Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 861 (2008) 160–170
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Generation of bioactive peptides by biological libraries☆

Christa Mersich, Alois Jungbauer*

Department of Biotechnology, University of Natural Resources and Applied Life Sciences Vienna, Muthgasse 18, Vienna, Austria

Received 19 March 2007; accepted 23 June 2007
Available online 3 July 2007

Abstract

Biological libraries are powerful tools for discovery of new ligands as well as for identification of cellular interaction partners. Since the first development of the first biological libraries in form of phage displays, numerous biological libraries have been developed. For the development of new ligands, the usage of synthetic oligonucletides is the method of choice. Generation of random oligonucleotides has been refined and various strategies for random oligonucleotide design were developed. We trace the progress and design of new strategies for the generation of random oligonucleotides, and include a look at arising diversity biases. On the other hand, genomic libraries are widely employed for investigation of cellular protein–protein interactions and targeted search of proteomic binding partners. Expression of random peptides and proteins in a linear form or integrated in a scaffold can be facilitated both *in vitro* and *in vivo*. A typical *in vitro* system, ribosome display, provides the largest available library size. *In vivo* methods comprise smaller libraries, the size of which depends on their transformation efficiency. Libraries in different hosts such as phage, bacteria, yeast, insect cells, mammalian cells exhibit higher biosynthetic capabilities. The latest library systems are compared and their strengths and limitations are reviewed.

Keywords: Biological library; Surface display library; Peptides; Affinity ligands

Contents

☆ This paper was presented at the Biochromatography and Nanotechnologies Conference, Vellore, Tamil Nadu, India, 12–15 February 2007.

* Corresponding author. Fax: +43 13697615.
E-mail address: alois.jungbauer@boku.ac.at (A. Jungbauer).

1570-0232/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2007.06.031

1. General aspects of biological libraries

1.1. Definition of biological library

A biological library consists of a pool of microorganisms expressing different polypeptides. Each microorganism carries only one encoding DNA or RNA sequence for a certain peptide, representing one clone. Each single clone of the library can be propagated and it will express the same peptide.

A polypeptide library construction starts with the design of the encoding DNA sequence. The source for this insert can be a pool of chemically synthesized degenerated oligonucleotides, cDNA, genomic DNA fragments, or mutagenized specific gene fragments. The library will be constituted by viral particles or by cells.

The next step is the screening of the library against the target molecule. Clones, identified as binders to the target substance, will be sequenced, and their coding regions will be translated into the particular peptide sequences.

1.2. Design of random oligonucleotides

One method of designing a random peptide library is the use of random oligonucleotides. In a fully degenerated oligonucleotide, each triplet will code for one of the 64 possible codons. At each coupling reaction, an equal mixture of all four nucleotides (N) will be used for all three positions in the triplet. In this way, the oligonucleotide will contain all 64 possible codons, and all 20 amino acids and three stop codons will be represented.

If it is necessary to avoid certain stop codons or amino acids, some positions in the triplet cannot be fully randomized. For one position of the triplet, a mixture of only two or three nucleotides will be used instead of a mixture of all four (see Table 1).

Table 1
Design of random oligonucleotides for generation of peptide libraries

Triplet	Function	Reference
NNK	All 20 amino acids possible Only 1 stop codon possible	[83,84]
NNS	All 20 amino acids possible Only 1 stop codon possible	[83,84]
NNY + RNN	No stop codon possible, but Cys and Gln missing	[37]
RNN + NNG + NHY	Cys missing	[85]

N = A, C, G, T; K = G, T; S = G, C; Y = C, T; R = A, G, H = A, C, T.

Another way to design randomized oligonucleotides was presented by LaBean and Kauffman [1]. This method minimizes stop codons and matches amino acid frequencies observed in 207 natural proteins. With the use of a refining-grid search algorithm, termination codons are minimized and amino acid compositions of the peptides get balanced. Three mixtures of nucleotides are designed, each corresponding to one of the three positions in the codon.

A different approach for the synthesis of randomized DNA was described by Neuner et al. [2]. The strategy is based on the use of dinucleotide phosphoramite building blocks in a resin-splitting procedure. Seven dinucleotide building blocks are required to encode all the 20 natural amino acids (see Fig. 1).

There is also a way to constrain peptides by introducing two codons for cysteine in both sides of the random region. The screening of pools of such cyclic libraries (CX_5C, CX_6C, CX_7C) resulted in the isolation of ligands to several integrins [3]. Cyclic and linear peptide libraries were also employed to screen for streptavidin binders. The analysis of the binding peptides showed, that the conformationally constrained cyclic peptides bound streptavidin three orders of magnitude higher than linear peptides [4]. The usage of split inteins [5,6] also allows the production of cyclic peptides.

1.3. Considerations for cloning steps

The introduction of DNA fragments into an appropriate vector and the transformation into microorganisms require optimized protocols to maximize the cloning efficiency, especially for the construction of large libraries.

Because the cloning of a DNA fragment requires compatible ends with the vector, the two DNAs must be cut with the same restriction enzymes. The vector DNA must be linearized and purified. A ligation reaction is set up, where degenerated DNA fragments are mixed at a molar excess with the vector, and are ligated together with the enzyme T4 DNA ligase.

The amount of double stranded DNA fragments required depends on the number of randomized nucleotides and on the expectation how many times a unique sequence should be represented in the library (library complexity). The other important parameter is the transformation efficiency (number of transformants obtained from 1 μg vector DNA) of the system used. Usually, a transformation efficiency in *E. coli* obtained with electroporation is around 10^9 transformants per microgram of supercoiled vector DNA, while the efficiency of a cut-and-religated vector, as in the case in building random sequence libraries, is about 10–100 times less. The ligation mix is used

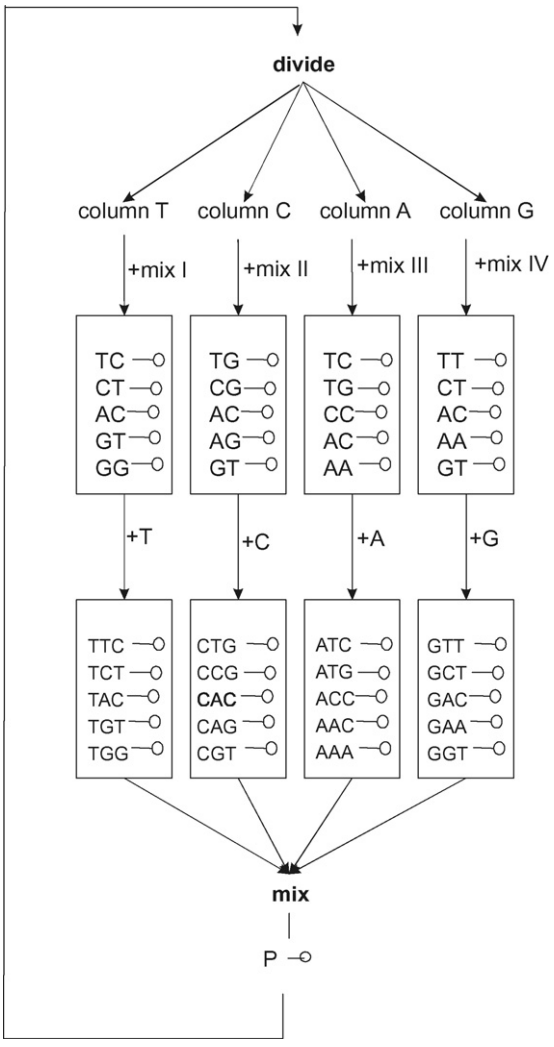

Fig. 1. Strategy for the synthesis of a randomized DNA sequence containing the 20 most abundant codons in highly expressed genes of *E. coli* (from [2]).

to transform competent *E. coli* cells in several separated transformations. An aliquot of these transformed cells is grown on a solid medium and counted in order to calculate library complexity. The library itself can be grown in liquid medium. The plasmids can be harvested and purified to transform the final host organism of the library.

