DOWNLOAD PDF SCREENING EXPRESSION LIBRARIES

Chapter 1 : Expression Library Screening (Procaryotic)

When trying to identify a clone within a cDNA library, which may contain a coat protein (CP) gene, one useful technique may by immunological screening, using antibodies raised against either purified.

Vortex the tube to release the phage particles into the SM buffer. In a 50ml conical tube combine: Add XL1-Blue cells and helper phage alone, no recombinant phage. Decant supernatant into a sterile tube. To plate the rescued phagemid, combine the following in two 15ml tubes: White colonies on the plate contain the pBluescript plasmid with the cloned DNA insert. The number of rescued colonies increases linearly with increased incubation time in step 3. If low numbers of colonies are rescued, increase the incubation time. If a rescue is unsuccessful, it may be necessary to make a high titer stock of the phage for the rescue procedure. About 10, phage are required for a successful rescue using the above protocol. Helper phage titers make little difference in rescue efficiencies. However, it was observed that different helper phage strains contain up to 10 fold variations in the rescue efficiency. A variety of helper phage strains are available from Stratagene. R is recommended for its stability during preparation of the helper phage stocks. If ampicillin resistent colonies grow slowly, it may be due to the presence of helper phage in the cell. Helper phage can be easily removed by increasing the ratio of bacteria to phage during the plating in step 6 above. Cloning of ligand targets: Systematic isolation of SH3 domain-containing proteins. Identification and characterization of Src SH3 ligands from phage- displayed random peptide libraries. Cold Spring Harbor Laboratory Press,

Chapter 2 : Phage Display Library Screening - Creative Biolabs

This session will review how to make a recombinant cDNA library and how to use this library to find a specific gene. This session will outline the differences between a genomic and a cDNA library, and discuss how to use a cDNA library to clone a gene of interest. To understand what a recombinant.

