequences Abλ5'1–4 [SEQ ID NO: 42–45]** |
| | 3' primers: | Sequences Abλ3'1[SEQ ID NO: 50] |
| Light Chain Primers for Cloning into a Site Upstream of GAL-4 AD: | | |
| λ chain | | |
| Fv, | 5' primers: | Sequences Vλ5'1'–9' [SEQ ID NO: 51–59] |
| | 3' primers: | Sequences Vλ3'1'–2' [SEQ ID NO: 60–61] |
| **Full length,** | **5' primers:** | **Sequences Abλ5'1'–9' [SEQ ID NO: 51–59]** |
| | 3' primers: | Sequences Abλ3'1'–2' [SEQ ID NO: 62–63] |
| κ chain | | |
| Fv, | 5' primers: | Sequences Vκ5'1'–4' [SEQ ID NO: 64–67] |
| | 3' primers: | Sequences Vκ3'1'–4' [SEQ ID NO: 68–71] |
| **Full length,** | **5' primers:** | **Sequences Abλ5'1'–4' [SEQ ID NO: 64–67]** |
| | 3' primers: | Sequences Abλ3'1' [SEQ ID NO: 72] |

Each of the heavy chain 5'-primers, which are the same for Fv, Fab and full length Ab, contains a Not I restriction site. Each of the heavy chain 3'-primer contains both Sac II and Sal I restriction sites. By using these primer sets for heavy chain regions listed in Table 2, the heavy chain library can be generated by PCR amplification of human antibody library to incorporate restriction sites at the 5' and 3' ends. This library can be cleaved by restriction digestion and directionally cloned into a yeast expression vector, such as a modified pACT2 vector, pACT-DC.

Each of the λ and κ light chain 5'-primers which are the same for Fv and full length Ab contains a 60-bp flanking sequence (underlined) that is designed to be homologous to a section at the 5' terminus of a linearized pACT2 or pACT-DC. Each of the the λ and κ light chain 5'-primers contains a 60-bp flanking sequence (underlined) homologous to a section at the 3' terminus of the linearizd pACT2 or pACT-DC. These primer sets are used in combination to amplify the light-chain regions of the human antibody gene pool from the cDNA library. The resulting PCR fragments can be used for subsequent insertion into the pACT2 or pACT-DC vector via homologous recombination. The plasmid map of vector pACT2 is shown in FIG. **12A**.

Each flanking sequence added to the primary PCR product is 60 bp in length. The design of the flanking sequence of primer is such that the reading frame of the light chain sequences are conserved with upstream GAL 4 reading frame that is encoded by the cloning vector. Depending on the cloning vector used in the next step, additional features such as epitope tags (for detection and purification) and unique restriction enzyme recognition sites (for subcloning) can also be integrated at this step by primer design.

The amplified heavy chain library can be directionally cloned into a modified pACT2 vector which is described below in bacteria. Subsequently, the amplified light chain library can be cloned into this vector in yeast via homologous recombination by following the schemed depicted in FIG. **3**.

The PCR reaction is done in the volume of 50 ul containing 5 ul of the cDNA synthesized from step 2, 20 pmol concentration of the mixed 5' and 3' primers, 250 uM dNTPs, 10 mM KCl, 10 mM $(NH4)_2SO_4$, 20 mM Tris.HCl (pH 8.8), 2.0 mM MgCl2, 100 mg/ml BSA, and 1 ul (1 unit) of AdvanTaq® DNA polymerase (Clontech, CA). The reaction mixture is subjected to 30 cycles of amplification using a Perkin-Elmer thermal cycler. The cycle is 94° C. for 1 min (denaturation), 57° C. for 1 min (annealing), and 72° C. for 2.5 min (extension). Vλ and Vκ chain PCR products are pooled together at this stage. The PCR products are checked by electrophoresis and purified from 1.0% agarose gel using Qiax affinity matrix (Qiagen, CA) and resuspended in 25 ul of $H_2O$.

3) Directional Cloning of Heavy Chain Library into a Two-hybrid AD Vector in Bacteria

The PCR fragments of the antibody heavy chain cDNA gene pool generated above are cloned into a modified pACT2 vector by directional cloning in bacteria.

The original pACT2 plasmid (FIG. **12**A) is modified by incorporating an expression cassette derived from the pBridge plasmid (FIG. **12**B). Since pACT2 has two Bgl 2 sites flanking the original multiple cloning site (MCS), the original MCS-II in pBridge that includes Not I and Bgl 2 needs to be modified. Two oligonucleotides (Sequences A1 and A2, SEQ ID NO: 73–74) with phosphate groups at their 5' ends are synthesized and annealed to each other. This annealed double-stranded DNA is ligated into the Bgl 2 site after pBridge plasmid is digested with Bgl 2 and dephosphorylated. Such modification results in a new vector (pBridge-1) that contains 3 new restriction sites (i.e. Sac 2, Pvu 2 and Sal I), but lacks Bgl 2 site.

```
                            Sac2  Pvu2  Sal I
Sequence A1           5'-pGATCCGCGGCAGCTGTCGAC-3'
[SEQ ID NO. 73]

Sequence A2                   3'-GCGCCGTCGACAGCTGCTAGp-5'
[SEQ ID NO. 74]
```

The expression cassette in pBridge-1 contains the MET25 promoter ($P_{MET25}$) followed by a nuclear localization signal (NLS), a HA-tag and a MCS (designated MCS-II), and the PGK terminator ($T_{PGK}$). The following oligos (Sequences A3 and A4, SEQ ID NO: 75–76) are used as primers to amplify the cassette (~1 kb) from pBridge-1 by PCR.

Sequence A3: oligo corresponding to the 5' end of ($P_{MET25}$) [SEQ ID NO: 75]

```
          Xho I
      5'-ACTCGAGCTTCTAATTCTTCCAACATAC
```

Sequence A4: oligo complementing to the 3' end of ($T_{PGK}$) [SEQ ID NO: 76]

```
          Xho I
      5'-ACTCGAGAACGCAGAATTTTCGAGTTATT
```

The cassette is then cloned into a cloning vector pGEM-T, and its sequence is confirmed by standard DNA sequencing methods.

FIG. **12**C depicts the overall process of modifying pACT2 to generate pACT-DC, a yeast expression vector having double expression cassettes. Briefly, the vector pACT2 is digested with Not I enzyme, and treated with Klenow fragment of *E. coli* DNA polymerase I in the presence of dCTP and dGTP. The vector is then self-ligated to produce a plasmid pACT3 that lacks Not I site and lox P site. The plasmid pACT3 is further digested with Spe I and treated with Klenow fragment in the presence of dNTP's. The vector is then self-ligated, resulting in pACT4 that does not contain Spe I.