Other methods of inserting double stranded oligonucleotides into the vector, that have been described as alternatives, include: methods based on gap repair [7]; and a process consisting of the ligation of a single stranded oligonucleotide to one end of the vector, the subsequent synthesis of the second strand via the Klenow fragment of DNA polymerase, and finally, a second ligation to the other end of the vector [8].

1.4. Directed evolution by secondary libraries

To improve the affinity of peptide ligands that have been isolated, a secondary library can be constructed by introducing either targeted or random mutations into the binder's coding sequence. This allows some fine-tuning of ligands: by using only those ligands that already display affinity to the target, and by directing the mutations, one should be able to generate ligands with progressively higher affinity to the target. In cassette mutagenesis, the target regions are substituted by a synthetic DNA duplex with the desired mutations [9]. In regional mutagenesis, mutations are introduced by chemical or enzymatic treatments at a controlled rate of alterations per nucleotide. This mutated DNA is then cloned [10]. The combinatorial approach to mutagenesis replaces a certain number of amino acids per peptide using the cassette method [11]. Lastly, spiked oligonucleotides can be synthesized by adding a predetermined amount of a mixture of different bases at specific preset locations. In this manner, the wild type bases are "spiked" [12].

1.5. Comparison of biological libraries over chemically synthesized peptide libraries

In general, biological libraries do not have specific advantages over chemical libraries and vice versa. The choice of a library should be guided by practical matters. Of course, technical considerations, such as the experience of operators and available equipment may well limit the choice. Beyond that, however, there are a set of specific factors that will make one of these systems the correct choice for a given application.

In chemical libraries the diversity is given by 20^n (20 is the number of different amino acids, n is the number of randomized positions). For example, a complete library, constituted of five amino acids, will have 3.2×10^6 different molecules. The longer the peptide sequences are, the synthesis will be more error-prone. In chemical libraries, there is no bias toward specific amino acids, whereas in biological libraries some amino acids are more represented than others, because of the codon degeneracy. Furthermore, the incorporation of non proteinogenic amino acids is only possible in chemically synthesized peptides.

In a fully degenerated oligonucleotide library the diversity is given by $(4 \times 4 \times 4)^n$, where 4 is the number of different nucleotides and n is the number of randomized codons. The only limits on the size of biological libraries are the microorganism's transformation efficiency and the amount of cells that can be handled. In *E. coli*, for example, the upper limit of the transformation efficiency is described as 10^9 transformants per 1 μg of vector DNA. Biological libraries can also consist of long random polypeptides. If the randomized amino acid positions total more than seven, the library will be incomplete (e. g. seven randomized amino acids result in 1.3×10^9 peptides). The representation of amino acids in a biological library encoded by a degenerated oligonucleotide is restricted by several biological exigencies. The codon degeneracy, or the encoding of some amino acids by multiple triplets, ensures that some amino acids will not follow an even, random distribution. Some peptides may prove toxic to the cell, others may be expressed by the cell less efficiently. The advantage of long random sequences, expressed in incomplete libraries, is that the binding region is limited to a few amino acid residues in most cases. Since a long variable peptide will contain within its sequence several short peptide sections, the total number of different short peptides will be higher than the number of different clones representing the library. Furthermore, long random sequences allow affinity selection of peptide ligands that

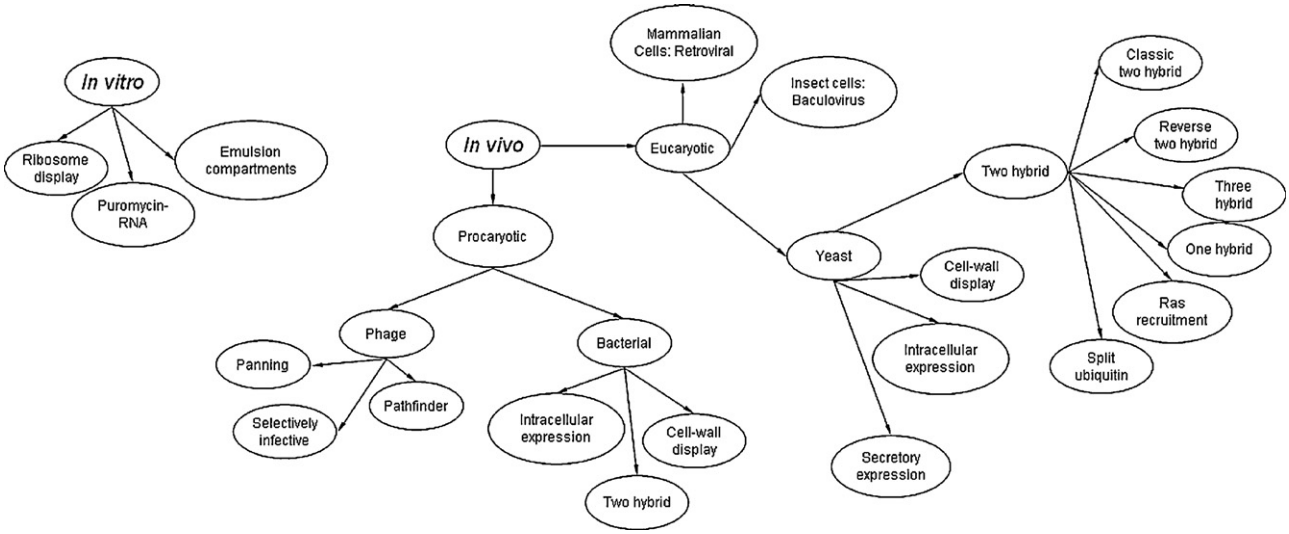

Fig. 2. Overview of available biological polypeptide-display library systems.

require the interaction of few residues spaced apart, or small structural elements.

1.6. Overview of biological libraries

The choice of a particular platform depends on the relative importance of library size, biosynthetic capability, and quantitative precision to the particular application at hand. A summary of available library systems is depicted in Fig. 2.

2. *In vitro* polypeptide-display libraries

In vitro systems represent a special case among biological libraries: they are not propagated in living cells. The two major advantages of *in vitro* transcription/translation systems are the potential to generate very large libraries up to 10^{15} by obviating a cell transformation step and the ability to control screening conditions independent of the maintenance of cell viability.