At the end of the page we will briefly summarize the reason that cDNAs can be extremely valuable in experimental design, although many of these should already be obvious to most readers. It is also true that there are dozens of different approaches to isolating cDNAs of interest and these will be briefly described in the second part of the section. We will begin by describing how a cDNA for a known protein can be isolated using amino acid sequence information, which, historically, was the first way that a cDNA encoding for a known protein was isolated. Let us begin with three definitions: First, what does it mean to clone? Cloning refers to the isolation of a genetically homogeneous strain of any organism. Within a clone, all organisms are identical to all other organisms at a genetic level. It is possible to clone bacteria or phage or even higher plants by isolating a single cell and allowing that single cell to produce a colony, or a plaque, or an entire plant. Since most plants are derived from a single cell with a unique genotype, the act of rooting leaves to produce a collection of identical African violets is cloning. In some cases cDNA cloning may simply refer to the isolation of any single cDNA, since, in some circumstances, an experimentalist may be interested in any cDNA produced by a particular tissue. More frequently, the challenge of cDNA cloning is not the isolation of any cDNA but the selection of a single cDNA that is of interest to the experimentalist for a particular reason. In the same way it is possible to isolate clones that are not cDNA clones but rather are genomic clones. Genomic clones are simply DNA derived directly from a genome. Genomic DNA would incorporate some sequences such as introns or regulatory sequences that would not be found in cDNAs. Likewise, the isolation of a monoclonal antibody refers to the isolation of a single cell that expresses a mRNA for a unique antibody. Thus, making monoclonal antibodies an exercise in cloning. The second concept that is important in understanding the strategy needed to isolate a cDNA clone, a genomic clone, or even a monoclonal antibody is the idea of a library. For experimental convenience, vectors are usually derivatives of viruses plasmids, bacteriophages, animal viruses, retroviruses. Since the essence of being able to isolate clones is the ability to replicate to make large amounts of biological material, the essence of a vector is that it must incorporate some mechanism of reproduction. Thus, one would expect that vectors would incorporate an origin of DNA replication. Since vectors are an important experimental approach, a considerable amount of effort has gone into designing vectors that are particularly easy to use in an experimental sense. It would be impossible to provide even a brief description of the tricks that have been incorporated into various classes of vectors. The incorporation of selectable markers is certainly a significant experimental advantage in many cases. The underlying experimental approach to cloning can be divided into four parts. First, it is necessary to produce or obtain a library including the sequence of interest. Second, it is necessary to isolate clones that may be of interest. Third, it is essential to develop a formal test to ensure that the clones that have been isolated are indeed the correct clones. Fourth, it is essential to put the cDNA that has been isolated to some interesting biological use. There are a number of different criteria that might be used to judge the quality of a cDNA library. A cDNA library is generally better if the size of the inserts that is the amount of continuous cDNA in each clone is large, ideally full-length. The library should be sufficiently large that it contains the cDNA of interest or,more precisely, it should have enough independently derived clones that it contains the cDNA of interest. In general this means that it should be representative of all the mRNAs present in a particular tissue. Of course, choosing a tissue that has a relatively large amount of the mRNA of interest is an important experimental choice. In general it is easier to isolate a cDNA from a library where it is represented many times than from a library where it is present rarely. Some characteristics of a library depend on the vector chosen. Vectors are frequently chosen because they allow the screening of a large number of independent members of the library with experimental ease. Some vectors are designed to express only the cDNAs, while others have been modified to express not only the cDNA but also to express it in a context so the cDNA is made into a protein or a fusion protein. Fusion proteins will be discussed below. Before using a cDNA library it is wise to determine if it is a good quality library. More than one student has wasted months of time screening a library that had no inserts or inserts so short that they were of little value. Like all DNA polymerases it cannot initiate synthesis de novo but depends on the presence of a primer. This two-step procedure has been optimized to maximize fidelity and length of cDNAs. Incorporating cDNA into the vector. One of the most convenient ways of doing this is to attempt to manipulate the cDNAs so that each one has a unique restriction site at those ends. To do this, the cDNAs are frequently methylated with a specific methyl transferase that incorporates a methyl group into particular restriction site to protect them from the restriction enzyme that will be used later. It is then possible to ligate a synthetic oligonucleotide to the ends of this cDNA. Blunt end ligation is generally a low efficiency process; but, by using a high concentration of these synthetic oligonucleotides, it is possible to drive the reaction to near completion. The value of producing an overhang is that it will facilitate the introduction of the cDNA into a vector. The vector can also be prepared by treating it with the same nuclease, or a nuclease that produces the same restriction site, to produce a single-stranded region that is complementary to the single-stranded region in the cDNA. Mixing the cDNA of interest with the vector in the presence of ligase allows incorporation of the cDNA into the vector. One of the experimental difficulties in doing this is that the vector itself will have a high tendency to re-ligate to form a vector without any cDNA insert. This is frequently minimized by treating the vector with the phosphatase to remove the terminal phosphates. These phosphates are required for ligase to act, so this strategy prevents this unwanted side reaction. The choice of the vector used also has an important impact on experimental outcome. Initially, plasmids were chosen as vectors and were modified to include markers that could be used to determine whether a plasmid had been introduced into a bacterial cell or whether there was a cDNA insert in the cloning site. More recently, derivatives of bacteria phage lambda has been made that can be effective vectors for cDNA cloning. The extent of understanding of lambda and lambda genetics has made it possible to isolate lambda derivatives where some non-essential genes have been removed making it possible to carry inserts of up to 11 kb of cDNA, which is a convenient size and sufficient for the isolation of most cDNAs. The lambda genome is a linear molecule when it is packaged into the bacteriophage and the cDNA can be incorporated into the central region of the DNA. The lambda "arms" the more distal parts of the DNA encode all the essential information for replication of lambda in an infectious cycle. If the lambda arms re-ligate in the absence of an insert, and an appropriate host is chosen hfl, for high frequency of lysogeny , then these particles will not form plaques. Thus only particles carrying an insert will form plaques. The remarkable power of bacteriophage lambda as a vector is that once the cDNA has been ligated into the lambda arms, the DNA can then be incorporated into a phage particle in vitro. Extracts prepared from cells that have all the necessary proteins for the assembly of lambda can then be mixed with the library DNA and ATP and particles will be assembled! These particles can then be used to infect E coli and each individual plaque is an independent clonal population which represents a single cDNA species. This ability can be used both to amplify the cDNA library which is somewhat dangerous because repeated amplification can lead to a loss of some cDNA sequences and for the screening of the cDNA library to isolate the cDNA of interest. A lambda -gt10 library can be conveniently screened by plating it at relatively high concentrations on a bacterial lawn of E coli. High density screening allows the experimentalist to screen between , and 1,, independent plaques on a single plate and makes it theoretically possible to screen for a cDNA that is present only at one copy per cell in a particular tissue. Screening is done by a "replica plating" procedure. After the phage infect E coli and form individual plaques, a perfect spatial representation of the infected plaque can be produced by placing a piece of nitrocellulose on top of the lawn of E-coli. Nitrocellulose binds DNA with great avidity and so some of the DNA of each plaque can be transferred to nitrocellulose paper or even several different nitrocellulose papers. The DNA from the library can then be cross-linked to the filter and extraneous protein can be washed off. The plaques of interest can then be screened using a hybridization assay. This takes us to the question of how a library can be screened to isolate candidates for the cDNA of interest. One of the most straight forward ways to do this is to take advantage of DNA hybridization. If one can design an oligonucleotide that is complementary to the mRNA of interest this can be used to screen the library. Such an oligonucleotide probe can be designed by sequence information from the amino acid sequence of a known protein. In the 50s and 60s biochemical methods were developed to produce amino acid sequence of overlapping fragments of known purified proteins. Our task is much simpler. It is now necessary only to know the amino acid sequence of a couple of regions of the protein. To do this, a purified protein is generally digested either with proteases or biochemical method to produce a series of peptides. Unlike proteins, which must be treated with care to ensure that they retain their native conformation, peptides can be treated as bio-organic molecules. They can be fractionated by fairly standard procedures using HPLC high pressure liquid chromatography which is capable of resolving individual peptides. If a series of individual peptides can be resolved, the sequence of those peptides can be determined, or at least partially determined, by Edmund degradation. This series of reaction cleaves individual amino acids one at a time from a peptide and the resultant amino acid derivatives can be identified. This procedure can produce sequence information on a series of peptides. To do this intelligently it is essential that each of the peptides is derived from a single protein molecule, and the criterion for insuring that this is likely to be the case were discussed in the section on protein purification. Edmund degradation works via removing single amino acids from the N-terminal end and can in some cases be applied to an intact protein, however, generally the N-terminal amino group is chemically modified so this approach usually fails. A probe is an oligonucleotide that is designed to be complementary to the mRNA of interest so that it can be used to screen a library. Of course, any mRNA produces a unique polypeptide when it is translated; but the reverse is not true. Because the triplet code is degenerate, there are many mRNA sequences that might produce the same amino acid sequences. Because of this the design of an oligonucleotide probe is not straight forward, but a clever experimentalist can make good choices in designing a probe. There are basically two strategies that can be used. Either the experimentalist can choose to design a relatively short oligonucleotide that hopefully will have a high degree of homology to the mRNA of interest or the experimentalist can choose to design a longer probe that is more likely to have some regions that are not complementary to the mRNA of interest but hopefully will have at least some sequences that can form a stable duplex. In many cases it makes sense to make a mixture of different probes, which are homologous, but have different bases in positions where it is not possible to make a good prediction of which one should be present. This is called degeneracy. The choice of which strategy depends on the amino acid sequences that are available. There are a number of other factors that should also be taken into consideration. In many organisms, there is a preference for the use of particular triplets over the use of other triplets codon utilization. Designing a probe that has homology to a known mRNA is generally not recommended since this may lead to the cloning of the wrong cDNA.

