RIC

ARTÍCULO DE REVISIÓN

Proteins in a DNA world: expression systems for their study

Jorge Mauricio Reyes-Ruiz,† Hugo Alberto Barrera-Saldaña*

* ULIEG-Facultad de Medicina de la UANL.

ABSTRACT

Every day, new proteins are discovered and the need to understand its function arises. Human proteins have a special interest, because to know its role in the cell may lead to the design of a cure for a disease. In order to obtain such information, we need enough protein with a high degree of purity, and in the case of the human proteins, it is almost impossible to achieve this by working on human tissues. For that reason, the use of expression systems is needed. Bacteria, yeast, animals and plants have been genetically modified to produce proteins from different species. Even "cell-free" systems have been developed for that purpose. Here, we briefly review the options with their advantages and drawback, and the purification systems and analysis that can be done to gain understanding on the function and structure of the protein of interest.

Key words. Expression systems. Functional proteomics. Protein purification. Bioreactors.

Proteínas en el mundo del DNA. Sistemas de expresión para su estudio

RESUMEN

Cada día, nuevas proteínas son descubiertas y surge la necesidad de caracterizarlas, siendo las de origen humano las que presentan un mayor interés. Conocer su función nos ayudará a entender padecimientos y diseñar una posible cura. Sin embargo, obtener suficiente cantidad de proteínas humanas en cantidad para llevar a cabo los análisis pertinentes, presenta una gran dificultad. Por tal razón, es necesario el uso de sistemas de expresión de proteínas heterólogas. Bacterias, levaduras, animales y plantas han sido modificados genéticamente para expresar proteínas de otras especies, e incluso sistemas in vitro *han sido desarrollados para producir proteínas. En este artículo se revisan brevemente las opciones con sus ventajas y desventajas, así como las estrategias de purificación y los análisis que se pueden llevar a cabo para avanzar en el conocimiento de la función y estructura de la proteína de interés.*

Palabras clave. *Sistemas de expresión. Proteómica funcional. Purificación de proteínas. Biorreactores.*

INTRODUCTION

DNA has been the center of attention for the past half century. The completion of the sequencing of the human genome has been the most important achievement of the human race in modern times and its implications will cause controversy for years to come for the ethical, social and even religious consequences that this discovery has brought. Now and then we read in newspapers or hear on the radio that a gene has been linked to a particular disease or affliction, even that some genes could be linked to criminal behavior, sexual preferences or drug-addictions. Consequently, we are gaining an understanding of the human genome, and will be able to attack these problems targeting the accused genes. But rarely in a headline is a protein the center of attention, when the proteins are actually the functional part of the great majority of the genes. Besides Mad Cow Disease, the importance of proteins is almost null in the everyday world. Fortunately, among scientists the story is different: with an actual estimate of about 30,000[1] genes and with as much as 90,000 – 120,000 pro-

† 2205 McGaugh Hall. University of California, Irving. Irvin, CA. 92697.

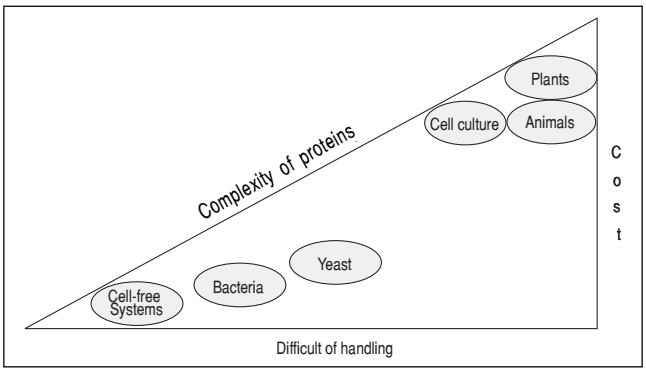

Figure 1. *Selection of the expression system. In order to make the right selection, one must consider the complexity of the protein to be expressed, the cost and the easiness of handling. Despite the low cost and easy manipulation of cell-free systems, bacteria and yeast, complex proteins will be hard to produce successfully; cell culture, on the other hand, will be able to produce complex proteins, but the cost and the need of specialized facilities make it less appealing, though necessary frequently. Animals and plants as bioreactors will produce great amounts of biologically active proteins, but the cost implied in the maintenance and other factors, such as the time needed to harvest and the cost of purification from complex tissues add up significantly.*