Another MCS (designated MCS-III) is added to pACT4 upstream of the GAL4-AD in pACT4. This is done by PCR using primers (sequences A5 & A6, SEQ ID NO: 77–78) with restriction sites added in the primers. As depicted in FIG. **12**C, The PCR product is digested with Spe I and self-ligated. There are five restriction sites added (SgrA I, Apa I, Spe I, Sph I and BssH2, in order) between T antigen NLS and Gal4-AD domain. Since ten codons are added, the Gal4-AD is still in-frame. Its sequence is confirmed by standard DNA sequencing methods. The resulting vector is designated pACT5.

```
Sequence A5 [SEQ ID NO: 77]:
       Spe I  Sph I BssH2
5'-ATATGACTAGTGGCATGCGCGCCAATTTTAATCAAAGTGGG

Sequence A6 [SEQ ID NO: 78]:
       Spe I  ApaI SgrA I
5'-ATATGACTAGTGGGCCCACCGGTGGCGGTACCCAATTCGACCTT
```

The expression cassette derived from pBridge is retrieved from pGEM-T by digestion with Xho I. As depicted in FIG. **12**C, the DNA fragment is then ligated into pACT5 that has been digested with Sal I and dephosphorylated. This ligation destries both Xho I and Sal I sites. The resulting plasmid will be confirmed by restriction digestions. This vector contains two different expression cassettes and is designated pACT-DC. Table 3 lists the oligonucleotides used to modify pACT2 to produce pACT-DC.

The library of expression vectors containing human antibody heavy chain library is constructed by directional cloning in bacteria. The heavy chain library amplified from human antibody gene pool (described in Section 2) of this Example) are cloned into MCS-II of pACT-DC at Not I site at the 5' end and Sac II or Sal I at the 3' end, such that expression of the heavy chain library is under the control of the promoter $P_{met25}$.

In order to avoid internal cutting of the heavy chain library by Sac II or Sal I, the PCR amplified heavy chain library is divided into two portions. The first portion is digested with Not I and Sac II, and then ligated into pACT2-DC digested with Not I and Sac II. The second portion is digested with Not I and Sal I, and then ligated into pACT2-DC digested with Not I and Sal I.

The ligated products are transformed into *E. coli* cells. Care is taken not to have high level of empty vector in the product. Plating density of the library is preferred to have no more than $0.2\times10^4$ colonies per 150 mm diameter plate.

Colonies of *E.coli* transformants are collected and used for plasmid preparation directly. Total volume of the *E.coli* colonies scraped from the plates should be sufficient for a plasmid prep at maxi-level. The total library DNA prepared is subjected to quality control analysis by using these tests: 1) determination of percentage of plasmid containing inserts (preferred to be above 95%); 2) verification of Fv, Fab, or full length heavy chain sequences; 3) determination of read through ability of the junction region sequence; and 4) determination of percentage of non-identical insert sequences from 2–3 dozens of clones. The complexity of the heavy chain library is preferred to be about $10^4$–$10^5$.

4) Cloning of Light Chain Library into pACT-DC via Homologous Recombination in Yeast

The library of expression vectors containing both heavy chain and light chain libraries under transcriptional control of different promoters is constructed through homologous recombination in yeast. The light chain library (including both λ and κ light chain) amplified from human antibody gene pool (described in Section 2 of this Example) are cloned into MCS (located downstream of GAL4 AD) or MCS-III (located upstream of GAL-4 AD) of pACT-DC, such that expression of the heavy chain library is under the control of the promoter $P_{ADH1}$.

The library of pACT-DC containing the heavy chain library is linearized with restriction enzymes digestion (e.g. BamH I, Xho I, preferably Sfi I) in the multiple cloning site (MCS). This is done in 20 ul volume containing the following reagents: 10 μg of vector DNA, 1–2 ul of restriction enzyme Sfi I, 2 ul of 10× buffer. Digestion is carried out at 37° C. overnight. The completion of the enzyme digestion is checked by electrophoresis. No further modification or purification of linearized vector is necessary.

Alternatively, the library of light chain fragments can be cloned into MCS-III site of the pACT-DC, such that the light chain expressed is fused with the N-terminus of GAL-4 AD, i.e. upstream of GAL-4 AD.

The linearized vector DNA (10 μg) is mixed with equal amount of the PCR amplified light chain fragments (described in Section 2 of this Example), preferably at about 5–10 molar excess of the insert fragment). The linearized vector DNA and the PCR fragments are co-transformed into competent yeast strain Y187 (α mating type, from Clontech).

Transformation is performed as the following. Yeast competent cells are prepared by LiAc protocol (Gietz et al. (1992) "Improved method for high efficiency transformation of intact yeast cells" Nucleic Acids Res. 20:1425), or obtained from a commercial source (Life Technology Inc., MD). Minimum yeast competency of 106 transformant/ug DNA may be required for library construction. Yeast competent cells derived from 1 liter culture of $OD_{600}$=0.2 are used for each transformation in 50 ml conical bottom tubes. Yeast cells are thawed at 4° C., washed with de-ionized water and resuspended in 8 ml of 1×TE/LiAc (1×TE/LiAc is made up of 40% polyethylene glycol 4000, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5, and 0.1 M lithium acetate). The mixture of DNA containing the linearized vector and PCR amplified inserts with extended ends is added to the tube and vortexed to mix. The tube is incubated at 30° C. for 30 min, with shaking (200 rpm). DMSO (Dimethyl sulfoxide, 700 ul) is added into the tube and mixed gently. The cells in the tube are heat shocked at 42° C. in a water bath for 15 minutes with occasional swirl. After the heat shock, the cells are pelleted by a brief centrifugation at 4° C. and washed one or two time with water. The cells are resuspended in 1.5 ml of 1×TBE buffer.

Yeast cells are plated into plates made up of selection medium. For Y187 strain of yeast, the SD/-Leu medium is used. Harper et al. (1993), supra. The library scale transformation requires approximately 100 large plates of 150 mm in diameter. Y187 transformed with either linearized vector without insert DNA fragment or vise versa is also plated onto the same selection plates as controls. Y187 transformed with unlinearized vector pACT2 is used as transformation efficiency control and is plated with series dilutions. The plates are incubated bottom up at 30° C. for 3 days or more. Colony number is examined and recorded. If the yeast control transformation with unlinearized pACT2 yields at least 1 million transformants, as expected, 10 millions of single chain library recombinant clones are expected to obtain from each such transformation. Any control transformation with either the linearized vector or insert DNA fragment alone is expected to yield only 1/10 or less number of colonies as compared with the combined vector/insert transformation. This single step of transformation is repeated until 100 million or more independent clones are obtained.

The yeast library recombinant colonies generated as described above are scraped from the final culture plates after growing for 5–7 days. The majority of the yeasts are mixed with 50% (volume) of glycerol and stored at −80° C. for future library screening use. A small fraction of the yeast clones are subjected to the following quality analyses:

a. Percentage of recombinant clones: PCR amplification of the light chain insert directly from yeast with a primer pair matched with flanking vector sequences (e.g., Long PCR primer pair for AD vectors supplied by Clontech) should reveal how many clones are recombinant. Since our design of extended homologous regions for recombination between the insert and cloning vector is sufficient long (about 60 bp), a high percentage of recombinant clone (higher than 95%) should be expected. Libraries with minimum of 90% recombinant clones are preferably to be saved for screening use.
b. Insert size: The same PCR amplification of selected clones should reveal the insert size. Although a small fraction of the library may contain double or other forms of multiple inserts, the majority (>95%) should have single insert with expected size.
c. Fingerprinting verification of sequence diversity: PCR amplification product with the correct size is fingerprinted with frequent digesting restriction enzymes, such as Bst NI or any other 34 base cutters. From the agarose gel electrophoresis pattern, one can determine whether clones analyzed are of the same identity or of the distinct or diversified identity. The PCR products can also be sequenced directly. This will reveal the identity of inserts and the fidelity of the cloning procedure, and will prove the independence and diversity of the clones. If 100 clones are sequenced, it should be expected that only small fraction (<5%) of clones will have multiple isolates.