Chapter 3 : cDNA library - Wikipedia

An immunological screening method employing protein A-peroxidase which does not require radiolabelled antibodies for detection of Escherichia coli colonies synthesizing foreign proteins in a cDNA expression library is described.

A method of constructing an expression library for expressing a peptide having a conformation sufficient for binding to a target protein or nucleic acid, said method comprising: The method of claim 1 further comprising selecting nucleic acid fragments from the fragments at a that have substantially different nucleotide sequences thereby enhancing nucleotide sequence diversity among the selected fragments compared to the diversity of sequences in the genome. The method of claim 1, further comprising introducing the recombinant construct into a host cell. The method according to claim 1 wherein producing nucleic acid fragments comprising amplifying nucleic acid from two or more microorganisms or eukaryotes comprising compact genomes. The method according to claim 1 wherein producing nucleic acid fragments comprises amplifying genomic DNA from two or more microorganisms or eukaryotes comprising compact genomes. The method according to claim 1 wherein step 1 c comprises inserting the nucleic acid fragments at b into a suitable expression construct in equimolar amounts thereby producing recombinant constructs, wherein each fragments is in operable connection with a promoter sequence that is capable of conferring expression on that fragment. The method according to claim 1 wherein the two or more microorganisms or compact eukaryotes are selected from the group consisting of Aeropyrum pernix, Anopheles gambiae, Arabidopsis thaliana, Aquifex aeolicus, Archaeoglobus fulgidis, Bacillus subtilis, Bordetella pertussis, Borrelia burgdorferi, Caenorhabditis elegans, Chlamydia trachomatis, Danio rerio, Drosophila melanogaster, Escherichia coli, Haemophilus influenzae, Helicobacter pylori, Methemobacterium thermoautotrophicum, Methanococcus jannaschii, Mycoplasma pneumoniae, Neisseria meningitidis, Pseudomonas aeruginosa, Pyococcus horikoshii, Saccharomyces cerevesiae, Schizosaccharomyces pombe, Synechocystis PCC , Takifugu rubripes, Thermoplasma volcanium, and Thermotoga maritima. The method according to claim 1 wherein the two or more microorganisms are selected from the group consisting of Archaeoglobus fulgidis, Aquifex aeolicus, Aeropyrum pernix, Bacillus subtilis, Bordetella pertussis TOX6, Borrelia burgdorferi, Chlamydia trachomatis, Escherichia coli K12, Haemophilus influenzae rd , Helicobacter pylori, Methanobacterium thermoautotrophicum, Methanococcus jannaschii, Mycoplasma pneumoniae, Neisseria meningtidis, Pseudomonas aeruginosa, Pyrococcus horikoshii, Synechocystis PCC , Theroplasma volcanium and Thermotoga maritima. The method according to claim 1 wherein the promoter sequence is operable in a yeast cell. The method according to claim 1 wherein the promoter sequence is operable in a bacterial cell. The method according to claim 1 wherein the expression construct is a yeast expression vector. The method according to claim 1 wherein the expression construct is a phage display vector. The method according to claim 3 wherein the host cell is a yeast cell. The method according to claim 3 wherein the host cell is a bacterial cell. FIELD OF THE INVENTION The present invention relates generally to methods for the production and of nucleic acid fragment libraries that express highly diverse peptides, polypeptides or protein domains and, in particular, methods for producing nucleic acid fragment libraries wherein the nucleic acid fragments of the libraries are derived from one and preferably from two or more prokaryote genomes or compact eukaryote genomes, such as, for example, organisms having diverse characterized genomes. In another embodiment, the nucleic acid fragments are expressed as protein domains capable of assuming a conformation that binds to a target protein or nucleic acid during library screening. The present invention further provides methods of screening such libraries to identify peptides, polypeptides or protein domains that bind to a target protein or nucleic acid such as, for example, to modulate the activity of the target protein or nucleic acid. Also provided are methods for identifying nucleic acid encoding such peptides, polypeptides or protein domains. The present invention extends to the nucleic acids, peptides, polypeptides and protein domains identified by the methods described herein. Unless the context requires otherwise or specifically stated to the contrary, integers, steps, or elements of the invention recited herein as singular integers, steps or elements clearly encompass both singular and plural forms of the recited integers, steps or elements. Unless specifically stated otherwise, each feature described herein with reference to a particular aspect or embodiment of the invention shall be taken to apply mutatis mutandis to each and every other aspect or embodiment of the invention. For example, any one or more features described herein with respect to methods for expression library construction shall apply to those embodiments relating to methods for screening expression libraries to identify a peptide or protein domain capable of binding a target protein or nucleic acid or nucleic acid encoding same. Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations or any two or more of said steps or features. The present invention is not to be limited in scope by the specific examples described herein. Functionally equivalent products, compositions and methods are clearly within the scope of the invention, as described herein. The present invention is performed without undue experimentation using, unless otherwise indicated, conventional techniques of molecular biology, microbiology, virology, recombinant DNA technology, peptide synthesis in solution, solid phase peptide synthesis, and immunology. Such procedures are described, for example, in the following texts: A Practical Approach, Vols. I and II D. A Practical Approach M. A Practical Approach B. Immobilized Cells and Enzymes: Methods In Enzymology S. Handbook of Experimental Immunology, Vols. Methods in Yeast Genetics: Guide to Yeast Genetics and Molecular Biology. Methods in Enzymology Series, Vol. Fink eds Academic Press, London, see whole of text. Description of the Related Art As a response to the increasing demand for new lead compounds and new target identification and validation reagents, the pharmaceutical industry has increased its screening of various sources for new lead compounds having a unique activity or specificity in therapeutic applications, such as, for example, in the treatment of neoplastic disorders, infection, modulating immunity, autoimmunity, fertility, etc. It is known that proteins bind to other proteins, antigens, antibodies, nucleic acids, and carbohydrates. Such binding enables the protein to effect changes in a wide variety of biological processes in all living organisms. As a consequence, proteins represent an important source of natural modulators of phenotype. Accordingly, peptides that modulate the binding activity of a protein represent attractive lead compounds drug candidates in primary or secondary drug screening. For example, the formation of a target biological interaction that has a deleterious effect eg. Similarly, the activity or expression of an antimicrobial target eg. Peptides that block the function of specific membrane channels, or disrupt cytoplasmic membranes of some organisms is represent attractive candidates for anti-microbial drugs. Antimicrobial effects have been demonstrated for certain