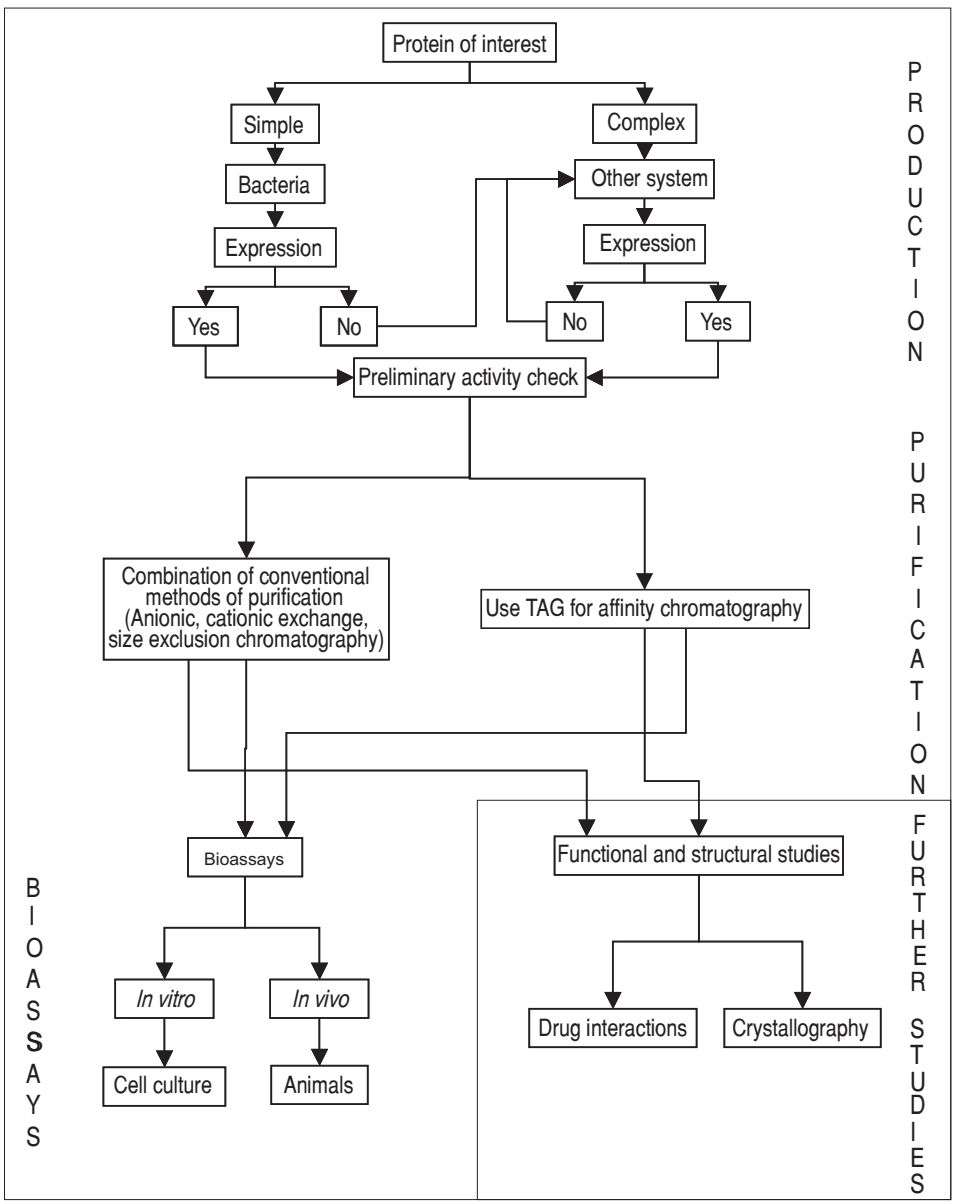

Figure 2. *Basic strategy to follow in protein studies. Choosing the right expression system can take some time and makes the difference between success and failure. Once the protein is expressed, a purification scheme can then be planned. The use of a TAG should not be overlooked. The ultimate goal will be to gain a major understanding of the function of the protein by performing bioassays and by exploring possible targets for drugs and structural composition.*