5) Alternative Design: Cloning of Light Chain Library into a Yeast Surface Display Vector (pYD1) via Homologous Recombination in Yeast

A library of yeast surface display vectors for expressing an antibody library can be constructed by following similar protocols as described above for construction of the library of two-hybrid expression vectors. Briefly, a yeast surface display vector, pYD1 (available from Invitrogen, San Diego, Calif.; Boder and Wittrup (1997) Nature Biotech. 15: 553–557), is used as the expression vector for expressing the cDNA library of human antibody described in Section 1 of this Example. The vector map of pYD1 is shown in FIG. **12**D. As shown in FIG. **12**D, the vector pYD1 encodes Aga2 subunit of the yeast cell wall protein, a-agglutinin (or a-agglutinin). Aga2 subunit forms a-agglutinin by interacting with Aga1 subunit of a-agglutinin through disulfide bonding. The protein complex formed between Aga1 and Aga2 subunits binds to the a-agglutinin yeast adhesion receptor on the yeast cell wall, thus being displayed on the surface of yeast cells.

Using a protocol similar to that for modifying pACT-2 vector, pYD1 is modified to include the MET25 expression cassette from pBridge vector. The modified pYD1 is designated pYD1-DC. PCR fragments of the antibody heavy chain cDNA gene pool are cloned into a site downstream of the $P_{MET25}$ promoter of pYD1-DC through directional cloning in bacteria. The light chain library (including λ and κ light chain) amplified from human antibody gene pool are cloned into the MCS site downstream of Aga2 domain through homologous recombination in yeast. The light chain is thus expressed as a fusion protein with Aga2. The library of yeast surface display vectors encoding human antibody library are transformed into *S. cervisiae* cells. The antibody formed between the heavy chain and Aga2-light chain fusion is displayed on the surface of the yeast cells through association of Aga1/Aga2 complex with the a-agglutinin yeast adhesion receptor on the yeast cell wall. This library of human antibodies displayed on yeast cell surface is screened against a fluorescence-labeled target molecule. Those cells displaying antibodies that bind to the target molecule are selected by FACS.

## Example 2

### Screening of Antibody Libraries in Yeast with the Two-hybrid System Against Defined Protein Antigens via Mating Between Two Yeast Strains

This example describes a procedure used to screen the antibody libraries generated in the Example 1. The human antibody libraries are generated in yeast strain with an α mating type. This mating type of yeast can be readily mated with an a type of yeast with simple mating procedure to form diploid yeast cells. Guthrie and Fink (1991) "Guide to yeast genetics and molecular biology" in Methods in Enzymology (Academic Press, San Diego) 194:1–932. The a-yeast contains the target (probe, or bait) plasmid.

The target plasmid contains a fusion formed between the GAL 4 DNA binding domain (BD) and any desired target protein that is to be used as a probe to fish out the antibodies as its affinity ligand. When the two types of yeast cell mate and form diploid cells, the probe plasmid and the library clone plasmid also come together in a same cell. Therefore, if a specific antibody clone recognizes and binds to the probe protein, each of these proteins or protein fragments should bring their fusion partners (GAL 4 AD and GAL 4 BD) to a close proximity in the promoter region of reporter(s). Under such a circumstance, the reporter(s) construct built in the yeast cells (the parental a-and/or α-type of haploid cells) should be activated by the active GAL 4 proteins. Thus the reporter is expressed and a positive signal in the library screen is detected. Certain reporter(s) are of nutritional reporter, which allows the yeast to grow on a specific selection medium plate.

In practice, equal volume of bait-containing yeast strain (a-type, e.g. AH109 strain) and the antibody library-containing yeast stain (α-type, e.g. Y187 strain) are inoculated into selection liquid medium and incubated with rigorous shaking at 30° C. for 20 hours. These cultures are then mixed in a single flask and allowed to grow in rich medium 1×YPD (20 g/l Difco peptone, 10 g/l yeast extract, and 2% glucose) for 12–16 additional hours with slow shaking at 30° C. Under the rich nutritional culture condition, the two haploid yeast strains encounter and mate to form diploid cells. At the end of this mating process, a good fraction—5–10% of the yeast population present in the mating pool will form diploids. Bendixen, C., Gangloff, S., and Rothstein, R. (1994) "A yeast mating-selection scheme for detection of protein-protein interactions" Nucleic Acids Res. 22:1778–1779.

After mating, the yeast cells are washed with $H_2O$ several times and plated into selection plates by using the SD/-Leu-Trp-His-Ade selections. The first two selections are for selection markers (Leu and Trp) expressed from the vectors and are for retaining both BD and AD vectors in the same yeast cells. The selected cells should be diploid cells, since either haploid cell only expresses one of these markers. The latter two markers are expressed by the reporter from the host strains and are for selection of clones that show positive interaction between the members of the antibody library and the target protein.

## Example 3

### Screening of Human Antibody Libraries Against a Library of Antigens in a Yeast Two-hybrid System

For small number of pre-selected probes (i.e. baits or targets), the procedure of individual mating screening as described above is sufficient. However, this procedure can also be modified to suit for screening against large number of probes. The following list describes the potential probes that are in large number and may not suitable for individual mating screening:

a. A collection of human EST clones, or total library of human EST. Such EST collection can be ordered from public resource in a library format with individually clones arrayed in 96-well or 384-well plates. The EST inserts from the original collection (usually in bacterial cloning and sequencing vectors) are PCR amplified with extended homologous sequences at both ends. The EST inserts can be PCR amplified and additional flanking sequences can be added to both ends of the ESTs by PCR for mediating homologous recombination in yeast. Then through the same homologous recombination procedure describe in Section 4) of Example 2, the EST insert can be cloned into the AD vector. A maximum of three homologous recombination events should be sufficient for the read-through fusion of each EST with the GAL4 AD. Hua, S. B. et al. (1998) "Construction of a modular human EST-derived yeast two-hybrid cDNA library for the human genome protein linkage map" Gene 215:143–152.

b. A collection of certain domain structures, such as zinc finger protein domains each having 18–20 amino acids. These domain structures may not be completely random. Synthetic oligonucleotides with characteristic conserved and random/degenerate residues can be made to cover most of the rational domain structures;

c. A completely random peptide library each having 16–20 amino acid residues. Such a library can also be made by random oligonucleotide synthesis. Such library has been constructed in an AD vector. Yang, M. et al. "(1995) "Protein-protein interactions analyzed with the yeast two-hybrid system" Nucleic Acids Res. 23:1152–1157. Such a library of probes can also be built in an BD vector. Each clone of such library represents a short peptide. The human antibody library (built in AD vector) is screened against this library of probes, peptide ligands for each antibody can be selected. Such peptides may have potential applications in rational design and structural improvement of antigens.