natural peptides produced by animals and insects, and for synthetic cationic peptides eg. A virulence determinant of a pathogen also presents an attractive target for identifying lead compounds having antimicrobial activity. For example, a peptide antagonist of an autoinducer of virulence in Staphylococcus aureus that controls the production of bacterial toxins involved in pathogenesis has been determined. In another example, differential gene expression between normal and diseased eg. Accordingly, the genes or proteins that are differentially expressed in diseased and normal cells, or the differential cellular processes between normal and diseased cells, form attractive targets for therapy. Similarly, cyclin proteins such as Cdc2, Cdc25, and cyclin-dependent kinases CDKs are attractive targets for cellular proliferation. Peptides that agonize or antagonize the expression of such target genes or target processes are suitable lead compounds for therapeutic applications. In yet another example, certain allergen proteins eg. It is widely recognized that there is a need to develop methods for determining novel compounds, including nucleic acid-based products and peptide-based products, that modulate an activity or function of a particular target. In such approaches, an activity of a target protein or nucleic acid is screened in the absence and presence of a potential lead compound, which is a peptide, and modified activity of the target is determined. Similarly, peptides can be used as dominant negative inhibitors or the validation of prospective drug targets using assays such as observing the phenotype resulting from over-expression of the peptides in ex-vivo assays or in transgenic mice. In one known approach to identify novel lead compounds, random peptide synthetic mimetic or mimotope libraries are produced using short random oligonucleotides produced by synthetic combinatorial chemistry. The DNA sequences are cloned into an appropriate vehicle for expression and the encoded peptide is then screened using one of a variety of approaches. However, the ability to isolate active peptides from random fragment libraries can be highly variable with low affinity interactions occurring between the peptide-binding partners. This is not surprising, considering that biological molecules appear to recognise shape and charge rather than primary sequence Yang and Honig J. Biol 3 , and that such random peptide aptamers are generally too small to comprise a protein domain or to form the secondary structure of a protein domain. To enhance the probability of obtaining useful bioactive peptides or proteins from random peptide libraries, peptides have previously been constrained within scaffold structures, eg. USA, 97, , or catalytically inactive staphylococcal nuclease Norman et al, Science, , , , to enhance their stability. Constraint of peptides within such structures has been shown, in some cases, to enhance the affinity of the interaction between the expressed peptides and its target, presumably by limiting the degrees of conformational freedom of the peptide, and thereby minimizing the entropic cost of binding. It is also known to tailor peptide expression libraries for identifying specific peptides involved in a particular process, eg. The amplified sequences are expressed using a bacterial display system, for screening with selected antigens to determine those antibody fragments that bind the antigens. However, the expression libraries described in U. Additionally, the antibody-encoding libraries described in U. Several attempts have been made to develop libraries based on naturally occurring proteins eg genomic expression libraries. Libraries of up to several thousand polypeptides or peptides have been prepared by gene expression systems and displayed on chemical supports or in biological systems suitable for testing biological activity. For example, genome fragments isolated from Escherichia coli MG have been expressed using phage display technology, and the expressed peptides screened to identify peptides that bind to a polyclonal anti-Rec A protein antisera Palzkill et al. Gene, , Additionally, as many bacteria comprise recA-encoding genes, the libraries described by Palzkill et al. The procedure described by Diversa Corp. Rare sequences, that are less likely to reanneal to their complementary strand in a short period of time, are isolated as single-stranded nucleic acid and used to generate a gene expression library. However, total normalization of each organism within such uncharacterized samples is difficult to achieve, thereby reducing the biodiversity of the library. Such libraries also tend to be biased toward the frequency with which a particular organism is found in the native environment. As such, the library does not represent the true population of the biodiversity found in a particular biological sample. In cases where the environmental sample includes a dominant organism, there is likely to be a significant species bias that adversely impacts on the sequence diversity of the library. Furthermore, as many of the organisms found in such samples are uncharacterized, very little information is known regarding the constitution of the genomes that comprise such libraries. Accordingly, it is not possible to estimate the true diversity of such libraries. Additionally, since the Diversa Corp. Accordingly, there remains a need to produce improved methods for constructing highly diverse and well characterized expression libraries wherein the expressed peptides are capable of assuming a secondary structure or conformation sufficient to bind to a target protein or nucleic acid, such as, for example, by virtue of the inserted nucleic acid encoding a protein domain. USA 91 , Proteins that fold well in nature have non-random hydrophobicity distributions Irback et al. USA 93, , In any native peptide, the distribution of amino acid residues according to their chemical properties eg hydrophobicity, polarity, etc is also non-random Baud and Karlin, Proc Natl Acad. USA 96, , Accordingly, the present inventors realized that random peptide libraries have a low frequency of naturally occurring or native peptide conformational structures or secondary structures, such as, for example, those structures formed by protein domains. It will be understood from the disclosure herein that the bioactive peptides or proteins expressed by individual library clones of such libraries are screened for an activity of the encoded peptide, particularly a binding activity, which said encoded protein has not been shown to possess in the context of the protein from which it was derived ie in its native environment. In the screening process, any library clone encoding a peptide that has the same activity as it would have in its native environment is excluded during the screening process, since an objective of the present invention is to isolate novel bioactive peptides or proteins. Peptides encoded by genomes which differ from the genome of the drug target organism eg. This is because in the evolution of the target organism itself, such high affinity peptide domains have been selected against other than the interaction interfaces which may exist in that organism for functional dimerization with natural partners. In one embodiment, the libraries described in the present invention are constructed from nucleic acid fragments comprising genomic DNA, cDNA, or amplified nucleic acid derived from one or two or more well-characterized genomes. Preferably, one or more well-characterized genomes is a compact genome of a eukaryote ie. In another embodiment, one or more well-characterized genomes is a compact genome of a prokaryote ie.