teins coded by the human genome (estimates say that because of alternative splicing, a gene on average can code for 3 or 4 proteins) and the function of 50% of these proteins remaining unknown, more studies on proteins rather than DNA are required.

Proteins being the targets of the top-selling prescription drugs, proteomics is expected to make a profit of over $5 billion in sales by the end of 2005, according to the Proteome Society.[2] Pharmaceuticals on the market target fewer than 500 human gene products[3] sug-gesting that there is an enormous untapped pool of human gene-based targets for therapeutic intervention.

Efforts to understand the relationship between protein structure and biological function have intensified, and the huge number of candidate proteins rendered by functional genomics has generated interest in all aspects of protein expression and purification. Proper expression and purification techniques are essential for the large-scale production of pure proteins, which can be used in subsequent analyses, i.e. high-throughput screening and 3-D structure determination. In addition, expression and purification systems that rapidly yield high levels of pure recombinant proteins are fundamental for the identification of target molecules for drug development. The keys to efficient, high production expression systems are good host strains, vectors, and growth conditions. Bacterial, yeast, insect and mammalian cells are the most common types of expression systems used by researchers today.[4]

Once the protein is over expressed, detection methodologies as well as functional tests are used to follow up the protein during another critical step of study: purification.

We will briefly review the most common expression systems (Figure 1) and discuss alternatives for detection, biological assays and purification of recombinant proteins (Figure 2).

EXPRESSION SYSTEMS

Bacteria

The first choice for the heterologous production of a human protein is *Escherichia coli,* because of the low cost of culture media, the availability of a great variety of strains and expression vectors, their easy handling and acceptable production yields of recombinant proteins in a short time.[5] The drawbacks of this system is that as a prokaryote organism, this bacterium is unable to complete some post-translational modifications, such as glycosylation and formation of disulfide bonds among others, required for many eukaryotic proteins to function properly. Likewise, its codon-usage differs from that of human cells,[6,7,8] which leads to low yields[9] or sometimes a truncated protein, a non-functional protein or no protein at all.[10] But now there are some improved commercially available strains to overcome some of these limitations. For example, if the protein contains disulfide bonds, proper folding is stimulated with a more oxidizing cytoplasmic environment and Novagen[11] has the AD4A4 and Origami strains mutated in the thioredoxin reductase and glutathione reductase genes to aid for this purpose. When the protein contains a high number of rare *E. coli* codons, one can try to express it using a strain that co-expresses the tRNAs for these rare codons, for example: CodonPlus-RIL, CodonPlus-RP and Rosetta are strains available from Stratagene[12] and Novagen that co-express some rare tRNAs. The regulation of expression can be constitutive or triggered by the addition of an inducer such as IPTG or L-arabinose[13], or just by a shift in temperature[14], to turn on the promoter (depending on the vector used). The time to harvest the expressed protein varies between 3 to 12 hrs, depending on the expression levels achieved, or if the protein is toxic for the host. Other bacteria have been used as expression hosts (for example, *B. subtilis,*[15] *Lactococcus, Lactobacillus, Leuconostoc, Pediococcus and Streptococcus spp*[16]) but *E. coli* remains the most widely used.