The library of probes are cloned into a DB vector and each is fused with GAL4 DB domain. This library are made as an arrayed clone library by depositing every clone obtained with BD-probe fusion into a well in 96 or 384 well plates. This arrayed format facilitates large scale library screening with machine-aided automation.

Prior to using the library of probes to screen against the human antibody library, the library of probes are transformed into yeast a-type of host strain to select out any self-activating clones. This pre-selection is to allow the yeast harboring only the probe plasmids to grow in a selection medium (SD/-Trp-His) and check for activation without the AD mating partner, the so-called self activation.

Alternatively, the pre-selection is conducted in selection medium with α- or β-galactosidase substrate. Any positive clones will produce a colored reaction and can be easily detected by naked eye or by instrument. The clone that send out positive signals indicating activation of the reporter gene(s) are self-activating clones which are excluded from the subsequent use as the targets for the antibody library.

The machine-aided automatic screening is performed by using 96- or 384-well plates. The target clones of a-strain are sequentially inoculated into a plate which is pre-seeded with an arrayed library of the antibody library of α-strain. The two haploid yeast strains mate in the rich medium and form diploid. The wells sending positive signals of reporter gene expression are detected. The screening process is similar to the individual target screening against a library in the mixed culture as described in Example 3. The difference in this case is that clonal mating (a mating between an individual target against an individual antibody) is performed here to enhance the efficiency when large numbers of targets and human antibodies are involved.

## Example 4

### Maturation of Primary Antibody Isolates by Random Mutagenesis in Vitro and Re-screening in Vivo in a Yeast Two-hybrid System

The antibody clones isolated from in Examples 3–4 can be of various degree of affinity. Although high affinity clones may be obtained with a low marginal possibility, the majority of the clones may need further modification to reach affinity compatible with natural antibodies (dissociation constant at $10^{-9}$ M or lower).

In this example, the sequences of primary clones are mutagenized in vitro to incorporate random mutations into the heavy chain and/or light chain regions, thereby creating a secondary library of antibodies with increased complexity. Complexity of the secondary library is expected to be at 104 or higher. So the combined diversity of primary and secondary libraries screened should be at $10^{14}$–$10^{18}$, no less than the natural antibody diversification through selection/maturation in an animal.

For example, coding sequences of the light chain regions of the selected antibodies are amplified from the corresponding antibody clones by PCR. The light chain region resides in the AD vector and is fused with GAL-4 AD domain. A pair of PCR primers are used to specifically amplify the light chain region out of the vector. The pair of primers are designed to match with the regions of the cloning vectors that flank the light chain genes. These regions contain sequences for homologous recombination between the cloning vector and the amplified product.

This primary PCR product is checked by agarose gel electrophoresis for correct size and amount. An aliquot of the primary PCR product is then subjected to a secondary PCR. This secondary PCR is designed to incorporate mutations into the product under these conditions: high concentration of $Mn^{2+}$ and over-proportionaly high concentration of one nucleotide substrate in the PCR reaction in the PCR reaction. $Mn^{2+}$ at a concentration of between 0.4 and 0.6 mM can efficiently cause Taq polymerase to incorporate mutations into the PCR product. This mis-incorporation is caused by the malfunction of Taq DNA polymerase. Single nucleotide (e.g., dGTP) at an extra higher concentration than the other 3 essential nucleotides (dATP, dTTP, and dCTP) causes the incorrect incorporation of this high concentration substrate into the template and produce mutations.

Besides the two conditions listed above, other condition may influence the rate of mis-incorporation of "wrong" nucleotide into the PCR product, including the number of PCR cycles, the species of DNA polymerase used, and the length of the template. In this example, a pre-made kit is used (Diversity PCR Random Mutagenesis Kit, Cat.#K1830-1, Clontech, Palo Alto, Calif.). This kit contains reagents necessary for optimizing the conditions for random mutation by PCR, such as dNTP Mix and additional dGTP solution, Manganese Sulfate, and control PCR template and primer mix.

As suggested by the user manual for this kit, the following condition is used for PCR mutagenesis: 640 uM $MnSO_4$, 200 uM dGTP. Under this condition, an average of 8 mutations is expected to be found in every 1000 bp, a rate that is sufficient for scFv diversification.

This secondary antibody library is reintroduced into yeast through homologous recombination and screened directly in yeast following similar procedures as in the primary screening described in Example 2 and Example 3, respectively. This whole process mimics the naturally occurring affinity maturation process that higher organisms including human are inherited.

TABLE 1

| Sequence of LoxP Sites | |
|---|---|
| LoxP WT | 5'-ATAACTTCGTATAATGTATGCTATACGAAGTTAT-3'<br>[SEQ ID NO: 1] |
| LoxP511 | 5'-ATAACTTCGTATAGTATACATTATACGAAGTTAT-3'<br>[SEQ ID NO: 2] |
| LoxC2 | 5'-ACAACTTCGTATAATGTATGCTATACGAAGTTAT-3'<br>[SEQ ID NO: 3] |
| LoxP1 | 5'-ATAACTTCGTATAATATATGCTATACGAAGTTAT-3'<br>[SEQ ID NO: 4] |
| LoxP2 | 5'-ATAACTTCGTATAGCATACATTATACGAAGTTAT-3'<br>[SEQ ID NO: 5] |
| LoxP3 | 5'-ATAACTTCGTATAATGTATACTATACGAAGTTAT-3'<br>[SEQ ID NO: 6] |
| LoxP4 | 5'-ATAACTTCGTATAATATAAACTATACGAAGTTAT-3'<br>[SEQ ID NO: 7] |
| LoxP5 | 5'-ATAACTTCGTATAATCTAACCTATACGAAGTTAT-3'<br>[SEQ ID NO: 8] |
| LoxP6 | 5'-ATAACTTCGTATAACATAGCCTATACGAAGTTAT-3'<br>[SEQ ID NO: 9] |
| LoxP7 | 5'-ATAACTTCGTATAACATACCCTATACGAAGTTAT-3'<br>[SEQ ID NO: 10] |
| LoxP8 | 5'-ATTACCTCGTATAGCATACATTATACGAAGTTAT-3'<br>[SEQ ID NO: 11] |
| LoxP9 | 5'-ATAACTTCGTATAGCATACATTATATGAAGTTAT-3'<br>[SEQ ID NO: 12] |
| LoxP10 | 5'-ATTACCTCGTATAGCATACATTATATGAAGTTAT-3'<br>[SEQ ID NO: 13] |

### TABLE 2

Sequence of PCR primers for amplifying heavy- and light-chain gene of human antibody.
(B = C/G/T; D = A/G/T; K = G/T; M = A/C; R = A/G; S = C/G; W = A/T; and Y = C/T)

a) Heavy-chain
5'-primers for Fv:
NotI
VH5'1: 5'-ACC AAG GAA AAA CAA GCG GCC GCA CAG GTG CAG CTG CAG GAG TCS G-3' [SEQ ID NO: 14]