Chapter 4 : DNA (Gene Libraries): Construction, Genomic Libraries and cDNA Libraries

A number of types of libraries, expression vec- tors, and screening methods have been described but workers in the field of parasitology have generally used antibody screening of cDNA libraries constructed in the bacteriophage Xgtll or its variants.

FIELD OF THE INVENTION The present invention relates generally to methods for the production and of nucleic acid fragment libraries that express highly diverse peptides, polypeptides or protein domains and, in particular, methods for producing nucleic acid fragment libraries wherein the nucleic acid fragments of the libraries are derived from one and preferably from two or more prokaryote genomes or compact eukaryote genomes, such as, for example, organisms having diverse characterized genomes. In another embodiment, the nucleic acid fragments are expressed as protein domains capable of assuming a conformation that binds to a target protein or nucleic acid during library screening. The present invention further provides methods of screening such libraries to identify peptides, polypeptides or protein domains that bind to a target protein or nucleic acid such as, for example, to modulate the activity of the target protein or nucleic acid. Also provided are methods for identifying nucleic acid encoding such peptides, polypeptides or protein domains. The present invention extends to the nucleic acids, peptides, polypeptides and protein domains identified by the methods described herein. Unless the context requires otherwise or specifically stated to the contrary, integers, steps, or elements of the invention recited herein as singular integers, steps or elements clearly encompass both singular and plural forms of the recited integers, steps or elements. Unless specifically stated otherwise, each feature described herein with reference to a particular aspect or embodiment of the invention shall be taken to apply mutatis mutandis to each and every other aspect or embodiment of the invention. For example, any one or more features described herein with respect to methods for expression library construction shall apply to those embodiments relating to methods for screening expression libraries to identify a peptide or protein domain capable of binding a target protein or nucleic acid or nucleic acid encoding same. Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations or any two or more of said steps or features. The present invention is not to be limited in scope by the specific examples described herein. Functionally equivalent products, compositions and methods are clearly within the scope of the invention, as described herein. The present invention is performed without undue experimentation using, unless otherwise indicated, conventional techniques of molecular biology, microbiology, virology, recombinant DNA technology, peptide synthesis in solution, solid phase peptide synthesis, and immunology. Such procedures are described, for example, in the following texts: A Practical Approach, Vols. I and II D. A Practical Approach M. A Practical Approach B. Immobilized Cells and Enzymes: Methods In Enzymology S. Handbook of Experimental Immunology, Vols. Methods in Yeast Genetics: Guide to Yeast Genetics and Molecular Biology. Methods in Enzymology Series, Vol. Fink eds Academic Press, London, see whole of text. Description of the Related Art As a response to the increasing demand for new lead compounds and new target identification and validation reagents, the pharmaceutical industry has increased its screening of various sources for new lead compounds having a unique activity or specificity in therapeutic applications, such as, for example, in the treatment of neoplastic disorders, infection, modulating immunity, autoimmunity, fertility, etc. It is known that proteins bind to other proteins, antigens, antibodies, nucleic acids, and carbohydrates. Such binding enables the protein to effect changes in a wide variety of biological processes in all living organisms. As a consequence, proteins represent an important source of natural modulators of phenotype. Accordingly, peptides that modulate the binding activity of a protein represent attractive lead compounds drug candidates in primary or secondary drug screening. For example, the formation of a target biological interaction that has a deleterious effect eg. Similarly, the activity or expression of an antimicrobial target eg. Peptides that block the function of specific membrane channels, or disrupt cytoplasmic membranes of some organisms is represent attractive candidates for anti-microbial drugs. Antimicrobial effects have been demonstrated for certain natural peptides produced by animals and insects, and for synthetic cationic peptides eg. A virulence determinant of a pathogen also presents an attractive target for identifying lead compounds having antimicrobial activity. For example, a peptide antagonist of an autoinducer of virulence in Staphylococcus aureus that controls the production of bacterial toxins involved in pathogenesis has been determined. In another example, differential gene expression between normal and diseased eg. Accordingly, the genes or proteins that are differentially expressed in diseased and normal cells, or the differential cellular processes between normal and diseased cells, form attractive targets for therapy. Similarly, cyclin proteins such as Cdc2, Cdc25, and cyclin-dependent kinases CDKs are attractive targets for cellular proliferation. Peptides that agonize or antagonize the expression of such target genes or target processes are suitable lead compounds for therapeutic applications. In yet another example, certain allergen proteins eg. It is widely recognized that there is a need to develop methods for determining novel compounds, including nucleic acid-based products and peptide-based products, that modulate an activity or function of a particular target. In such approaches, an activity of a target protein or nucleic acid is screened in the absence and presence of a potential lead compound, which is a peptide, and modified activity of the target is determined. Similarly, peptides can be used as dominant negative inhibitors or the validation of prospective drug targets using assays such as observing the phenotype resulting from over-expression of the peptides in ex-vivo assays or in transgenic mice. In one known approach to identify novel lead compounds, random peptide synthetic mimetic or mimotope libraries are produced using short random oligonucleotides produced by synthetic combinatorial chemistry. The DNA sequences are cloned into an appropriate vehicle for expression and the encoded peptide is then screened using one of a variety of approaches. However, the ability to isolate active peptides from random fragment libraries can be highly variable with low affinity interactions occurring between the peptide-binding partners. This is not surprising, considering that biological molecules appear to recognise shape and charge rather than primary sequence Yang and Honig J. To enhance the probability of obtaining useful bioactive peptides or proteins from random peptide libraries, peptides have previously been constrained within scaffold structures, eg. USA, 97, , or catalytically inactive staphylococcal nuclease Norman et al, Science, , , , to enhance their stability. Constraint