Yeast

Next in the list of the most commonly used expression hosts is an expanding group of yeasts. Yeasts are attractive hosts for production of mammalian proteins requiring posttranslational processing, because they offer advantages of both eukaryotic biosynthetic mechanisms and bacteria-like growth and handling.[17] Thus for the past years yeasts have been used to overcome the shortcomings of bacterial expression systems. They can provide intracellular as well as extracellular expression by applying short signaling sequences.[18] *Saccharomyces cerevisiae* is usually the first yeast of choice and since it tends to hyperglycosylate secreted proteins, it is primarily used for intracellular production.[19] Other yeast strains (i.e., *Pichia pastoris, Hansenula polymorpha, Schizosaccharomyces pompe, Klyveromyces lactis*, and *Yarrowia lipolytica*) are more suited for secretion.[19,20] Among the best-developed systems for large-scale protein production are those based on *P. pastoris*.[21] Notable advantages of *P. pastoris* compared with *S. ce-*

revisiae are that the former may produce more abundant heterologous protein 10-100-fold and it does not hyperglycosylate them as the latter does upon secretion.[21] Both yeasts have a majority of N-linked glycosylation of the high-mannose type, but in *Pichia* the average length of the oligosaccharide chain is 8-14 mannose residues per side chain, whereas in *Saccharomyces* it is between 50-150 mannoses. In addition, core oligosaccharides in *S. cerevisiae* have terminal α1,3 glycan linkages (believed to be responsible for the antigenic nature of proteins produced by secretion from *S. cerevisiae)*, while *P. pastoris* has none.[22]

One of the most important drawbacks of yeast is the cell wall if the protein is not to be secreted. Breaking the cell wall and recovering the interior of the cell intact can be a major headache. The typical way to do it is by using glass beads, but this has limitations on the size of the sample. For a larger sample, a French Press capable of reaching 24,000 psi or a Bead Mill are better options.

Mammalian cultured cells

Mammalian cells in culture are the next choice. By using a cell line from higher eukaryotes one can usually avoid most of the problems encountered in the bacteria or yeast systems, but at a steep price: the cost is sky-high when compared to bacteria or yeast cultures, handling requires trained personnel and appropiate facilities, there are safety risks and the yield is not necessarily high. Nevertheless, mammalian host cell lines, although non-human, are the protein production vehicles of choice as they are capable of producing proteins with almost human-like glycosylation. Glycosylation is of particular importance for the recognition of "foreign" proteins by the immune system of higher mammals.[23] Pharmacokinetic analysis has shown that insufficient or inconsistent glycosylation results in accelerated clearance of these proteins by carbohydrate specific receptors in the human liver preventing them from fulfil-ling their therapeutic purpose.[24]

Insect cells

Insect cells are valued for their ability to express some recombinant proteins at a higher level than most mammalian expression systems.[25] The baculovirus-based system is a eukaryotic expression system and is therefore able to produce overexpressed recombinant proteins with proper folding, disulfide bond formation and oligomerization.[26] This system is also capable of performing many post-translational modifications including O-linked and N-linked glycosylation, phosphorylation, acylation, amidation, carboxymethylation and cleavage of certain proteins to their active forms.[27] The most frequently used insect cells that are susceptible to baculovirus infections are Sf9 and Sf21 cell lines. Both of these lines are originally established from ovarian tissues of *Spodoptera frugiperda* larvae. These cell lines may be grown in suspension and can therefore be used in a bioreactor.[26] High Five cells, derived from *Trichoplusia ni* egg cell homogenates, have been shown to be capable of expressing significantly higher levels of secreted recombinant proteins compared to other insect cells.[28] They can be transformed by baculovirus or by using liposomes.