Ribosome display was first described by Mattheakis et al. [13] for short peptides. It involves the preservation of a polypeptide–ribosome–mRNA ternary complex as a genetic unit. Ribosome display has been adapted in order to screen larger proteins, e.g. single chain antibodies [14]. Optimization of the *in vitro* transcription/translation reactions and protein folding conditions can deliver a system that allows the identification of proteins with improved expression, stability, and affinity [15] (see Fig. 3). In the first step, a DNA library is amplified by PCR, whereas a T7 promoter, ribosome-binding site, and stem-loops are introduced and then transcribed to DNA. After purification, mRNA is translated *in vitro* in an *E. coli* S-30 system

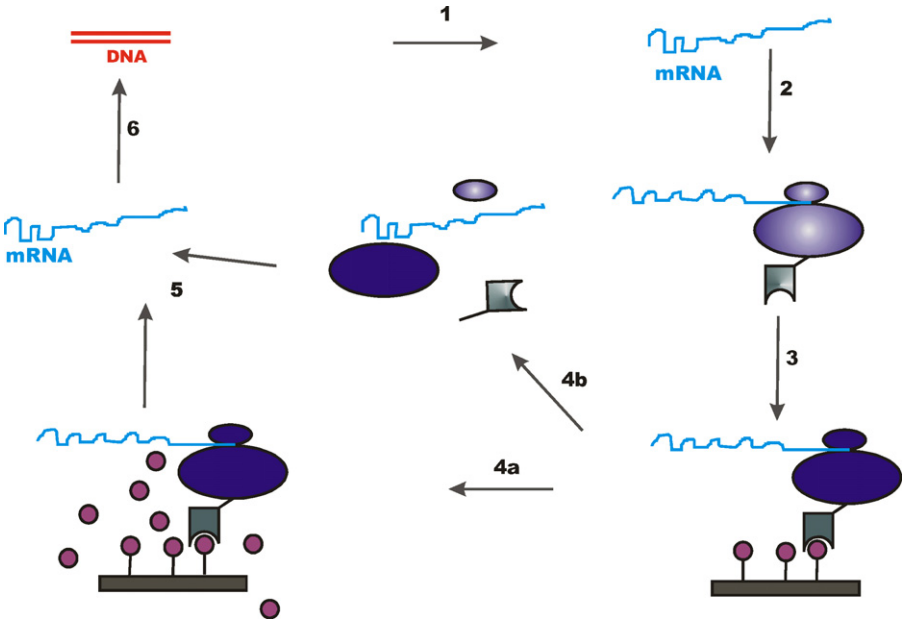

Fig. 3. Principle of *in vitro* ribosome display for screening native protein libraries for ligand binding from Hanes and Pluckthun [15].

in the presence of different factors that enhance the stability of ribosomal complexes and improve the folding of the protein on the ribosomes. The translation process is terminated by cooling on ice, and the ribosome complexes are stabilized by increasing the magnesium concentration. In the next step, the desired ribosome complexes are affinity selected from the translation mixture by the binding of the native protein to the immobilized ligand. Unspecifically bound ribosome complexes are removed by intensive washing. After washing, the bound ribosome complexes can be dissociated by EDTA; alternatively, the entire complex can be specifically eluted with its ligand. The RNA is later isolated from the complexes. This isolated mRNA is then reverse transcribed into cDNA, and the cDNA is amplified by PCR. The amplified cDNA is used for the next cycle of enrichment; additionally, a portion of the newly minted cDNA can be analyzed by cloning and sequencing, and/or by ELISA.

Puromycin-linked peptide–RNA systems were first described by Roberts and Szostak [16]. It consists of a nucleotide covalently linked to a polypeptide. Covalent RNA–peptide complexes are formed by linkage with puromycin in an *in vitro* transcription/translation reaction. One advantage of these covalent complexes is that they may be subjected to harsh biochemical treatments and screening conditions that would inactivate polysome complexes, viral particles, or cells.

To localize the phenotypic effects of a mutated enzyme, Tawfik and Griffiths [17] dispersed an *in vitro* transcription/translation reaction in an oil–water emulsion creating aqueous compartments with cellular dimensions. This technique is called emulsion compartments. In this method, the *in vitro* transcription/translation reaction mixture contains a library of genes linked to a substrate that will be enzymatically converted into a marker for the desired reaction. Once the mixture is dispersed, it forms a water-in-oil emulsion consisting of compartments that each contains one gene and the apparatus for transcription and translation. In the next step, the genes are both transcribed and translated within their compartments. At this point, proteins with enzymatic activities convert the substrate into a marker product linked to the gene. After breaking the emulsion, the liberated gene–product complexes are selectively enriched and amplified.

3. *In vivo* polypeptide-display libraries

In vivo library display platforms can be created in hosts ranging from simple phage particles to whole cells, and prokaryotic microorganisms and on through eukaryotic cells which can be induced to display complex proteins on their surface.

3.1. Prokaryotic display

3.1.1. Phage display

Phage display delivers proteins that are displayed as fusions to a phage coat protein (see Fig. 4). Phage particles, propagated in *E. coli*, are isolated by "panning" against a ligand bound to a solid-phase support. This method was first described by Smith [18].

The filamentous phage's minor coat protein pIII is the most widely used display protein. Present at 3–5 copies per virion, it is synthesized with an N-terminal signal peptide that is cleaved during the translocation through the inner membrane. The mature pIII protein is 406 amino acids long; the C-terminus is in the cytosol, the N-terminus and the peptide are in the periplasm, and a single membrane-spanning domain anchors the protein to the membrane. Library sequences can be inserted into the N-terminal domain, which tolerates insertion of long fragments: Sparks et al. [19] demonstrated the expression of 36-mer peptides in this domain.

Further, the major capsid protein pVIII of the filamentous phage can also be used for peptide display. This protein is syn-

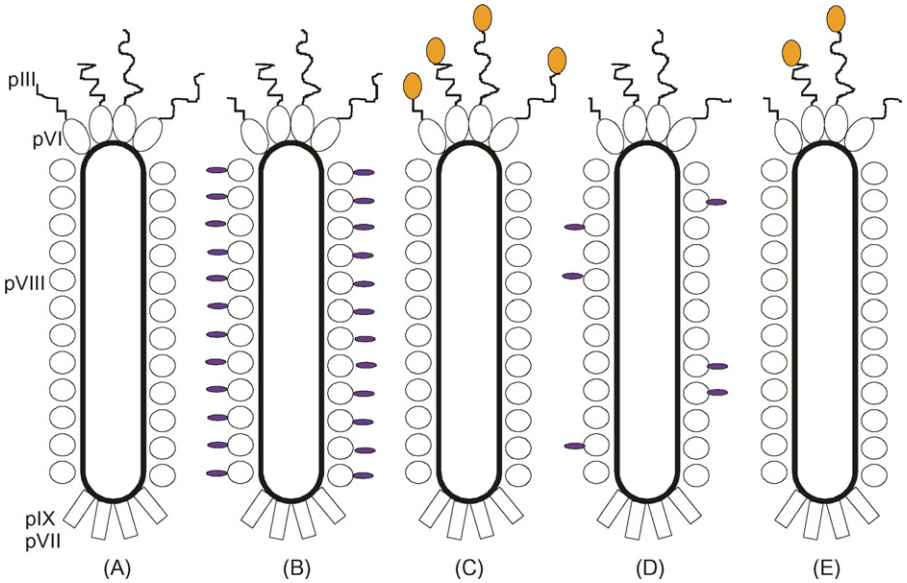

Fig. 4. Schematic representation of filamentous phage and phagemid display. (A) Wild type phage. (B) Phage display of peptides fused to pVIII. (C) Phage display of peptides fused to pIII. (D) Phagemid display of peptides fused to pVIII. (E) Phagemid display of peptides fused to pIII.

thesized as a precursor with an N-terminal leader sequence of 23 amino acids—this sequence is necessary for the insertion of the pVIII into the host bacterium's cytoplasmatic membrane. During phage assembly, 2700 copies aggregate around the virus' DNA, forming a helical array with the N-terminus exposed to the medium. The mature protein is 50 amino acids long; up to 6 amino acids can be inserted into pVIII without disrupting the protein coat assembly [20,21].

