Protein Microarrays

Issues in protein expression

I want to start this discussion with the problems associated with protein expression. In many cases if you are going to make a protein micro array you are going to need to express proteins on a genomic scale and therefore the chances of problems is much greater that working with s few proteins.

To begin with there are two different type of expression, homologous and heterologous.

Homologous expression occurs either in the same organism. For eukaryotic expression we have several systems, yeast, insects, fungi, and cultured cells. In each case an expression vector is needed to transfer the gene into the host cells and to regulate the expression of the protein. For systems outside of these it can be difficult to transform the cells. In plant systems agrobacterium is used to transfer in DNA.

Heterologous expression is when the protein of interest is expressed in a different organism such as a mammalian protein expressed in E. coli. With heterologous expression there are many possible problems. One we do not often think of is toxic proteins. Often in an expression system we overproduce a certain protein. The expressed protein may be toxic to the system. In this case you may never find a clone unless the expression of the protein is tightly regulated. Also in some cases the toxic effects are only seen at high expression and a lower copy number plasmid and less active promoter can help.

Missing interaction can also result in inclusion bodies. In this case the proteins necessary to add the cofactors, metals, etc are not present in the expression system and so the proteins do not become active properly folded proteins. In a few cases is has been shown that the processing in the Golgi apparatus is required to complete the folding process properly. If the protein is expressed in an eukaryote this will not happen. In cases where proteins are found in complexes the absence of the other subunits can result in an inactive and not properly folded protein.

The last problem is that the codons used in the gene you want to express are not commonly used in the expression system and represent rare codons. Rare codon usage usually means that the tRNA for that codon will also be rare. This results in inefficient expression of the protein. To compensate for this the codons in the gene will either need to be changed to the more common codons or the tRNA for that codon need to be expressed in the system. There are E. coli cell lines that have a plasmid containing the tRNA for the rare codons.

The most common problem with protein expression is due to the improper folding of the expressed protein. There are many factors that can result in this. The cytoplasm of E coli tends to be a reducing environment. This can result in the reduction of the disulfide bonds in the protein and the subsequent oxidation of the sulfur attaching it to

other proteins. When this occurs you get a blob of protein precipitating out in what is called an inclusion body. Another cause of inclusion body formation is that the protein is expressed so rapidly that the cells folding system does not work quickly enough. In this case coexpressing the proteins of the foldosome, GROeL and GROeS, can increase the rate of folding. Alternatively controlling the "rate" of expression can help with folding. In this case the activation of the promoter is controlled by the amount of stimulating compound added. In most cases when inclusion bodies are found there is always some of the protein properly folded in the cytoplasm but the majority is in the inclusion body. When a protein is expressed as an inclusion body greater than 90% of the expressed protein may be found in the inclusion body.

A protein expressed in an inclusion body can be recovered if it can be refolded. The difficulty with this is that partial refolding occurs and that some proteins can not be refolded due to processing or other factors.

Protein Micro-arrays

Now let's look at protein micro arrays. In almost all protein micro arrays you are going to need an antibody. There are three common protein micro arrays which have different uses. The first type is called an analytical micro array. In this system an antibody raised against a molecule or molecules of interest (protein, zenobiotic, ect) are arrayed on some surface. The solution that you wish to probe is placed on the surface for binding to occur. After washing, a solution of second antibodies, which are either linked to an enzyme or fluorescent molecule, is reacted with the array followed by an analysis by fluorescence or a chemical reaction. Currently fluorescence is used due to the prevalence of micro array readers.

The next time of array the antibodies are again arrayed on the surface but here the all of the molecules in the mixture to be probed are fluorescently labeled. This solution of labeled molecules is then applied to the array, binding allowed to occur, washed, and detected through fluorescence.

The next type of array is called functional arrays. Here recombinant or purified proteins are attached to the surface. The molecules in the probe solution are labeled and then placed on the array. It is washed and then imaged. This is a functional array since we are looking for interactions that are part of the function of the proteins. Another example of a functional array is to array proteins and use them to probe for antibody presence such as would be found in serum or for cross reactivity of the antibodies.

The last type is a reverse phase array. In this case a complex sample, such as cell lysate, serum, etc are arrayed on the surface and it is probed with a multitude of antibodies to look for a pattern that can be associated with some condition. This type of array has been used to come up with markers associated with disease states and prongnosis of the disease.

What is needed

So now that we have the types of arrays the next question is what is need to make the array. The first thing we are going to need is some antibodies. As you may have noticed out of the six arrays that I showed before five of them require at least one or two antibodies. In the other we need the proteins to be studied to be purified. Next we need a surface to attach the proteins to and some mechanism to detect the proteins.