In contrast to the baculovirus system, the *Drosophila* system relies on stable cell lines that express the foreign protein.[29] Since the *Drosophila* cells will spontaneously incorporate hundreds of copies of the transfected genetic information, the expression levels should be very high, and no extensive testing of transfected clones for expression is required. Stable cell lines can be established after co-transfection with a plasmid that confers resistance to the antibiotic hygromycin B.[30] The expression of the foreign protein is under the control of the metallothionein promoter, and the system can be induced by the addition of non-toxic concentrations of metal ions such as copper.[30]

OTHER INTERESTING ALTERNATIVES

Cell-free systems

Cell-free protein expression removes the need for cell culture: protein is expressed *in vitro* using cellular extracts that contain all the machinery and biochemical constituents required for transcription and translation. The cellular extract may be obtained either from prokaryotic or eukaryotic cells.[31] The use of *in vitro* translation systems can have advantages over *in vivo* gene expression when the over-expressed product is toxic to the host cell, when the product is insoluble or forms inclusion bodies, or when the protein undergoes rapid proteolytic degradation by intracellular proteases. In principle, it should be possible to prepare a cell-free extract for *in vitro* translation of mRNAs from any type of cell. In practice, only a few cell-free systems have been developed for *in vitro* protein synthesis. The most frequently used cell-free translation systems consist of extracts from rabbit reticulocytes, wheat germ and

Escherichia coli.[32,33] All are prepared as crude extracts containing all the macromolecular components (70S or 80S ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation and termination factors, etc.) required for translation of exogenous RNA. To ensure efficient translation, each extract must be sup-plemented with amino acids, energy sources (ATP, GTP), energy regenerating systems (creatine phosphate and creatine phosphokinase for eukaryotic systems, and phosphoenol pyruvate and pyruvate kinase for the *E. coli* lysate), and other co-factors (Mg_2^+, K^+, etc.). There are two approaches to *in vitro* protein synthesis based on the starting genetic material: RNA or DNA. Standard translation systems, such as reticulocyte lysates and wheat germ extracts, use mRNA as a template; whereas "coupled" and "linked" systems start with cDNA templates.[34] *Escherichia coli*-coupled transcription–translation system yields as much as 6 mg of protein per milliliter of reaction volume.[35]

Plants

The use of plants as expression systems for recombinant proteins is currently receiving a lot of attention all over the world. Plant expression systems have advantages over other *in vitro* expression systems in terms of low production costs. In addition, contamination of endotoxin and animal viruses has been associated with bacterial and eukaryotic cultured cell expression systems for recombinant proteins. Although transgenic plants may contain plant-specific viruses or bacteria, they are not known to be harmful to humans and animals. Thus, using the plant expression systems, several biologically active proteins and peptides for potential pharmaceutical applications, such as vaccine antigens and immunomodulators have been expressed in plants.[36,37,38] However, the production of recombinant proteins in higher plants has drawbacks, which can make this system less attractive from an economic standpoint. The first drawback is the substantial length of time required from the initial transformation event to small-scale evaluation and production: often 2 years or more. This length is primarily due to the relatively slow growth rates of terrestrial plants relative to other organisms used for protein expression. Hence, the initial generation of transformants, their propagation to flowering, genetic crosses, and production of seed stocks are handicapped by this intrinsically slow rate of growth. Higher plant expression systems are also associated with complex processing issues, because recombinant proteins are produced and deposited in specific organs such as leaves, fruits, and seeds. These proteins must be purified to homogeneity out of a complex mixture of tissues and cell types, a requirement that can add significantly to the costs of purification.

Animals as bioreactors

Expression of recombinant human proteins in the milk of transgenic dairy animals offers a source of clinically important proteins that cannot be produced as efficiently in adequate quantities by other methods. Milk expression has been reported for at least 17 different proteins in five livestock species, 11 of them at commercially feasible levels of >1 g/l.[39] Concomitant advances have been made in purifying proteins from raw milk. Heterologous proteins have been expressed in the obvious dairy species: cows, sheep and goats. However, pigs and rabbits are also used for selected applications because of their large litters and shorter generation times. The size of the production herd required for a particular application will depend primarily upon total annual needed, the recombinant-protein expression level and the recovery efficiency.