VH5'2: 5'-ACC AAG GAA AAA CAA GCG GCC GCA CAG GTA CAG CTG CAG CAG TCA-3' [SEQ ID NO: 15]

VH5'3: 5'-ACC AAG GAA AAA CAA GCG GCC GCA CAG GTG CAG CTA CAG CAG TGG G-3' [SEQ ID NO: 16]

VH5'4: 5'-ACC AAG GAA AAA CAA GCG GCC GCA GAG GTG CAG CTG KTG GAG WCY-3' [SEQ ID NO: 17]

VH5'5: 5'-ACC AAG GAA AAA CAA GCG GCC GCA CAG GTC CAG CTK GTR CAG TCT GG-3' [SEQ ID NO: 18]

VH5'6: 5'-ACC AAG GAA AAA CAA GCG GCC GCA CAG RTC ACC TTG AAG GAG TCT G-3' [SEQ ID NO: 19]

VH5'7: 5'-ACC AAG GAA AAA CAA GCG GCC GCA CAG GTG CAG CTG GTG SAR TCT GG-3' [SEQ ID NO: 20]

3'-primers for Fv:
Sac II Sal I
VH3'1: 5'-ATC CAC CGC GGT CGA CTA TGA GGA GAC RGT GAC CAG GGT G-3' [SEQ ID NO: 21]

VH3'2: 5'-ATC CAC CGC GGT CGA CTA TGA GGA GAC GGT GAC CAG GGT T-3' [SEQ ID NO: 22]

VH3'3: 5'-ATC CAC CGC GGT CGA CTA TGA AGA GAC GGT GAC CAT TGT-3' [SEQ ID NO: 23]

VH3'4: 5'-ATC CAC CGC GGT CGA CTA TGA GGA GAC GGT GAC CGT GGT CC-3' [SEQ ID NO: 24]

VH3'5: 5'-ATC CAC CGC GGT CGA CTA GGT TGG GGC GGA TGC ACT CC-3' [SEQ ID NO: 25]

VH3'6: 5'-ATC CAC CGC GGT CGA CTA SGA TGG GCC CTT GGT GGA RGC-3' [SEQ ID NO: 26]

3'-primer for IgG CH1 region of Fab:
FabH3'1: 5'-ATC CAC CGC GGT CGA CTA ACA TGG TTT GVR CTC AAC TBT CTT GTC CAC-3' [SEQ ID NO: 27]

3'-primer for IgG CH3 region of Ab:
AbH3'1: 5'-ATC CAC CGC GGT CGA CTA TTT ACC CRG AGA CAG GGA GAG GCT-3' [SEQ ID NO: 28]

b) Light-chain Vλ for cloning into a site downstream of GAL-4 AD
5'-primers for Fv:
Vλ5'1: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT CAG TCT GTS BTG ACG CAG CCG CC-3' [SEQ ID NO: 29]

Vλ5'2: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT TCC TAT GWG CTG ACW CAG CCA C-3' [SEQ ID NO: 30]

Vλ5'3: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT TCC TAT GAG CTG AYR CAG CYA CC-3' [SEQ ID NO: 31]

Vλ5'4: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT CAG CCT GTG CTG ACT CAR YC-3' [SEQ ID NO: 32]

Vλ5'5: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT CAG DCT GTG GTG ACY CAG GAG CC-3' [SEQ ID NO: 33]

TABLE 2-continued

| Sequence of PCR primers for amplifying heavy- and light-chain gene of human antibody. (B = C/G/T; D = A/G/T; K = G/T; M = A/C; R = A/G; S = C/G; W = A/T; and Y = C/T) |
|---|

Vλ5'6: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT CAG CCW GKG CTG ACT CAG CCM CC-3' [SEQ ID NO: 34]

Vλ5'7: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT TCC TCT GAG CTG AST CAG GAS CC-3' [SEQ ID NO: 35]

Vλ5'8: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT CAG TCT GYY CTG AYT CAG CCT-3' [SEQ ID NO: 36]

Vλ5'9: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT AAT TTT ATG CTG ACT CAG CCC C-3' [SEQ ID NO: 37]

3'-primers for Fv:
Vλ3'1: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TAG GAC GGT SAS CTT GGT CC-3' [SEQ ID NO: 38]

Vλ3'2: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC GAG GAC GGT CAG CTG GGT GC-3' [SEQ ID NO: 39]

3'-primer for Cλ1 region
Abλ3'1: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TTA TGA ACA TTC TGC AGG GGC MAC TGT-3' [SEQ ID NO: 40]

3'-primer for Cλ2 region
Abλ3'2: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TTA AGA GCA TTC TGC AGG GGC CAC TGT-3' [SEQ ID NO: 41]

c) Light-chain Vκ for cloning into a site downstream of GAL-4 AD
5'-primers for Fv:
Vκ5'1: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT GAC ATC CRG DTG ACC CAG TCT CC-3' [SEQ ID NO: 42]

Vκ5'2: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT GAA ATT GTR WTG ACR CAG TCT CC-3' [SEQ ID NO: 43]

Vκ5'3: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT GAT ATT GTG MTG ACB CAG WCT CC-3' [SEQ ID NO: 44]

Vκ5'4: 5'-CCA CCA AAC CCA AAA AAA GAG ATC TGT ATG GCT TAC CCA TAC GAT GTT CCA GAT TAC GCT GAA ACG ACA CTC ACG CAG TCT C-3' [SEQ ID NO: 45]

3'-primers for Fv:
Vκ3'1: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TTT GAT TTC CAC CTT GGT CC-3' [SEQ ID NO: 46]

Vκ3'2: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TTT GAT CTC CAS CTT GGT CC-3' [SEQ ID NO: 47]

Vκ3'3: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TTT GAT ATC CAC TTT GGT CC-3' [SEQ ID NO: 48]

Vκ3'4: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TTT AAT CTC CAG TCG TGT CC-3' [SEQ ID NO: 49]

3'primer for Cκ:
Abκ3'1: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC

TABLE 2-continued

Sequence of PCR primers for amplifying heavy- and light-chain gene of human antibody. (B = C/G/T; D = A/G/T; K = G/T; M = A/C; R = A/G; S = C/G; W = A/T; and Y = C/T)

GGG GTT TTT CAG TAT CTA CGA TTC CTA GCA CTC TCC CCT GTT GAA GCT-3' [SEQ ID NO: 50]

d) Light-chain Vλ for cloning into a site upstream of GAL-4 AD
5'-primers for Fv:
Vλ5'1: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC CAG TCT GTS BTG ACG CAG CCG CC-3' [SEQ ID NO: 51]

Vλ5'2: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC TCC TAT GWG CTG ACW CAG CCA C-3' [SEQ ID NO: 52]

Vλ5'3: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC TCC TAT GAG CTG AYR CAG CYA CC-3' [SEQ ID NO: 53]

Vλ5'4: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC CAG CCT GTG CTG ACT CAR YC-3' [SEQ ID NO: 54]

Vλ5'5: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC CAG DCT GTG GTG ACY CAG GAG CC-3' [SEQ ID NO: 55]