Less often used as scaffolds for peptide display are the filamentous phage's minor coat protein pVI [22] and the D protein of bacteriophage λ [23].

Two systems are used to create libraries in filamentous phages: the polyvalent display ("one-gene system") and the monovalent display ("two-gene system"). In the polyvalent display system, the DNA fragments coding for the peptides are inserted into the phage vector, usually between a particular coat protein and its single peptide. Each coat protein molecule will be fused to the peptide; consequently, the number of displayed peptides will correspond to the number of coat proteins.

In the monovalent display system, the phage genome is modified by the deletion of the viral genes with the retention of the sequences needed for packing the phage into virions, including the modified coat protein genes. This defective phage is called a phagemid. When a cell harboring a phagemid is infected by a filamentous helper phage (which supplies the missing genes), virions are produced. These virions display a mixture of recombinant coat proteins, encoded by the phagemid's gene, and the corresponding wild-type proteins, encoded by the helper phage's gene. If the ultimate goal is the display of larger peptides or if pVIII fusions are desired, then the two-gene system is preferred.

3.1.2. Selectively infective phage technology

Phage library screening can be converted into a true genetic growth selection by linking the binding event of the library to the infectivity of the phage particle. A "selectively infective phage" consists of two components: a filamentous phage particle made non-infective by replacing the N-terminal domains of its gene3 protein (g3p) with a ligand-binding protein, and an "adapter" molecule which contains the ligand linked to the phage's missing g3p N-terminal domains. Infectivity is only restored when the displayed protein binds the ligand, and thereby attaching the missing N-terminal domains of g3p to the phage particle [24]. Phage propagation becomes strictly dependent on this protein–ligand interaction.

3.1.3. Pathfinder selection phage display

This method was developed to identify antibodies that bind in the vicinity of an initial target by catalyzing biotinylation of antibody–phage particles, which bind near a horseradish peroxidase-conjugated lead-binding molecule [25,26].

3.1.4. Display on bacterial surfaces and cytoplasmatic expression

Several fusion protein-based strategies leading to the display of relatively short peptides on the surface of Gram-negative bacteria have been described. When fused into surface exposed loops of outer membrane proteins (Omps) from enteric bacteria, peptides of less than 60 amino acid residues can be displayed on the cell surface. Displaying peptides in bacteria requires several elements: including a leader sequence, one or more hydrophobic membrane spanning regions, and a signal peptide that directs the protein to the outer membrane. Bacterial membrane proteins that have been used for the display of peptides on the cell surface include: the *E. coli* OmpA [27] (see Fig. 5A), LamB [28] and PhoE [29], *E. coli* lipoproteins [30] and lipoprotein-OmpA fusions Lpp′OmpA [31] (see Fig. 5B), the *Pseudomonas* OprF [32], the *Shigella* VirG$_\beta$ [33] and the *Neisseria* IgA$_\beta$ [34].

Extracellular appendages, such as pili and flagella, have also served as successful sites for the display of peptides. Lu et al. [35] developed the FLITRX system, in which an *E. coli* display vector based on the major structural component of the *E. coli* flagellum FliC is utilized. The small protein thioredoxin (TrxA) was introduced into a dispensable region of FliC. In the next step, peptides were inserted into the TrxA domain of the FliC fusion. Subsequent analysis revealed that the TrxA domain containing the peptide was surface exposed.

Expression systems for the display of proteins in Gram-positive bacteria have also been developed. Hansson et al. [36] used fusions to the cell-wall bound X-domain of protein A, permitting the display of peptides up to 88 amino acids long to the surface of *Staphylococcus* strains.

Expression of random peptide libraries as free proteins [37] in the cytoplasm of bacteria has been demonstrated via fusions with a DNA binding protein [38], and by fusions to ubiquitin [39]. Bussow et al. [40,41] used an *E. coli* expression vector for the construction of a human cDNA library. For the expression of proteins in a range of 15–100 kDa, they used a high-throughput system in which growth, induction, cell lysis, and screening were all performed on filter membranes.

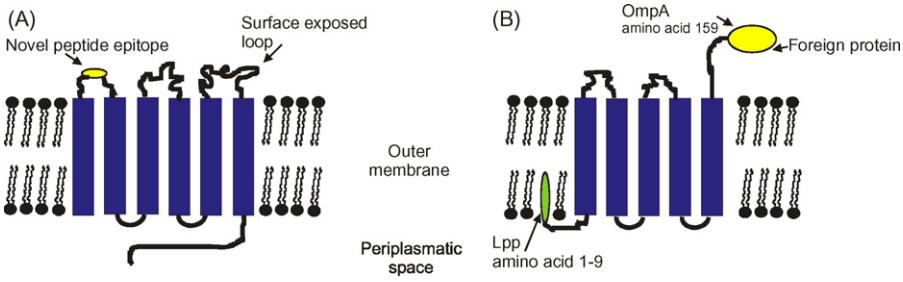

Fig. 5. Schematic presentation of *E. coli* OmpA and the chimeric Lpp′OmpA. Rectangles represent membrane-spanning β-strands of OmpA. (A) A novel peptide is expressed in a surface exposed loop. (B) A protein is fused to the C-terminus of the Lpp′OmpA construct.

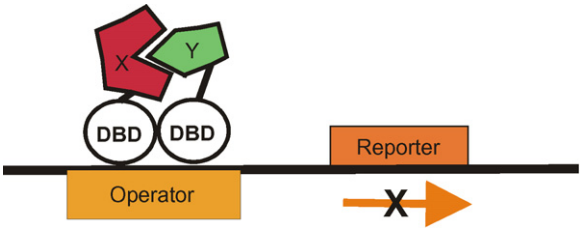

Fig. 6. Bacterial two-hybrid systems based on chimeric transcriptional repressors. λcI or LexA protein consist of an N-terminal DNA-binding domain (DBD) and a C-terminal dimerization domain. They only act as transcriptional repressors in a dimeric form. When native dimerization domains of λcI or LexA are replaced by a pair of interacting proteins X and Y, the resulting chimeric complex will bind to the operator and repress the transcription of the reporter gene.

3.1.5. Bacterial two-hybrid systems

Originally described by Fields and Song [42] for yeast, two-hybrid systems are a powerful approach to screen for *in vivo* protein–protein interactions. They employ transcriptional activity as a measure of protein–protein interaction. Bacterial two-hybrid based on chimeric transcription repressors were described by Hu et al. [43] using the λ repressor and Kornacker et al. [44] using the LexA repressor (see Fig. 6). A two-hybrid system based on chimeric RNA polymerase was described by Dove et al. [45]. Further variations of these systems include the reconstitution of cAMP signaling cascade [46] and functional complementation of mouse DHFR [47]. Their main advantages over yeast two-hybrid methods comprise the faster growth of *E. coli*, the higher transformation efficiency and the lack of nuclear localization of the hybrid proteins as in yeast systems.

4. Eukaryotic systems

A general advantage of eukaryotic systems is the capacity for high fidelity folding of mammalian extracellular proteins and domains.