Another approach is the use of the hen, which prom-ises to be a low cost, high-yield bioreactor. More than half of the egg white protein content derives from the ovalbumin gene with four other proteins (lysozyme, ovomucoid, ovomucin and conalbumin) present at levels of 50 milligrams or greater.[40] Taking advantage of the promoter of this gene to express a recombinant protein could bring yields of up to a gram or more. Since modern layers produce 300 eggs per year, three or four hens producing one gram per egg would yield a kilogram of raw product annually.[41] The naturally sterile egg[42] also allows a long shelf life of recombinant protein without loss in activity.[43] In spite of these advantages, transgenic procedures for the bird have lagged far behind those of other organisms. Gene transfer into the avian genome has been achieved by the use of retroviruses,[44] microinjection of DNA into cytomplasm of fertilized zygotes,[45]and the use of Drosophila transposons.[46] However, the production levels have been only 3 to 38 μg of protein per egg.[43] Hence, development of more robust and efficient transgenic methods will be necessary before at-will modification of avian genetic material is feasible.

DETECTION

Once the protein is expressed, one must find a way to detect it throughout the procedure. Western

blot or immunodetection is the number one choice. The protein is separated in an SDS-PAGE electrophoretically transferred to a membrane and then exposed to antibodies. The first antibody detects the protein of interest, while a second is conjugated with either alkaline phosphatase, or horseradish peroxidase and detects the immunoglobulin used as first antibody. Using the appropriate substrates for each enzyme, a color-devel-oping reaction takes place in the membrane where the proteins were transferred and one is able to detect it. Another option for greater sensibility is the use of chemiluminiscence (CL), a technique based upon production of light in the form of enhanced CL catalyzed by an enzyme. The products are analyzed in CL by using luminol and enhancer as substrates for peroxidase conjugates and 1,2-dioxetane and enhancer for alkaline phosphatase conjugates.

The easiest way to detect the protein being investigated is to use the appropriate antibody usually avail-able in the market; however, in most of the cases of newly investigated proteins, the antibodies that detect them are not available. One can produce the antibody, but requires that the research group has experience and specialized facilities. It is time- and mon-ey- consuming, making the use of molecular tags the best option. More and more often we find the use of tags to be extremely helpful. By adding a small sequence coding for a polypeptide (in most cases it does not interfere with the biological activity of the protein) whose antibody is available on the market, one can save both time and money. Some of them are: c-myc - a 10 amino acid segment of the human protooncogene myc (EQKLISEEDL); V5 epitope is derived from a small epitope (Pk) present on the P and V proteins of the paramyxovirus simian virus 5 (GKPIPNPLLGLDST); HA – derived from the haemagglutinin protein from human influenza virus (YPYDVPDYA); FLAG - Synthetic FLAG peptide (DYKDDDDKC); His6 - if six histidines are placed in a row, they form a structure that binds the element Nickel. The latter is especially useful for affinity chromatography (see below) but can also be used as an epitope tag. The availability and low cost of antibodies and affinity col-umns against these tags facilitates this application.

PURIFICATION

Once we have expressed the protein at satisfactory levels and are able to detect it from the raw sample, one must develop a purification strategy. This is true specially if the purpose of our research is to uncover physical characteristics related to function, or simply because we need it in a purer form (to be injected into experimental animals). It is a prerequisite to know as much as possible of its physical properties (such as isoelectric point). Then one can start to plan which of the available chromatography protocols to use. First, some differential centrifugations or precipitations may be needed to enrich a fraction and to put the sample in an appropriate state to interact with the solid phase (resins), i.e. resuspend the proteins in a suitable buffer. It is usually very important to work fast, using protease inhibitors and proceeding at 4 °C to avoid unwanted degradation of the sample. For protein separation, high performance liquid chromatography (HPLC) offers the best resolution, and may be very useful to separate proteins from a complex mixture. However, preparative HPLC is not the first choice, mainly because of limitations in the size of the sample that can be injected, not to mention the high price of the preparative columns. Moreover they are easily clogged (leaving them useless), and the pumps required for the preparative procedures are very expensive. For the above reasons, fast protein liquid chromatography (FPLC) remains the most used meth-od in preparative chromatography, despite its lower resolution.