Vλ5'6: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC CAG CCW GKG CTG ACT CAG CCM CC-3' [SEQ ID NO: 56]

Vλ5'7: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC TCC TCT GAG CTG AST CAG GAS CC-3'[SEQ ID NO: 57]

Vλ5'8: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC CAG TCT GYY CTG AYT CAG CCT-3' [SEQ ID NO: 58]

Vλ5'9: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC AAT TTT ATG CTG ACT CAG CCC C-3' [SEQ ID NO: 59]

3'-primers for Fv:
Vλ3'1: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC TAG GAC GGT SAS CTT GGT CC-3' [SEQ ID NO: 60]

Vλ3'2: 5'-GAG ATG GTG CAC GAT GCA CAG TTG AAG TGA ACT TGC GGG GTT TTT CAG TAT CTA CGA TTC GAG GAC GGT CAG CTG GGT GC-3' [SEQ ID NO: 61]

3'-primer for Cλ1 region
Abλ3'1: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TGG ATT AAA ATT GGC TGA ACA TTC TGC AGG GGC MAC TGT-3' [SEQ ID NO: 62]

3'-primer for Cλ2 region
Abλ3'2: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TGG ATT AAA ATT GGC AGA GCA TTC TGC AGG GGC CAC TGT-3' [SEQ ID NO: 63]

c) Light-chain Vκ for cloning into a site upstream of GAL-4 AD
5'-primers for Fv:
Vκ5'1: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC GAC ATC CRG DTG ACC CAG TCT CC-3' [SEQ ID NO: 64]

Vκ5'2: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC GAA ATT GTR WTG ACR CAG TCT CC-3' [SEQ ID NO: 65]

TABLE 2-continued

Sequence of PCR primers for amplifying heavy- and light-chain gene of human antibody.
(B = C/G/T; D = A/G/T; K = G/T; M = A/C; R = A/G; S = C/G; W = A/T; and Y = C/T)

Vκ5'3: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC GAT ATT GTG MTG ACB CAG WCT CC-3' [SEQ ID NO: 66]

Vκ5'4: 5'-GAT AAA GCG GAA TTA ATT CCC GAG CCT CCA AAA AAG AAG AGA AAG GTC GAA TTG GGT ACC GCC GAA ACG ACA CTC ACG CAG TCT C-3' [SEQ ID NO: 67]

3'-primers for Fv:
Vκ3'1: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TTG ATT AAA ATT GGC TTT GAT TTC CAC CTT GGT CC-3' [SEQ ID NO: 68]

Vκ3'2: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TTG ATT AAA ATT GGC TTT GAT CTC CAS CTT GGT CC-3' [SEQ ID NO: 69]

Vκ3'3: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TTG ATT AAA ATT GGC TTT GAT ATC CAC TTT GGT CC-3' [SEQ ID NO: 70]

Vκ3'4: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TTG ATT AAA ATT GGC TTT AAT CTC CAG TCG TGT CC-3' [SEQ ID NO: 71]

3'primer for Cκ:
Abκ3'1: 5'-GTT AGT GAA AGT GAA GGA CAA TGA GCT ATC AGC AAT ATT CCC ACT TTG ATT AAA ATT GGC GCA CTC TCC CCT GTT GAA GCT-3' [SEQ ID NO: 72]

TABLE 3

Sequence of oligonucleotides for modifying the cloning vector pACT2.

a) Oligos for modifying the original MCS in pBridge:
Sac2 Pvu2 Sal I
Sequence A1 [SEQ ID NO. 73] 5'-pGATCCGCGGCAGCTGTCGAC-3'

Sequence A2 [SEQ ID NO. 74] 5'-pGTACGTCGACAGCTGCCGCG-3'

b) Oligos for amplifying the $P_{met25}$ expression cassette in pBridge:
Sequence A3: oligo corresponding to the 5' end of ($P_{MET25}$) [SEQ ID NO: 75]
Xho I
5'-ACTCGAGCTTCTAATTCTTCCAACATAC

Sequence A4: oligo complementing to the 3' end of ($T_{PGK}$) [SEQ ID NO: 76]
Xho I
5'-ACTCGAGAACGCAGAATTTTCGAGTTATT

C) Oligos for adding restriction sites to pACT upstream of GAL-4 AD to produce MCS-III
Sequence A5 [SEQ ID NO: 77]
Spe I Sph I BssH2
5'-ATATGACTAGTGGCATGCGCGCCAATTTTAATCAAAGTGGG