4.1. Yeast two-hybrid system

The two-hybrid system [42,48] is a genetic method that uses transcriptional activity as a measure of protein–protein interaction. It relies on the modular nature of many site-specific transcriptional activators that consist of a DNA-binding domain and a transcriptional activation domain. The DNA-binding domain targets the activator to the specific genes that are to be expressed, and the activation domain contacts other proteins of the transcriptional machinery in order to enable transcription. In the two-hybrid system, the two domains of the activator are not covalently linked. They can however, be brought together by the interaction of any two proteins. To test if two proteins X and Y interact, both are expressed as fusion proteins to a transcriptional activation domain ("prey"), and to a DNA-binding domain ("bait"). The two vector constructs, each of which harbors either prey or bait, are transformed to yeast strains of opposite mating types. The yeast strains are mated, combining the two fusion proteins into one diploid cell, which contains a reporter gene. If the proteins X and Y interact, they bring the activation domain close to the DNA-binding domain, consequently creating a functional activator. Successful interactions are characterized by the expression and subsequent detection of the reporter gene and its product (see Fig. 7).

The yeast two-hybrid method has been undergoing continual refinement and extension since its invention, resulting in a variety of variants, including reverse two-hybrid, three-hybrid, and one-hybrid systems (reviewed in [49,50]).

Examination of secreted or cell-surface proteins in generally not possible in this system: because the yeast two-hybrid method requires nuclear localization and transcriptional activation.

4.1.1. The ras recruitment system

This variant of the two-hybrid system uses a temperature sensitive yeast mutant in the *ras* pathway, whose growth can be rescued by the Sos protein. Thus, a fusion protein with Sos in combination with a myristolated protein can artificially recruit Sos to the membrane, where it rescues the genetic defect [51]. In contrast to the original two-hybrid system, this system does not require nuclear localization and could have applications with membrane protein interactions.

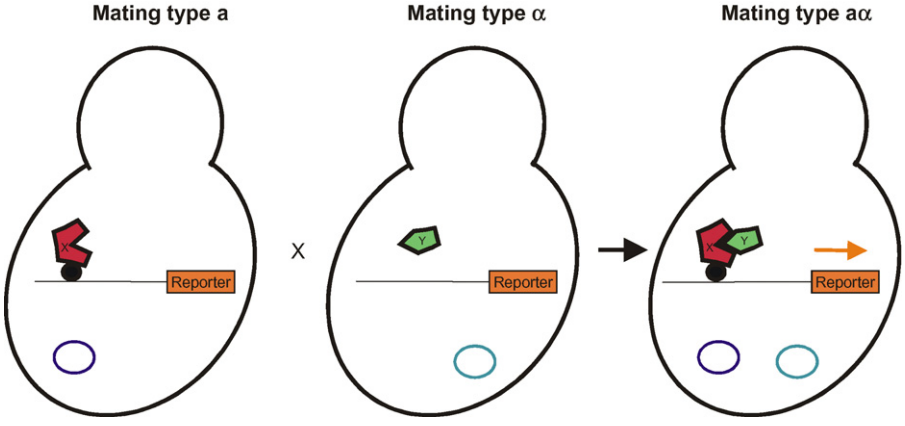

Fig. 7. Classical yeast two-hybrid system. A protein of interest X is expressed in yeast as a fusion to a DNA-binding domain. Another protein of interest Y is fused to a transcriptional activation domain. The two yeast strains are mated to combine the two fusion proteins in one cell. If X and Y interact in the resulting diploid cells, they reconstitute a transcription factor which activates a reporter gene.

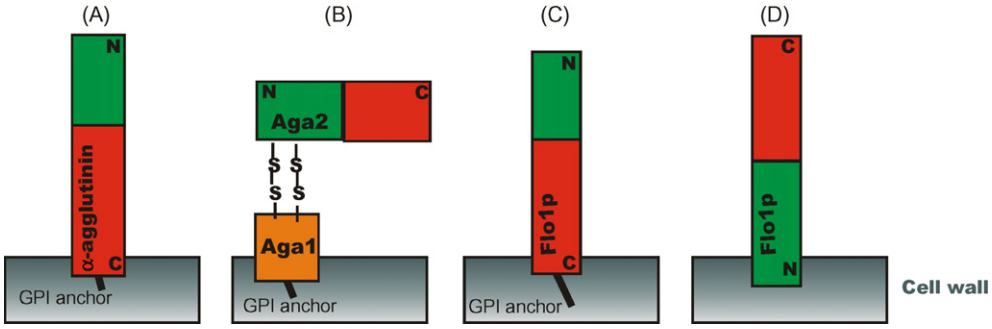

Fig. 8. Yeast cell surface display systems after Ueda and Tanaka [82]. (A) α-Agglutinin system; (B) a-agglutinin system; (C) C-terminal region of Flo1p; (D) N-terminal region of Flo1p.

4.1.2. Split ubiquitin system

Johnsson and coworkers [52,53] developed a cytoplasmatic two-hybrid assay based on ubiquitin. When the C-terminal fragment of ubiquitin is fused to a reporter gene and co-expressed with the amino terminal fragment, the two halves will reconstitute the native ubiquitin. The reconstituted ubiquitin will cleave the reporter protein. For its adaptation to detect protein-protein interactions, a mutant N-terminal fragment – which is not able to interact with the C-terminus on its own – was fused to one protein and a carboxy terminal fragment reporter hybrid was fused to its prospective interaction partner. This interaction results in the reconstitution of ubiquitin which leads to cleavage and release of the reporter gene.

4.2. Yeast surface display

Scaffolds for peptide display in yeast have been successfully created by using some of the glycosylphosphatidylinositol (GPI) anchored proteins on the yeast cell surface. Agα1p and Aga1p are mating-type specific agglutinins which mediate the direct cell–cell adhesion between cells during mating and are assumed to be located on the outermost surface [54]. Fusions to the C-terminal half of α-agglutinin are used to anchor heterologous proteins on the yeast surface, since these proteins are covalently linked with glucan (see Fig. 8A). In the case of α-agglutinin, the secretion-type Aga2p, the binding subunit is linked by disulfide bridges to the core protein Aga1p [55,56]. The heterologous proteins are fused to the C-terminus of the 69 amino acid binding subunit Aga2p. The Aga2 fusion protein and Aga1p associate within the secretory pathway, they are exported to the cell surface, and covalently linked to the cell wall (Fig. 8B).

The flocculin Flo1p, a lectine-like well wall protein of *S. cerevisiae*, plays a major role in flocculation [57]. Due to high levels of *N*- and *O*-glycosylation, it is believed to form stem-like structures. A repetitive region of 1200 amino acids gives Flo1p the length it needs to transverse the cell wall, providing for the design of variable-length anchors. Flo1p consists of several functional domains; as a secretion signal, flocculation functional domain, GPI-anchor attachment signal, and membrane-anchoring domain. Reversible flocculation of cells occurs when the Flo1p flocculation functional domain recognizes and adheres non-covalently to cell wall components, such as α-mannan carbohydrates [58]. Two types of cell surface display using Flo1p have been developed (see Fig. 8C and D) [59,60].

4.3. Yeast intracellular expression

Several high-throughput applications that make use of the intracellular expression of cDNA libraries in yeast have been reported. Lueking et al. [61] described a dual vector system for the expression of a human fetal brain cDNA library in *P. pastoris* and *E. coli*. Yeast expressed 29 soluble proteins, while *E. coli* produced only 9 proteins under native conditions. Holz et al. [62] successfully expressed a cDNA library in *S. cerevisiae*.