Again, the use of tags is a very powerful tool in the purification procedure. The main commercially available affinity resins have high quality standards and by simply modifying the expression vectors one can add a tag suitable for purification. The most common are: Histidine tags –6-10 histidines. GST – Glutathione affinity chromatography. MBP – Maltose affinity chromatography. Protein A –IgG affinity chromatography. Calmodulin binding peptide (CBP, 4kDa)–Calmodulin affinity chromatography. TAP – Combination protein A and calmodulin binding peptide. IMPACT –Chitin binding domain. FLAG - Synthetic FLAG peptide. Although the tags hardly ever interfere with the biological activity of the protein, if the protein is to be expressed for structural study purposes, it may be necessary to remove the tag after the purification procedures. Removal of the tag from a protein of interest can be accomplished with a site-specific protease, and cleavage should not reduce protein activity.[47] It is important to be sure that the protein of interest lacks the recognition site of the protease selected. The most commonly used proteases are: enterokinase, tobacco etch virus (TEV), factor Xa, and thrombin. Enterokinase is often the choice for N-terminal fusions, since it specifically recognizes a five- amino-acid polypeptide (D-

D-D-D-K-X_1) and cleaves at the carboxyl site of lysine. TEV protease is a site-specific protease that has a seven-amino-acid recognition site: E-X-X-Y-X-Q-S, (being E-N-L-Y-F-Q-S the optimal sequence)[48,49] and X can be various amino acid residues, but not all, and cleavage occurs between the conserved glutamine and serine.[50] Factor Xa cleaves at the carboxyl side of the four-amino-acid peptide I-E[D]-G-R- X_1,[51] where X_1 can be any amino acid except arginine and proline, making it a useful tool to completely remove N-terminal affinity tags. Thrombin is a protease widely used to cleave tags, but in contrast to enterokinase and factor Xa, thrombin cleavage results in the retention of two amino acids on the C-terminal side of the cleavage point. The optimal cleavage site for thrombin has the structures of X_4-X_3-P-R[K]-X_1'-X_2', where X_4 and X_3 are hydrophobic amino acids and X_1', X_2' are non-acidic amino acids.[52,53] Elimination of the tag without using a protease has also been used by introducing a self-splicing intein.[54]

FUNCTIONAL STUDIES

Since the protein of our interest is not being expressed in native conditions (after all, we did choose a heterologous expression system, right?), one must be sure that the protein being expressed is biological-ly active and the tests are very different from one protein to another. The protein assays are quantitative methods to determine the amount of a given activity. Some of them are: enzymatic reactions, when one is able to quantify a specific product after a reaction involving the protein of interest, the use of a radioactive-labeled substrate that changes its physical properties after interacting with our protein, measuring the ability to stimulate cell death, cell proliferation or other effect on development *in vivo*, or the retention in a DNA column, suggesting that the protein interacts with nucleic acids. Whichever assay is chosen, one must be aware that it should be experimentally convenient, meaning that it is reason-ably sensitive and easy to perform, and the assay employed must be specific for the activity of interest.

FURTHER STUDIES

The ultimate goal of studies of a protein is gaining understanding of its structure, and being able to correlate it with, for example, possible interactions with drugs. Several techniques are available to aid in this purpose, with mass spectrometry and 2-D electrophoresis being the most used ones. Mass spectroscopy is the method of choice for protein identification and for the characterization of post-translational modifications. The two preferred methods for the ionization of proteins and peptides are electrospray ionization (ESI) treatment and matrix-assisted laser desorption ionization (MALDI) due to their effective application to a wide range of proteins and peptides. In addition, in a more elaborated and time-consum-ing process called tandem mass spectroscopy (MS/MS), the protein to be analyzed is digested with enzymes into peptide fragments and subjected to further fragmentation and mass spectroscopy. [55, 56, 57] Obtaining a 3-D model of the protein is the final step to the journey into the understanding of the structure of a protein. The most common experimental method for obtaining a detailed picture of a protein or protein complex is to interpret the diffraction of X rays. Determining a protein's structure by X-ray crystallography consists of growing high-quality crystals of the purified protein, measuring the directions and intensities of X-ray beams diffracted from the crystals, and using computers to transform the X-ray measurements. This method produces an image of the crystal's contents that must be interpreted, which involves computer graphics to display the electron density of atoms in the molecule and the construction of a consistent molecular model.[58, 59]