Sequence A6 [SEQ ID NO: 78]
Spe I ApaI SgrA I
5'-ATATGACTAGTGGGCCCACCGGTGGCGGTACCCAATTCGACCTT

## SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 80

<210> SEQ ID NO 1
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP WT

<400> SEQUENCE: 1

ataacttcgt ataatgtatg ctatacgaag ttat 34

<210> SEQ ID NO 2
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP511

<400> SEQUENCE: 2

ataacttcgt atagtataca ttatacgaag ttat 34

<210> SEQ ID NO 3
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxC2

<400> SEQUENCE: 3

acaacttcgt ataatgtatg ctatacgaag ttat 34

<210> SEQ ID NO 4
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP1

<400> SEQUENCE: 4

ataacttcgt ataatatatg ctatacgaag ttat 34

<210> SEQ ID NO 5
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP2

<400> SEQUENCE: 5

ataacttcgt atagcataca ttatacgaag ttat 34

<210> SEQ ID NO 6
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP3

<400> SEQUENCE: 6

ataacttcgt ataatgtata ctatacgaag ttat 34

<210> SEQ ID NO 7
<211> LENGTH: 33

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP4

<400> SEQUENCE: 7

ataacttcgt ataatataaa ctatacgaag tta                                 33


<210> SEQ ID NO 8
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP5

<400> SEQUENCE: 8

ataacttcgt ataatctaac ctatacgaag ttat                                34


<210> SEQ ID NO 9
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP6

<400> SEQUENCE: 9

ataacttcgt ataacatagc ctatacgaag ttat                                34


<210> SEQ ID NO 10
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP7

<400> SEQUENCE: 10

ataacttcgt ataacatacc ctatacgaag ttat                                34


<210> SEQ ID NO 11
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP8

<400> SEQUENCE: 11

attacctcgt atagcataca ttatacgaag ttat                                34


<210> SEQ ID NO 12
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP9

<400> SEQUENCE: 12

ataacttcgt atagcataca ttatatgaag ttat                                34


<210> SEQ ID NO 13
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LoxP10

<400> SEQUENCE: 13
```

```
attacctcgt atagcataca ttatatgaag ttat                              34


<210> SEQ ID NO 14
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 14

accaaggaaa aacaagcggc cgcacaggtg cagctgcagg agtcsg                  46


<210> SEQ ID NO 15
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 15

accaaggaaa aacaagcggc cgcacaggta cagctgcagc agtca                   45


<210> SEQ ID NO 16
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 16

accaaggaaa aacaagcggc cgcacaggtg cagctacagc agtggg                  46


<210> SEQ ID NO 17
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 17

accaaggaaa aacaagcggc cgcagaggtg cagctgktgg agwcy                   45


<210> SEQ ID NO 18
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 18

accaaggaaa aacaagcggc cgcacaggtc cagctkgtrc agtctgg                 47


<210> SEQ ID NO 19
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 19

accaaggaaa aacaagcggc cgcacagrtc accttgaagg agtctg                  46


<210> SEQ ID NO 20
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
```

```
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 20

accaaggaaa aacaagcggc cgcacaggtg cagctggtgs artctgg                   47


<210> SEQ ID NO 21
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 21

atccaccgcg gtcgactatg aggagacrgt gaccagggtg                           40


<210> SEQ ID NO 22
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 22

atccaccgcg gtcgactatg aggagacggt gaccagggtt                           40


<210> SEQ ID NO 23
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 23

atccaccgcg gtcgactatg aagagacggt gaccattgt                            39


<210> SEQ ID NO 24
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 24

atccaccgcg gtcgactatg aggagacggt gaccgtggtc c                         41


<210> SEQ ID NO 25
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 25

atccaccgcg gtcgactagg ttggggcgga tgcactcc                             38


<210> SEQ ID NO 26
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 26

atccaccgcg gtcgactasg atgggccctt ggtggargc                            39
```

<210> SEQ ID NO 27
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 27

atccaccgcg gtcgactaac atggtttgvr ctcaactbtc ttgtccac 48

<210> SEQ ID NO 28
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 28

atccaccgcg gtcgactatt tacccrgaga cagggagagg ct 42

<210> SEQ ID NO 29
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 29

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

cagtctgtsb tgacgcagcc gcc 83

<210> SEQ ID NO 30
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 30

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

tcctatgwgc tgacwcagcc ac 82

<210> SEQ ID NO 31
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 31

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

tcctatgagc tgayrcagcy acc 83

<210> SEQ ID NO 32
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 32

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

cagcctgtgc tgactcaryc 80

-continued

<210> SEQ ID NO 33
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 33

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

cagdctgtgg tgacycagga gcc 83

<210> SEQ ID NO 34
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 34

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

cagccwgkgc tgactcagcc mcc 83

<210> SEQ ID NO 35
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 35

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

tcctctgagc tgastcagga scc 83

<210> SEQ ID NO 36
<211> LENGTH: 81
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 36

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

cagtctgyyc tgaytcagcc t 81

<210> SEQ ID NO 37
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 37

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

aattttatgc tgactcagcc cc 82

<210> SEQ ID NO 38
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 38

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc 60

taggacggts ascttggtcc 80

<210> SEQ ID NO 39
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 39

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc 60

gaggacggtc agctgggtgc 80

<210> SEQ ID NO 40
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 40

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc 60

ttatgaacat tctgcagggg cmactgt 87

<210> SEQ ID NO 41
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 41

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc 60

ttaagagcat tctgcagggg ccactgt 87

<210> SEQ ID NO 42
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 42

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

gacatccrgd tgacccagtc tcc 83

<210> SEQ ID NO 43
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 43

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct 60

gaaattgtrw tgacrcagtc tcc 83

<210> SEQ ID NO 44
<211> LENGTH: 83

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 44

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct     60

gatattgtgm tgacbcagwc tcc                                             83


<210> SEQ ID NO 45
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 45

ccaccaaacc caaaaaaaga gatctgtatg gcttacccat acgatgttcc agattacgct     60

gaaacgacac tcacgcagtc tc                                              82


<210> SEQ ID NO 46
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 46

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60

tttgatttcc accttggtcc                                                 80


<210> SEQ ID NO 47
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 47

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60

tttgatctcc ascttggtcc                                                 80


<210> SEQ ID NO 48
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 48

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60

tttgatatcc actttggtcc                                                 80


<210> SEQ ID NO 49
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 49

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60
```

-continued

```
tttaatctcc agtcgtgtcc                                                 80


<210> SEQ ID NO 50
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 50

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60

ctagcactct cccctgttga agct                                            84


<210> SEQ ID NO 51
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 51

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcccagtctg tsbtgacgca gccgcc                                          86


<210> SEQ ID NO 52
<211> LENGTH: 85
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 52

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcctcctatg wgctgacwca gccac                                           85


<210> SEQ ID NO 53
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 53

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcctcctatg agctgayrca gcyacc                                          86


<210> SEQ ID NO 54
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 54

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcccagcctg tgctgactca ryc                                             83


<210> SEQ ID NO 55
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer
```

-continued

```
<400> SEQUENCE: 55

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcccagdctg tggtgacyca ggagcc                                          86


<210> SEQ ID NO 56
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 56

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcccagccwg kgctgactca gccmcc                                          86


<210> SEQ ID NO 57
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 57

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcctcctctg agctgastca ggascc                                          86


<210> SEQ ID NO 58
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 58

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gcccagtctg yyctgaytca gcct                                            84


<210> SEQ ID NO 59
<211> LENGTH: 85
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 59

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gccaatttta tgctgactca gcccc                                           85


<210> SEQ ID NO 60
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 60

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60

taggacggts ascttggtcc                                                 80


<210> SEQ ID NO 61
```

-continued

```
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 61

gagatggtgc acgatgcaca gttgaagtga acttgcgggg tttttcagta tctacgattc     60

gaggacggtc agctgggtgc                                                 80


<210> SEQ ID NO 62
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 62

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc     60

tgaacattct gcaggggcma ctgt                                            84


<210> SEQ ID NO 63
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 63

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc     60

agagcattct gcaggggcca ctgt                                            84


<210> SEQ ID NO 64
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 64

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gccgacatcc rgdtgaccca gtctcc                                          86


<210> SEQ ID NO 65
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 65

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gccgaaattg trwtgacrca gtctcc                                          86


<210> SEQ ID NO 66
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 66

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60
```

-continued

```
gccgatattg tgmtgacbca gwctcc                                            86


<210> SEQ ID NO 67
<211> LENGTH: 85
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 67

gataaagcgg aattaattcc cgagcctcca aaaaagaaga gaaaggtcga attgggtacc     60

gccgaaacga cactcacgca gtctc                                             85


<210> SEQ ID NO 68
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 68

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc     60

tttgatttcc accttggtcc                                                   80


<210> SEQ ID NO 69
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 69

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc     60

tttgatctcc ascttggtcc                                                   80


<210> SEQ ID NO 70
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 70

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc     60

tttgatatcc actttggtcc                                                   80


<210> SEQ ID NO 71
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 71

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc     60

tttaatctcc agtcgtgtcc                                                   80


<210> SEQ ID NO 72
<211> LENGTH: 81
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
```

<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 72

gttagtgaaa gtgaaggaca atgagctatc agcaatattc ccactttgat taaaattggc 60

gcactctccc ctgttgaagc t 81

<210> SEQ ID NO 73
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligo for mutation

<400> SEQUENCE: 73

gatccgcggc agctgtcgac 20

<210> SEQ ID NO 74
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligo for mutation