4.4. Yeast secretory expression

Our group recently developed a library system in *S. cerevisiae*, in which the random peptide sequence is fused to the carboxy terminus of the 17 kDa eukaryotic initiation factor 5a (eIF5a) and secreted to the culture supernatant [63,64]. In this system gene expression is auto-induced by the alcohol-dehydrogenase promoter. The yeast mating pheromone alpha-leader sequence upstream of the gene fusion site facilitates secretion of the recombinant protein into the culture supernatant. The N-terminal octapeptide FLAG-tag DYKDDDDK enables rapid detection of the recombinant protein by monoclonal antibodies [65]. A real time biosensor was developed to establish a quantitative assay for FLAG fusion proteins using FLAG tagged bacterial alkaline phosphatase as standard. A range of FLAG tagged bacterial alkaline phosphatase concentrations were injected over the anti-FLAG M2 antibody surface of the biosensor and used as standards to determine the concentration of different FLAG-tagged proteins with a molecular mass of 18.1 kDa, respectively, 49.3 kDa from yeast culture supernatants. The M2 immobilized chip was found to retain binding capacity following regeneration for at least 120 cycles. This real time biosensor method allows the quantitation of proteins from culture supernatants using a calibration curve obtained with a different protein. Further benefits include the short assay time of approximately 5 min, the small amount of sample required (35 μl per injection) and the ability to monitor the binding event in real time [66].

We have applied this system for two different models. We were able to generate peptides that restored FVIII activity in the

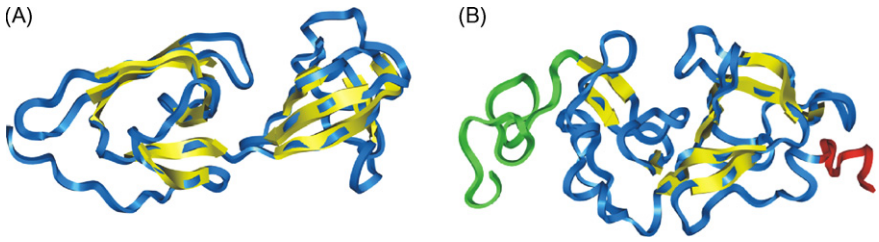

Fig. 9. Presentation of a random sequence as C-terminal extension of eIF5a secreted to the culture supernatant. (A) The eIF5a molecule. (B) The eIF5a molecule modified by an amino-terminal FLAG-tag and a carboxy-terminal random sequence.

presence of an inhibitory antibody. Their specificity was confirmed by displacement assays. Two peptides showed the ability to restore the factor VIII activity from 33% up to approximately 90% in functional tests performed *in vitro*. As a second example of this versatile approach we developed a ligand for the human IgG-Fc fragment. Ligands binding IgG-Fc have therapeutic potential, benefits in the large-scale purification of antibodies and applications in diagnostic tests. Through the screening of only 6160 clones, we identified a ligand – a peptide with an affinity constant of $3.9 \times 10^5\ M^{-1}$. Structure modeling indicates that the random peptide is ideally exposed on the outside of the core molecule accessible for protein interactions (see Fig. 9). Singularized cells are cultivated in microwells and following screening can be performed on nitrocellulose membranes or in an ELISA format independently of the cell viability allowing the use of harsh environmental conditions. This system provides a high applicability for fully automated high-throughput screening platforms. For further characterization of candidate peptides whereas higher concentrations are needed, the cells are simply grown at a larger scale without the need to alternate the expression system.

4.5. Baculovirus display

Foreign proteins have been displayed on the surface of insect cells [67,68], in occlusion bodies [69] and on the surface of the baculovirus [70]. Fusion proteins with baculoviral envelope protein gp64 [70], with the gp64 anchor sequence [71] (see Fig. 10),

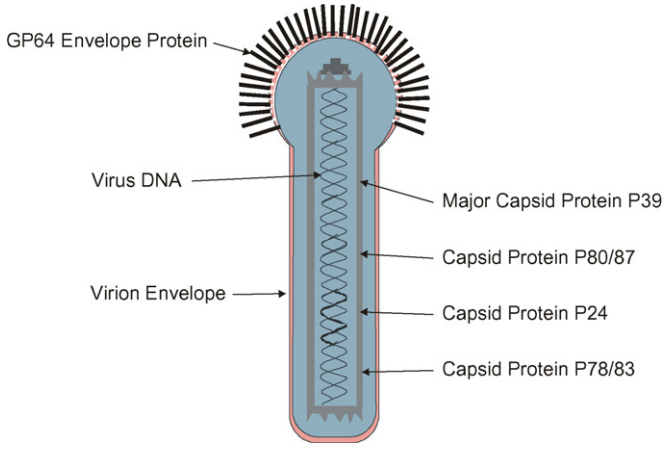

Fig. 10. *Autographa californica* Nuclear Polyhedrosis Virus used in Baculovirus surface display. Drawing kindly provided by R. Grabherr.

as well as foreign membrane proteins (such as the influenza virus hemagglutinin [72]), have all demonstrated targeting to the surface of infected insect cells. Ernst et al. [73] described the first insect-cell-based library screening.

4.6. Mammalian display

Peptides have also been displayed by using several eukaryotic RNA viruses that permit the insertion of short peptides into their native envelope proteins at distinct locations. The development of phage-like methodologies with the benefits of posttranslational modifications is possible, due to the identification of coat protein fusions that do not interfere with the retroviral infectivity.

Smith and coworkers [74,75] used human rhinovirus for the generation of peptide display libraries. Buchholz et al. [76] have demonstrated, that a short peptide library may be displayed and screened on Moloney murine leukaemia virus. Recent publications describe the use of this system for different screening applications, such as those by Urban et al. [77], who used this virus to present a peptide library for the selection of functional human antibodies. Wolkowicz et al. [78] described a library fused to the mammalian cell membrane chemokine receptor CCR5.

5. Considerations in selecting a screening platform

Given the growing number of options available for both, creating polypeptide libraries and screening them, the selection of the appropriate technology for any given application depends on careful consideration of selection criteria (see Table 2), which include: available size of the library, peptide size, biosynthetic capabilities of the system, and quantitative discrimination from false screening positives. The biosynthetic abilities are of utmost importance in the fields of mammalian functional genomics, where authentic posttranslational processing is crucial.

Biopanning is the method of choice for phage displayed libraries. The target molecule is immobilized on a polystyrene surface (immunotubes, microplate wells, or beads) and aspecific sites are blocked. The use of liposomes [79,80] may enhance structure stability of proteins which tend to collapse during the direct adsorption to the surface. Kumada et al. [81] recently presented an anti styrene peptide tag allowing affinity adsorption of the proteo-liposomes. The display library is then incubated with the target and the unbound clones are removed by washing steps. Bound clones are specifically eluted, amplified and used for further rounds of selection.