of peptides within such structures has been shown, in some cases, to enhance the affinity of the interaction between the expressed peptides and its target, presumably by limiting the degrees of conformational freedom of the peptide, and thereby minimizing the entropic cost of binding. It is also known to tailor peptide expression libraries for identifying specific peptides involved in a particular process, eg. The amplified sequences are expressed using a bacterial display system, for screening with selected antigens to determine those antibody fragments that bind the antigens. However, the expression libraries described in U. Additionally, the antibody-encoding libraries described in U. Several attempts have been made to develop libraries based on naturally occurring proteins eg genomic expression libraries. Libraries of up to several thousand polypeptides or peptides have been prepared by gene expression systems and displayed on chemical supports or in biological systems suitable for testing biological activity. For example, genome fragments isolated from Escherichia coli MG have been expressed using phage display technology, and the expressed peptides screened to identify peptides that bind to a polyclonal anti-Rec A protein antisera Palzkill et al. Gene, , Additionally, as many bacteria comprise recA-encoding genes, the libraries described by Palzkill et al. The procedure described by Diversa Corp. Rare sequences, that are less likely to reanneal to their complementary strand in a short period of time, are isolated as single-stranded nucleic acid and used to generate a gene expression library. However, total normalization of each organism within such uncharacterized samples is difficult to achieve, thereby reducing the biodiversity of the library. Such libraries also tend to be biased toward the frequency with which a particular organism is found in the native environment. As such, the library does not represent the true population of the biodiversity found in a particular biological sample. In cases where the environmental sample includes a dominant organism, there is likely to be a significant species bias that adversely impacts on the sequence diversity of the library. Furthermore, as many of the organisms found in such samples are uncharacterized, very little information is known regarding the constitution of the genomes that comprise such libraries. Accordingly, it is not possible to estimate the true diversity of such libraries. Additionally, since the Diversa Corp. Accordingly, there remains a need to produce improved methods for constructing highly diverse and well characterized expression libraries wherein the expressed peptides are capable of assuming a secondary structure or conformation sufficient to bind to a target protein or nucleic acid, such as, for example, by virtue of the inserted nucleic acid encoding a protein domain. USA 91 , Proteins that fold well in nature have non-random hydrophobicity distributions Irback et al. USA 93, , In any native peptide, the distribution of amino acid residues according to their chemical properties eg hydrophobicity, polarity, etc is also non-random Baud and Karlin, Proc Natl Acad. USA 96, , Accordingly, the present inventors realized that random peptide libraries have a low frequency of naturally occurring or native peptide conformational structures or secondary structures, such as, for example, those structures formed by protein domains. It will be understood from the disclosure herein that the bioactive peptides or proteins expressed by individual library clones of such libraries are screened for an activity of the encoded peptide, particularly a binding activity, which said encoded protein has not been shown to possess in the context of the protein from which it was derived ie in its native environment. In the screening process, any library clone encoding a peptide that has the same activity as it would have in its native environment is excluded during the screening process, since an objective of the present invention is to isolate novel bioactive peptides or proteins. Peptides encoded by genomes which differ from the genome of the drug target organism eg. This is because in the evolution of the target organism itself, such high affinity peptide domains have been selected against other than the interaction interfaces which may exist in that organism for functional dimerization with natural partners. In one embodiment, the libraries described in the present invention are constructed from nucleic acid fragments comprising genomic DNA, cDNA, or amplified nucleic acid derived from one or two or more well-characterized genomes. Preferably, one or more well-characterized genomes is a compact genome of a eukaryote ie. In another embodiment, one or more well-characterized genomes is a compact genome of a prokaryote ie. Wherein the nucleic acid fragments are from mixtures of organisms, it is preferred that those organisms are not normally found together in nature. In accordance with this embodiment of the invention, the process of combining nucleic acid fragments derived from diverse organisms not normally found together in nature enhances and controls diversity of the expression library produced using such nucleic acid fragments. It is to be understood that the nucleic acid fragments used in the production of the expression libraries of the present invention are generated using art-recognized methods such as, for example, a method selected from the group consisting mechanical shearing, digestion with a nuclease and digestion with a restriction endonuclease. Combinations of such methods can also be used to generate the genome fragments. In a particularly preferred embodiment, copies of nucleic acid fragments from one or two or more genomes are generated using polymerase chain reaction PCR using random oligonucleotide primers. The nucleic acid fragments or cDNA or amplified DNA derived therefrom are inserted into a suitable vector or gene construct in operable connection with a suitable promoter for expression of each peptide in the diverse nucleic acid sample. The construct used for the expression of the diverse nucleic acid fragment library is determined by the system that will be used to screen for those peptides that have a conformation sufficient for binding to a target protein or nucleic acid. Thus, consideration is generally given to an expression format suitable for screening the library. In one embodiment, the vector or gene construct is suitable for in vitro display of an expressed peptide. Preferred in vitro display formats include, ribosome display, mRNA display or covalent display. In another embodiment, the vector or gene construct is suitable for expressing a peptide in a cellular host. Preferred cellular hosts in this context are capable of supporting the expression of exogenous or episomal DNA such as, for example, a cellular host selected from the group consisting of a bacterial cell, yeast cell, insect cell, mammalian cell, and plant cell. In another embodiment, the vector or gene construct is suitable for expressing a peptide in a multicellular organism. Accordingly, one aspect of the present invention provides a method of constructing an expression library for expressing a peptide having a conformation sufficient for binding to a target protein or nucleic acid, said method comprising: By way of exemplification, FIG.