Three general fomats are glass slides, nano wells, and nitrocellulose membranes. Nitrocellulose binds proteins non-specifically but tightly. It has been used for years for western blots and is a proven technology. The difficulty is that the samples have to be spaced fairly far apart since they tend to spread out. Glass slides are similar to the DNA micro arrays with the same advantages and disadvantages. Lastly the nano-wells require special handling equipment.

Antibody production

Since we are going to need a significant number of antibodies, how are we going to produce them? The first step is to get a purified protein, inject it into an animal (usually a rabbit or mouse) and bleed the animal at a later time for antibodies. Mice produce very little blood but can be used to make monoclonal antibodies. Rabbits can be use to make quite a bit of antibodies but are not good for industrial use. Goats on the other hand can be used continuously for antibody production on a large scale. The difficulty with this is that you are making one antibody at a time. You could inject a mixture of proteins or molecules and get multiple antibodies or monoclonal antibodies. Here the work goes into selecting the antibodies you want.

Since the above methods involve animals they are expensive and time consuming. In addition there is the fact that many people feel that this is inhumane to the animals. In vitro methods are being developed such as phage display and an expression library.

An expression library is a means of selection a diverse pool of antibodies that can later be selected for specific proteins. This is one of many techniques that are being worked on to generate quickly a diverse array of antibodies for making protein arrays. In this process you take a pool of cDNA from IgG and IgA producing cells. Then using primers specific to the complementarity determination region, this portion of the cDNA is amplified. This is the region that determines the specificity of the antibody. This pool of CDRs are then mixed with framework cDNA and allowed to bind and then overlap extension is used to generate a gene for the antibody. This is done for the heavy and light chains separately. By linking light and heavy chains together full antibody gene is generated. This is expressed in E. coli and the various antibodies tested against different molecules. In this study they took B cells from many different individuals to get the greatest diversity. The random reassembly of the variable regions results in a large number of new antibodies. By using cDNA from B cells the sequences have been checked by the cells for self antigenicity and so the generation of antihuman antibodies is much reduced.

With any antibody based system there are several limitations. The first is specificity. Unfortunately specificity cannot be predicted. Just because a region of a protein has a similar sequence that does not guarantee that there will be cross reactivity and in many cases unlike sequences do show cross reactivity. The second issue is the binding strength or dissociation constant. The sensitivity of an assay will be limited by how tightly it binds to the antigen. The lower the dissociation constant the tighter it binds and the less you can measure. There is of course a limit due to the ability to visualize a limited number of molecules. The last problem is production. We have discussed one system being developed to make random antibodies for different compounds. Another is phage display in which phages with antibody structures encoded in their genes are mutates and screened for binding to a compound. Ones that show some binding are then further mutated and screened. After several round antibodies with reasonable binding constants are found.

Genomic Protein Expression

Another aspect of the proteins microarrays is to clone large number of different proteins to look for their interactions with other proteins or as substrates for enzymes such as kinases or phosphatases. There are three major methods used to clone on a genomic scale, the gateway system, gap repair-mediated recombination and ligation independent cloning. These are all similar in many ways.

This slide shows the different methods. In each figure the blue indicates the region being cloned and the green and red indicate common 3' and 5' sequences. In "A", gap repair-mediated ligation, the amplified gene with common ends is mixed with the linearized vector having similar ends and then transformed into yeast. Gap repair occurs and the DNA is now inserted into the plasmid and ready for expression.

In ligation independent cloning the 3' primers for each direction used to clone the gene are generated without on of the bases, in this case C. This is then mixed with T4 and dCTP which results in the digestion of the product until a C occurs. The same is done with the plasmid with a complementary strand but this time dGTP is used. After the digestions the two are mixed and cloned directly into E. coli.

The gateway system of Invitrogen is uses the integration/excision of λ phage. In integration/excision of lambda phage there is a homologous recombination between the λ att P site and the host att B site. This recombination results in an integrated phase with att L and att R sites. During excision recombination between the att L and aat R sites results in regeneration of the att B and att P sites.

In the gateway system two different att b sites are added to the sequence to be inserted. This segment is then mixed with a plasmid containing two different att P sites along with the enzymes for insertion. This plasmid can then be used for sequencing and propagation of the clone. If the researcher decides to put it in an expression vector this is

done by mixing the vector with the expression vector containing two different att R sites and the enzymes. This allows for one clone to be put in different expression vectors.

Purification of Expressed Proteins

Purification on a genomic scale is going to require a rapid, simple, efficient and accurate method. The methods used today are based on the affinity of some tag to another molecule that is either tethered to a column matrix or magnetic beads.