Table 2
Comparison of most common biological peptide display systems

Property	Ribosome display	Phage display	Bacterial display	Yeast display	Mammalian cell-based display
Theoretical upper limit of library size	10^{15}	$<10^{11}$	10^{9}	10^{8}	10^{8}
Expression host	*In vitro*	Prokaryote	Prokaryote	Yeast cell	Mammalian cell
Linkage	Non covalent or covalent	Viral capsid	Cell	Cell	Cell
Insert size restriction	+	+	−	−	−
Folding machinery	−	Nonnative	Nonnative	Native	Native
Posttranslational modifications	−	−	−	±	+

Fluorescence-activated cell sorting (FACS), in which incubation with fluorescently labeled molecules permits the separation of those cells able to bind the target, gives cell surface displayed systems several advantages. Cell sorting can highly enrich positive clones and also discriminate between clones of different affinity and specifity. FACS allows screening with the target molecule in solution, obviating the need for elution steps. This avoids both the elution problem of very tightly binding clones and sidesteps the necessity to isolate clones that bind unspecifically to the solid support. Cell surface displayed libraries can also be enriched by magnetic particle technology.

The detection of proteins, expressed soluble in cell's cytoplasm, usually requires a lysis step to access the intracellular products. In this process, single colonies are transferred to membranes, where they are lysed and incubated with the target molecule. The bound target molecule is usually detected by use of a labeled second ligand. The use of a system where the protein is secreted from cells renders a lysis step needless and allows straight forward screening.

6. Conclusions

The growing interest of researchers and biopharmaceutical companies in protein–protein interactions has led to a demand for sophisticated methods which allow the rapid identification, characterization and potential improvement of interaction partners. Random libraries are mainly employed in the fields of ligand development and peptide-based drug design. Genomic libraries, on the other hand, are powerful tools when used in such in genomic applications as the detection of gene expression linkage, identification of molecular markers and the search for insight into intracellular signal transduction mechanisms.

The choice of an adequate platform is a crucial decision, and one must consider the often conflicting requirements of different screening applications. None of the existing biological library system can be seen as "gold standard" superior to other systems. Each host- or expression-system provides diverse assets and drawbacks. First, each system is restricted by specific limitations and biases. Second, the screening environment has a formidable influence on the binding event. Desired parameters like extreme pH, increased temperature, high salt concentration, or the presence of denaturing agents may conflict with the physiological needs of living cells. Third, the available laboratory equipment may constrict the number of usable systems. Simple screening procedures, like phage panning on polystyrene surfaces, are far easier to facilitate than fully automated high-throughput methods.

Until now, biological libraries have contributed to the development of novel therapeutics mainly through the well established phage display, yeast two-hybrid, ribosome display and yeast display systems. The future development and refinement of individual platforms aside, the sophisticated combination of alternative systems may enable discovery of powerful bioactive peptides that previously had gone undetected.

Peptides intended for use as ligands for affinity chromatography are very difficult to select. It is not fully clear, why these strategies do not yield ligands with sufficient affinity. It is very likely that these peptides nestle at surface of the protein and do not find a cavity. Furthermore small peptides may form large aggregates through β formation. A tight but unspecific interaction with such ligands is the result. During screening a much lower ligand density is applied, thus the attractive possibility of creating small affinity ligands rapidly could not be realized until now. These peptides seem to nestle on the surface of a protein and do not provide enough sites for biospecific recognition. Ongoing research is necessary to fully understand these phenomena and to exploit these technologies for protein purification applications.

References

[1] T.H. LaBean, S.A. Kauffman, Protein Sci. 2 (1993) 1249.
[2] P. Neuner, R. Cortese, P. Monaci, Nucleic Acids Res. 26 (1998) 1223.
[3] E. Koivunen, B. Wang, E. Ruoslahti, Biotechnology (N Y) 13 (1995) 265.
[4] L.B. Giebel, R.T. Cass, D.L. Milligan, D.C. Young, R. Arze, C.R. Johnson, Biochemistry 34 (1995) 15430.
[5] C.P. Scott, E. Abel-Santos, A.D. Jones, S.J. Benkovic, Chem. Biol. 8 (2001) 801.
[6] C.P. Scott, E. Abel-Santos, M. Wall, D.C. Wahnon, S.J. Benkovic, Proc. Natl. Acad. Sci. USA 96 (1999) 13638.
[7] D.J. DeMarini, C.L. Creasy, Q. Lu, J. Mao, S.A. Sheardown, G.M. Sathe, G.P. Livi, Biotechniques 30 (2001) 520.
[8] M.T. Worthington, J. Pelo, R.Q. Lo, Anal. Biochem. 294 (2001) 169.
[9] J.F. Reidhaar-Olson, R.T. Sauer, Science 241 (1988) 53.
[10] L. Holm, A.K. Koivula, P.M. Lehtovaara, A. Hemminki, J.K. Knowles, Protein Eng. 3 (1990) 181.
[11] E. Merino, J. Osuna, F. Bolivar, X. Soberon, Biotechniques 12 (1992) 508.
[12] J.D. Hermes, S.M. Parekh, S.C. Blacklow, H. Koster, J.R. Knowles, Gene 84 (1989) 143.
[13] L.C. Mattheakis, R.R. Bhatt, W.J. Dower, Proc. Natl. Acad. Sci. USA 91 (1994) 9022.
[14] M. He, M.J. Taussig, Nucleic Acids Res. 25 (1997) 5132.
[15] J. Hanes, A. Pluckthun, Proc. Natl. Acad. Sci. USA 94 (1997) 4937.
[16] R.W. Roberts, J.W. Szostak, Proc. Natl. Acad. Sci. USA 94 (1997) 12297.
[17] D.S. Tawfik, A.D. Griffiths, Nat. Biotechnol. 16 (1998) 652.
[18] G.P. Smith, Science 228 (1985) 1315.
[19] A.B. Sparks, L.A. Quilliam, J.M. Thorn, C.J. Der, B.K. Kay, J. Biol. Chem. 269 (1994) 23853.
[20] J. Greenwood, A.E. Willis, R.N. Perham, J. Mol. Biol. 220 (1991) 821.
[21] J. Greenwood, G.J. Hunter, R.N. Perham, J. Mol. Biol. 217 (1991) 223.