Chapter 5 : Methods of constructing and screening diverse expression libraries - Phylogica Limited

Immunological screening for protein expression is also possible using plasmid expression vectors. Both vectors allow for high levels of inducible fusion protein expression; once positive clones are identified and isolated, the characterization procedure of the isolated positives will depend on the particular vector used.

A DNA library is a set of cloned fragments that collectively represent the genes of a particular organism. Particular genes can be isolated from DNA libraries, much as books can be obtained from conventional libraries. The secret is knowing where and how to look. There are two general types of gene library: The choice of the particular type of gene library depends on a number of factors, the most important being the final application of any DNA fragment derived from the library. If the ultimate aim understands the control of protein production for a particular gene or its architecture, then genomic libraries must be used. However, if the goal is the production of new or modified proteins, or the determination of tissue-specific expression of timing patterns, cDNA libraries are more appropriate. The main consideration in the construction of genomic or cDNA libraries is, therefore, the nucleic acid starting material. Since the genome of an organism is fixed, chromosomal DNA may be isolated from almost any cell type in order to prepare genomic DNA. Thus, it is important to consider carefully the cell or tissue type from which the mRNA is to be deriver in the construction of cDNA libraries. There are a variety of cloning vectors available, many based on naturally occurring molecules such as bacterial plasmids or bacteria-infecting viruses. The choice of vector also depends on whether a genomic library or cDNA library is constructed. After genomic DNA has been isolated and purified, it is digested with restriction endonucleases. These enzymes are the key to molecular cloning because of the specificity they have for particular DNA sequences. It is important to note that every copy of a given DNA molecule from a specific organism will give the same set of fragments when digested with a particular enzyme. DNA from different organisms will, in general, give different sets of fragments when treated with the same enzyme. By digesting complex genomic DNA from an organism it is possible to reproducibly divide its genome into a large number of small fragments, each approximately the size of a single gene. Some enzymes cut straight across the DNA to give flush or blunt ends. Other restriction enzymes make staggered single-strand cuts, producing short single-stranded projections at each end of the digested DNA. These ends are not only identical but complementary and will base-pair with each other; they are, therefore, known as cohesive or sticky ends. The choice of which enzyme to use depends on a number of factors. For example, the recognition sequence of 6 bp will occur, on average, every 46 bases, assuming a random sequence of each of the four bases. Enzymes with 8 bp recognition sequences produce much longer fragments. This makes subsequent steps more manageable, since a smaller number of those fragments need to be cloned and subsequently analyzed. The DNA products resulting from restriction digestion to form sticky ends may be joined to any other DNA fragments treated with the same restriction enzyme. There will, of course, also be pairing of fragments derived from the same starting DNA molecules, termed re-annealing. All these pairing are transient, owing to the weakness of hydrogen bonding between the few bases in the sticky ends, but they can be stabilized by use of an enzyme, DNA ligase, in a process termed ligation. However, long reaction times are needed to compensate for the low activity of DNA ligase in the cold. It is also possible to join blunt ends of DNA molecules, although the efficiency of this reaction is much lower than in sticky-ended ligations. In this way a DNA fragment can be cloned to provide sufficient material for further detailed analysis or for further manipulations. Thus, all of the DNA extracted from an organism and digested with a restriction enzyme will result in a collection of clones. This collection of clones is known as a gene library. For example, if the organism is a mammal whose entire genome encompasses some kbp and the gene is 10 kbp, then the gene represents only 0. It is impractical to attempt to recover such rare sequences directly from isolated nuclear DNA because of the overwhelming amount of extraneous DNA sequences. Instead, a genomic library is prepared by isolating total DNA from the organism, digesting it into fragments of suitable size, and cloning the fragments into an appropriate vector. This approach is called shotgun cloning because the strategy has no way of targeting a particular gene but instead seeks to clone all the genes of the organism at one time. The intent is that at least one recombinant clone will contain at least part of the gene of interest. This can be achieved by partial restriction digestion with an enzyme that recognizes tetra nucleotide sequences. Complete digestion with such an enzyme would produce a large number of very short fragments, but, if the enzyme is allowed to cleave only a few of its potential restriction sites before the reaction is stopped, each DNA molecule will be cut into relatively large fragments. Average fragment size will depend on the relative concentrations of DNA and restriction enzyme and, in particular, on the conditions and durations of incubation. It is also possible to produce fragments of DNA by physical shearing, although the ends of the fragments may need to be repaired to make them flush ended. This is achieved by using a modified DNA polymerase termed Klenow polymerase. The mixture of DNA fragments is then ligated with a vector, and subsequently cloned. If enough clones are produced there will be a very high chance that any particular DNA fragment, such as a gene, will be present in at least one of the clones. To keep the number of clones to a manageable size, fragments about 10 kb in length are needed for prokaryotic libraries, but the length must be increased to about 40 kb for mammalism libraries. Genomic libraries have been prepared from hundreds of different species. Many clones must be created