There are a large number of different tags each with their advantages and disadvantages. Several are proteins that are added to the protein of interest and these tags bind to various columns. Protein that are used as tags include glutathione s-transferease, calmodulin binging protein, protein A, and the maltose binding protein. Usually in cloning the tag onto the protein of interest a protease site is included for removing the tag from the desired protein product. This usually leaves anywhere from 5 to 10 extra amino acids. The other types of tags are short amino acid sequences that have an affinity for another protein or molecule. The hexyl histidine tags bind to nickel chelated columns. The FLAG tag is a sequence designed to improve water solubility and to be very antigenic. In this case the protein is isolated by an antibody column. The strep tag binds to a modified avidin but not as tightly as biotin and can be removed by desthiobiotin bound to a column. The column can be regenerated 3-5 times but with diminishing results.

Expression Systems

We have discussed some of the problems with expression proteins. The systems that are available are listed here. This is not an all inclusive list but covers the major systems. The yeast and insect cell lines are eukaryotic and so allow for protein processing and modifications. Of course E. coli is the easiest to grow but for eukaryotic systems it can be the most difficult to get good expression. The yeast system has only one drawback and that is that the yeast have an unpleasant odor. The other systems require cell culturing which can be expensive.

There several things to consider when choosing an expression system. One of the most important may be cost. The other reasons are centered on modifications and processing as well as solubility and toxicity. By appropriate choice of plasmid and expression systems most proteins can be expressed. The only problem comes when there is a special processing system for the specific protein.

Attachment of Proteins

Now that we have our proteins we need to make the arrays. In making an array you have to maintain the functionality of the protein and get the most protein bound. The amount of protein bound will partially determine the sensitivity of the assay. In thinking of the density we need as many proteins linked as possible but need to have space for the proteins to not be distorted by each other and so that they do not interfere with the

function of their neighbors. If the proteins structure is altered we may not see the activity that we want. In maintaining the structure we must keep in mind that the site of action cannot be obscured by is attachment. In some cases we are not going to be able to allow the protein to dry and therefore nanowells may be required.

For surface attachment it can be random of ordered. Random attachments usually come about when the proteins are linked with chemistry that is non-specific or will react with a single amino acid but can be located anywhere on the surface. The most common covalent surfaces are going to be amines, aldehydes and expoxy derivativzed glass. In this case the protein is linked through the modification of an amino acid. In many cases a linker is not used and so the protein is in direct contact with the surface. If the protein attaches at two sites, they can cause distortion of the protein and the loss of activity. This can also cause the protein to occupy more space and reduce the density of the proteins.

Non-covalent attachment can be through nitrocellulose, gel pads or poly lysine. Nitrocellulose binds proteins fairly tightly and at random but keeps the structure sufficient for antibodies to bind. Gel pads are gels with the proteins encased in the gel which allows molecules to diffuse through them. Poly-lysine is binding the proteins by their negative charge and so will be salt dependent and work only with negatively charged molecules.

Affinity tagged surfaces will orient the proteins and depending on the tag and the binding molecule lead to weak or very tight binding. Chelated nickel would be able to bind hexyl histidine tagged proteins. Since this is a common tag for purification this is a convenient type surface. The use of antibodies to attach the proteins requires the antibodies by attached first by one of the methods above. The advantage of antibodies is that it gets the protein of interest away from the surface reducing surface effect.

Detection

How are we going to determine if something has bound to the array or something has changed the bound proteins. One is to label the molecules binding to the array and use some means of detecting the label. Fluorescence is one common and relatively sensitive method. Radioisotopes is another that is quite sensitive. In some cases we can use light to induce a covalent binding of the bound molecule.

There are a few label free methods of detecting the bound molecules. One method is to use MALDI MS to ionize the molecules attached and determine their nature by their mass spectrum. This is very useful with protein-protein interactions when the other possible interaction proteins is unknown. A second method is to use surface plasmon resonance. This method measures difference in the index of refraction which will change when a large molecule binds to another. In this case the protein array is places on a thin gold film. A polarized colaminated white light is shown o the gold film and the reflected light is measured through a polarized filter. If something is bound there will be a change in the light. This method though is a 100 times less sensitive than fluorescence. A system using carbon nanotubes and nanowires looks for changes in the conductace through these structures. Lastly are micromechanical systems which are now being developed.

Uses of Protein Micro-arrays

The limitation to the uses of protein micro-arrays is the imagination of the developer. So far they have been used to measure protein-protein, protein-DNA/RNA, protein-lipid, and protein drug interactions. They are also used to determine disease markers and indicators of prognosis of a disease. They are also used for things such as kinase/phosphatase substrate determination and antibody cross-reactivity.