[22] G.F. Simons, R.N. Konings, J.G. Schoenmakers, Proc. Natl. Acad. Sci. USA 78 (1981) 4194.
[23] R. Imber, A. Tsugita, M. Wurtz, T. Hohn, J. Mol. Biol. 139 (1980) 277.
[24] C. Krebber, S. Spada, D. Desplancq, A. Krebber, L. Ge, A. Pluckthun, J. Mol. Biol. 268 (1997) 607.
[25] J.K. Osbourn, J.C. Earnshaw, K.S. Johnson, M. Parmentier, V. Timmermans, J. McCafferty, Nat. Biotechnol. 16 (1998) 778.
[26] J.K. Osbourn, E.J. Derbyshire, T.J. Vaughan, A.W. Field, K.S. Johnson, Immunotechnology 3 (1998) 293.
[27] R. Freudl, S. MacIntyre, M. Degen, U. Henning, J. Mol. Biol. 188 (1986) 491.
[28] M. Hofnung, Methods Cell. Biol. 34 (1991) 77.
[29] M. Agterberg, H. Adriaanse, A. van Bruggen, M. Karperien, J. Tommassen, Gene 88 (1990) 37.
[30] P. Fuchs, F. Breitling, S. Dubel, T. Seehaus, M. Little, Biotechnology (N Y) 9 (1991) 1369.
[31] J.A. Francisco, C.F. Earhart, G. Georgiou, Proc. Natl. Acad. Sci. USA 89 (1992) 2713.
[32] R.S. Wong, R.A. Wirtz, R.E. Hancock, Gene 158 (1995) 55.
[33] T. Suzuki, M.C. Lett, C. Sasakawa, J. Biol. Chem. 270 (1995) 30874.
[34] T. Klauser, J. Pohlner, T.F. Meyer, Embo. J. 9 (1990) 1991.
[35] Z. Lu, K.S. Murray, V. Van Cleave, E.R. LaVallie, M.L. Stahl, J.M. McCoy, Biotechnology (N Y) 13 (1995) 366.
[36] M. Hansson, S. Stahl, T.N. Nguyen, T. Bachi, A. Robert, H. Binz, A. Sjolander, M. Uhlen, J. Bacteriol. 174 (1992) 4239.
[37] W. Mandecki, Protein Eng. 3 (1990) 221.
[38] M.G. Cull, J.F. Miller, P.J. Schatz, Proc. Natl. Acad. Sci. USA 89 (1992) 1865.
[39] T.H. LaBean, S.A. Kauffman, T.R. Butt, Mol. Divers. 1 (1995) 29.
[40] K. Bussow, D. Cahill, W. Nietfeld, D. Bancroft, E. Scherzinger, H. Lehrach, G. Walter, Nucleic Acids Res. 26 (1998) 5007.
[41] K. Bussow, E. Nordhoff, C. Lubbert, H. Lehrach, G. Walter, Genomics 65 (2000) 1.
[42] S. Fields, O. Song, Nature 340 (1989) 245.
[43] J.C. Hu, E.K. O'Shea, P.S. Kim, R.T. Sauer, Science 250 (1990) 1400.
[44] M.G. Kornacker, B. Remsburg, R. Menzel, Mol. Microbiol. 30 (1998) 615.
[45] S.L. Dove, J.K. Joung, A. Hochschild, Nature 386 (1997) 627.
[46] G. Karimova, J. Pidoux, A. Ullmann, D. Ladant, Proc. Natl. Acad. Sci. USA 95 (1998) 5752.
[47] J.N. Pelletier, F.X. Campbell-Valois, S.W. Michnick, Proc. Natl. Acad. Sci. USA 95 (1998) 12141.
[48] C.T. Chien, P.L. Bartel, R. Sternglanz, S. Fields, Proc. Natl. Acad. Sci. USA 88 (1991) 9578.
[49] R.K. Brachmann, J.D. Boeke, Curr. Opin. Biotechnol. 8 (1997) 561.
[50] E.M. Phizicky, S. Fields, Microbiol. Rev. 59 (1995) 94.
[51] Y.C. Broder, S. Katz, A. Aronheim, Curr. Biol. 8 (1998) 1121.
[52] N. Johnsson, A. Varshavsky, Proc. Natl. Acad. Sci. USA 91 (1994) 10340.
[53] I. Stagljar, C. Korostensky, N. Johnsson, S. te Heesen, Proc. Natl. Acad. Sci. USA 95 (1998) 5187.
[54] P.N. Lipke, J. Kurjan, Microbiol. Rev. 56 (1992) 180.
[55] E.T. Boder, K.D. Wittrup, Nat. Biotechnol. 15 (1997) 553.
[56] M.C. Kieke, B.K. Cho, E.T. Boder, D.M. Kranz, K.D. Wittrup, Protein Eng. 10 (1997) 1303.
[57] B.L. Miki, N.H. Poon, A.P. James, V.L. Seligy, J. Bacteriol. 150 (1982) 878.
[58] M. Bony, D. Thines-Sempoux, P. Barre, B. Blondin, J. Bacteriol. 179 (1997) 4929.
[59] T. Matsumoto, H. Fukuda, M. Ueda, A. Tanaka, A. Kondo, Appl. Environ. Microbiol. 68 (2002) 4517.
[60] N. Sato, T. Matsumoto, M. Ueda, A. Tanaka, H. Fukuda, A. Kondo, Appl. Microbiol. Biotechnol. 60 (2002) 469.
[61] A. Lueking, C. Holz, C. Gotthold, H. Lehrach, D. Cahill, Protein Expr. Purif. 20 (2000) 372.
[62] C. Holz, A. Lueking, L. Bovekamp, C. Gutjahr, N. Bolotina, H. Lehrach, D.J. Cahill, Genome Res. 11 (2001) 1730.
[63] C. Mersich, W. Billes, I. Pabinger, A. Jungbauer, Biotechnol. Bioeng. 98 (2007) 12.
[64] C. Mersich, A. Jungbauer, Biotechnol. J. 2 (2007) 672.
[65] K.S. Prickett, D.C. Amberg, T.P. Hopp, Biotechniques 7 (1989) 580.
[66] C. Mersich, A. Jungbauer, J. Biochem. Biophys. Methods 70 (2007) 555.
[67] D.L. Domingo, I.S. Trowbridge, J. Biol. Chem. 263 (1988) 13386.
[68] H. Ghiasi, A.B. Nesburn, S.L. Wechsler, Virology 185 (1991) 187.
[69] J.H. McLinden, V.A. Ploplis, M.J. Fraser, E.D. Rosen, Vaccine 10 (1992) 231.
[70] Y. Boublik, P. Di Bonito, I.M. Jones, Biotechnology (N Y) 13 (1995) 1079.
[71] R. Grabherr, W. Ernst, O. Doblhoff-Dier, M. Sara, H. Katinger, Biotechniques 22 (1997) 730.
[72] R.D. Possee, Virus Res. 5 (1986) 43.
[73] W. Ernst, R. Grabherr, D. Wegner, N. Borth, A. Grassauer, H. Katinger, Nucleic Acids Res. 26 (1998) 1718.
[74] A.D. Smith, D.A. Resnick, A. Zhang, S.C. Geisler, E. Arnold, G.F. Arnold, J. Virol. 68 (1994) 575.
[75] D.A. Resnick, A.D. Smith, A. Zhang, S.C. Geisler, E. Arnold, G.F. Arnold, AIDS Res. Hum. Retroviruses 10 (Suppl. 2) (1994) S47.
[76] C.J. Buchholz, K.W. Peng, F.J. Morling, J. Zhang, F.L. Cosset, S.J. Russell, Nat. Biotechnol. 16 (1998) 951.
[77] J.H. Urban, R.M. Schneider, M. Compte, C. Finger, K. Cichutek, L. Alvarez-Vallina, C.J. Buchholz, Nucleic Acids Res. 33 (2005) e35.
[78] R. Wolkowicz, G.C. Jager, G.P. Nolan, J. Biol. Chem. 280 (2005) 15195.
[79] L.K. Jespersen, A. Kuusinen, A. Orellana, K. Keinanen, J. Engberg, Eur. J. Biochem. 267 (2000) 1382.
[80] Y. Kumada, M. Nogami, N. Minami, M. Maehara, S. Katoh, J. Chromatogr. A 1080 (2005) 22.
[81] Y. Kumada, S. Katoh, H. Imanaka, K. Imamura, K. Nakanishi, J. Biotechnol. 127 (2007) 288.
[82] M. Ueda, A. Tanaka, Biotechnol. Adv. 18 (2000) 121.
[83] J.K. Scott, G.P. Smith, Science 249 (1990) 386.
[84] G.P. Smith, J.K. Scott, Methods Enzymol. 217 (1993) 228.
[85] M. Scalley-Kim, P. Minard, D. Baker, Protein Sci. 12 (2003) 197.