to be confident that the genomic library contains the gene of interest. For example, if the library consists of 10 kbp fragments of the E. The need for cloning vectors capable of carrying very large DNA inserts becomes obvious from these numbers. Specific recognition and binding of other molecules is a defining characteristic of any protein or nucleic acid. Often, target ligands of a particular protein are unknown, or, in other instances, a unique ligand for a known protein may be sought in the hope of blocking the activity of the protein or otherwise perturbing its function. These strategies are also applicable to the study of nucleic acids. Unlike genomic libraries, combinatorial libraries consist of synthetic oligomers. Arrays of synthetic oligonucleotides printed as tiny dots on miniature solid supports are known as DNA chips. Specifically, combinatorial libraries contain very large numbers of chemically synthesized molecules such as peptides or oligonucleotides with randomized sequences or structures. Such libraries are designed and constructed with the hope that one molecule among a vast number will be recognized as a ligand by the protein or nucleic acid of interest. If so, perhaps that molecule will be useful in a pharmaceutical application, for instance as a drug to treat a disease involving the protein to which it binds. An example of this strategy is the preparation of a synthetic combinatorial library of hexapeptides. One approach to simplify preparation and screening possibilities for such a library is to specify the first two amino acids in the hexapeptide while the next four are randomly chosen. Screening these libraries with the protein of interest reveals which of the libraries contains a ligand with high affinity. Selection for ligand binding, again with the protein of interest, reveals the best of these 20, and this particular library is then varied systematically at the fourth position, creating 20 more libraries each containing or hexapeptides. This cycle of synthesis, screening, and selection is repeated until all six positions in the hexapeptide are optimized to create the best ligand for the protein. Thrombin is a major target for the pharmacological prevention of clot formation in coronary thrombosis. The protocol is similar for phage-based libraries except that bacteriophage plaques, not bacterial colonies, are screened. In a typical experiment, host bacteria containing either a plasmid based or bacteriophage-based library are plated out on a petri dish and allowed to grow overnight to form colonies or in the case of phage libraries, plaques Fig 4. A replica of the bacterial colonies or plaques is then obtained by overlaying the plate with a nitrocellulose disc. The disc is removed, treated with alkali to dissociate bound DNA duplexes into single-stranded DNA, dried, and placed in a sealed bag with labelled probe. The probe and target DNA complementary sequences must be in a single stranded form if they are to hybridize with one another. Probes for Southern Hybridization: Clearly, specific probes are essential reagents if the goal is to identify a particular gene against a background of innumerable DNA sequences. The oligonucleotides are synthesized so that different bases are incorporated at sites where degeneracies occur in the codons. The final preparation thus consists of a mixture of equal-length oligonucleotides whose sequences vary to accommodate the degeneracies. A piece of DNA from the corresponding gene in a related organism can also be used as a probe in screening a library for a particular gene. Such probes are termed heterologous probes because they are not derived from the homologous same organism. Problems arise if a complete eukaryotic gene is the cloning target; eukaryotic genes can be tens or even hundreds of kilo-base pairs in size. Genes of this size are fragmented in most cloning procedures. However, most cloning strategies are based on a partial digestion of the genomic DNA, a technique that generates an overlapping set of genomic fragments. This being so, DNA segments from the ends of the identified clone can now be used to probe the library for clones carrying DNA sequences that flanked the original isolate in the genome. Repeating this process ultimately yields the complete gene among a subset of overlapping clones. These libraries present an alternative strategy for gene isolation, especially eukaryotic genes. Ligation of blunt-ended DNA fragments is not as efficient as ligation of sticky ends; therefore, with cDNA molecules additional procedures are undertaken before ligation with cloning vectors. One approach is to add cDNA small, double stranded molecules with one internal site for a restriction endonuclease; these are termed nucleic acid linkers. Numerous linkers are commercially available with internal restriction for many of the most commonly used restriction enzymes. Linkers are blunt end ligated to cDNA but since they are added much in excess of the cDNA, the ligation process is reasonably successful. This process may be made easier by the addition of adaptors rather than linkers, which are identical except that the sticky ends are performed and so there is no need of restriction digestion following ligation. Once a cDNA derived from a particular gene has been identified, the cDNA becomes an effective probe for screening genomic libraries for isolation of the gene itself. Because different cell types in eukaryotic organisms express selected subsets of genes, RNA preparations from cells or tissues in which genes of interest are selectively transcribed are enriched for the desired mRNAs.

Chapter 6 : Expression cloning - Wikipedia

Screening an expression library with a ligand probe: isolation and sequence of a cDNA corresponding to a brain calmodulin-binding protein. Sikela JM, Hahn WE. The use of cloning vectors that express inserted cDNA as fusion protein has led to the isolation of genes encoding a variety of eukaryotic proteins.

Chapter 7 : Cloning and Molecular Analysis of Genes

The process is repeated until wells carryinghomogeneous clones corresponding thegene of interest have been calendrierdelascience.comSION LIBRARIES calendrierdelascience.com a DNA library is established using expressionvectors, each individual clone can beexpressed to yield a calendrierdelascience.com type of screening is important where theDNA sequence of the target.