<400> SEQUENCE: 74

gtacgtcgac agctgccgcg 20

<210> SEQ ID NO 75
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 75

actcgagctt ctaattcttc caacatac 28

<210> SEQ ID NO 76
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 76

actcgagaac gcagaatttt cgagttatt 29

<210> SEQ ID NO 77
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 77

atatgactag tggcatgcgc gccaatttta atcaaagtgg g 41

<210> SEQ ID NO 78
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 78

atatgactag tgggcccacc ggtggcggta cccaattcga cctt 44

-continued

```
<210> SEQ ID NO 79
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: semi-rigid linker

<400> SEQUENCE: 79

Pro Gln Pro Gln Pro Lys Pro Gln Pro Gln Pro Gln Pro Gln Pro Lys
1               5                   10                  15

Pro Gln Pro Lys Pro Glu Pro Glu
            20


<210> SEQ ID NO 80
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 80

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1               5                   10                  15
```

What is claimed is:

**1**. A method for generating a library of yeast expression vectors encoding a library of protein complexes, comprising:

- transforming into yeast cells a library of insert nudeotide sequences that are linear and double-stranded, and a library of linearized yeast expression vectors, each having a 5'- and 3'-terminus sequence at the site of linearization; and
- having homologous recombination occur between the vector and the insert sequence such that the insert sequence is included in the vector in the transformed yeast cells,
- wherein
    - each of the linearized yeast expression vectors in the vector library comprises a first polynucleotide sequence encoding a first polypeptide subunit which varies within the vector library and has a diversity of at least 1×10$^3$;
    - the insert sequences of the insert library comprise a second nucleotide sequence encoding a second polypeptide subunit which varies within the insert library, each of the insert sequences comprising a 5'- and 3'-flanking sequence at the respective ends of the insert sequence and being sufficiently homologous to the 5'- and 3'-terminus sequences of the linearized yeast expression vector, respectively, to enable homologous recombination to occur, and
    - the first and second polypeptide subunits are capable of being expressed as separate proteins and assembling to form the library of protein complexes of the first and second polypeptide subunits and having a diversity of at least 1×10$^6$.

**2**. The method of claim **1**, wherein the 5'- or 3'-flanking sequence of the insert nucleotide sequence is between about 20–120 bp in length.

**3**. The method of claim **1**, wherein the 5'- or 3'-flanking sequence of the insert nucleotide sequence is between about 40–90 bp in length.

**4**. The method of claim **1**, wherein the 5'- or 3'-flanking sequence of the insert nucleotide sequence is between about 45–55 bp in length.

**5**. The method of claim **1**, wherein the yeast expression vector is a 2$\mu$ plasmid vector.

**6**. The method of claim **1**, wherein the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody heavy chain region, and the second nucleotide sequence comprises a coding sequence of an antibody light chain region.

**7**. The method of claim **1**, wherein the library of insert nucleotide sequences are inserted into a site of the vector such that expression of the first and second polypeptide subunits is under the transcriptional control of separate promoters.

**8**. The method of claim **1**, wherein the library of insert nucleotide sequences are inserted into a site of the vector such that the first and second polypeptide subunits are expressed bicistronically from the same promoter.

**9**. The method of claim **1**, wherein the first and second polypeptide subunits when expressed are expressed as separate proteins that self-assemble in cells into which the library of expression vectors are introduced.

**10**. The method of claim **1**, wherein the first polypeptide subunit comprises an antibody heavy-chain region, and the second polypeptide subunit comprises an antibody light-chain region.

**11**. The method of claim **10**, wherein the first polypeptide subunit further comprises an antibody heavy chain constant 1 region, and the second polypeptide subunit further comprises an antibody light chain constant region.

**12**. The method of claim **10**, wherein the source of the coding sequences of the antibody light-chain and heavy-chain regions is from human, non-human primate, or rodent DNA.

**13**. The method of claim **10**, wherein the source of the coding sequences of the antibody light-chain and heavy-chain variable regions is from one or more non-immunized animals.

**14**. The method of claim **10**, wherein the source of the coding sequences of the antibody light-chain and heavy-chain variable regions are selected from the group consisting of human fetal spleen, fetal liver, bone marrow, lymph nodes and peripheral blood cells.

**15**. The method of claim **1**, wherein the diversity of the first and the second polypeptide subunits each independently is at least $1\times10^3$.

**16**. The method of claim **1**, wherein the diversity of the first and the second polypeptide subunits each independently is at least $1\times10^4$.

**17**. The method of claim **1**, wherein the diversity of the first and the second polypeptide subunits each independently is at least $1\times10^5$.

**18**. The method of claim **1**, wherein the diversity of the protein complexes is at least $1\times10^7$.

**19**. The method of claim **1**, wherein the diversity of the protein complexes is at least $1\times10^{10}$.

**20**. The method of claim **1**, wherein the first polypeptide subunit further comprises an activation domain or a DNA binding domain of a transcription activator.

**21**. The method of claim **20**, wherein the transcription activator is selected from the group consisting of GAL4, GCN4, and ADR1 transcription activator.

* * * * *

| | | | |
|---|---|---|---|
| 专利名称(译) | 在酵母中组装和筛选高度复杂和完全人类的抗体库 | | |
| 公开(公告)号 | US6610472 | 公开(公告)日 | 2003-08-26 |
| 申请号 | US09/703399 | 申请日 | 2000-10-31 |
| [标]申请(专利权)人(译) | 泰世基因公司 | | |
| 申请(专利权)人(译) | GENETASTIX CORPORATION | | |
| 当前申请(专利权)人(译) | ADIMAB , LLC | | |
| [标]发明人 | ZHU LI<br>HUA SHAOBING BENJAMIN<br>SHERIDAN JAMES<br>LIN YU HUEI | | |
| 发明人 | ZHU, LI<br>HUA, SHAOBING BENJAMIN<br>SHERIDAN, JAMES<br>LIN, YU-HUEI | | |
| IPC分类号 | C07K16/00 C12Q1/68 G01N33/68 C12N1/19 C40B30/04 G01N33/53 G01N33/569 | | |
| CPC分类号 | C07K16/00 C12Q1/6897 C40B30/04 G01N33/6845 C07K2317/21 C07K2317/55 C12Q2565/201 | | |
| 审查员(译) | 王 , ANDREW | | |
| 外部链接 | Espacenet USPTO | | |

摘要(译)

提供了组合物，方法和试剂盒，用于有效地产生和筛选高度多样化的蛋白质复合物的文库，因为它们能够结合其他蛋白质或寡核苷酸序列。在本发明的一个方面，提供了表达载体文库，用于表达蛋白质复合物文库。该文库包含编码第一多肽亚基的第一核苷酸序列;和编码第二多肽亚基的第二核苷酸序列。第一和第二核苷酸序列各自独立地在表达载体文库中变化。此外，第一和第二多肽亚基表达为单独的蛋白质，其自组装形成蛋白质复合物，例如双链抗体片段（dcFv或Fab）和完全组装的抗体，在其中的文库中。表达载体被引入。通过同源重组可以在酵母细胞中有效地产生表达载体文库;可以通过体内或体外的高通量筛选来选择对其靶分子具有高结合亲和力的编码蛋白质复合物。

FIGURE 1A