PERSPECTIVES

Being that proteins are the main target for drug discovery, an important effort is dedicated to understanding their function. The field of proteomics has advanced considerably and its common applications are: target identification and validation (identifying proteins whose expression levels or activities change in disease states), identification of biomarkers from biological fluids (used to assess whether target modulation has occurred), and research of mechanisms of drug action or toxicity (screen compounds in preclinical studies for target organ toxicities as well as later on in development during clinical trials). But the practice of proteomics ranges from the identification of thousands of proteins in a particular model system, to the detailed analysis of their 3D structure, possible modifications/isoforms, and the function of a single protein. Hence producing proteins in high amounts and purity can aid in the study of function and structure of novel proteins or help discover un-known properties of long-known ones that ultimately will be fundamental in the development of highly specific drugs.

ACKNOWLENGMENTS

The authors thank Drs. William Dubinsky and Illarian V. Turko for their critical reading of the manuscript.

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921.
2. The Proteome Society www.proteome.org
3. Ward SJ. Impact of genomics in drug discovery. *Biotechniques* 2001; 31: 626.
4. Rai M, Padh H. Expression systems for production of heterologous proteins. *Curr Sci* 2001; 80: 1121-8.
5. Panda AK. Bioprocessing of therapeutic proteins from the inclusion bodies of *Escherichia coli*. *Adv Biochem Eng Biotechnol* 2003; 85: 43-93.
6. Kurland CG. Codon bias and gene expression. *FEBS Lett* 1991; 285: 165-9.
7. Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol* 1996; 260: 649-63.
8. Humphreys DP, Sehdev M, Chapman AP, Ganesh R, Smith BJ, King LM, Glover DJ, Reeks DG, Stephens PE. High-level periplasmic expression in *Escherichia coli* using a eukaryotic signal peptide: importance of codon usage at the 5' end of the coding sequence. *Protein Expr Purif* 2000; 20: 252-64.
9. Del Tito BJ Jr, Ward JM, Hodgson J, Gershater CJ, Edwards H, Wysocki LA, Watson FA, Sathe G, Kane JF. Effects of a minor isoleucyl tRNA on heterologous protein translation in *Escherichia coli*. *J Bacteriol* 1995; 177: 7086-91.
10. McNulty DE, Claffee BA, Huddleston MJ, Kane JF. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr Purif* 2003; 27: 365-74.
11. Novagen www.novagen.com
12. Stratagene www.stratagene.com
13. Newman JR, Fuqua C. Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator. *Gene* 1999; 227: 197-203.
14. Kuczynska-Wisnik D, Laskowska E, Taylor A. Transcription of the ibpB heat-shock gene is under control of sigma(32)- and sigma(54)-promoters, a third regulon of heat-shock re-sponse. *Biochem Biophys Res Commun* 2001; 284: 57-64.
15. Schallmey M, Singh A, Ward OP. Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* 2004; 50: 1-17.
16. de Vos WM. Gene expression systems for lactic acid bacteria. *Curr Opin Biotechnol* 1999; 2: 289-95.
17. Cereghino GP, Cregg JM. Applications of yeast in biotechnology: protein production and genetic analysis. *Curr Opin Biotechnol* 1999; 10: 422-7.
18. Brake AJ. Alpha-factor leader-directed secretion of heterologous proteins from yeast. *Methods Enzymol* 1990; 185: 408-21.
19. Buckholz RG, Gleeson MA. Foreign gene expression in yeast: a review. *Bio-Technology* 1991; 9: 1067-72.
20. Romanos MA, Scorer CA, Clare JJ. Recombinant DNA technology and heterologous gene expression in yeasts. *Yeast* 1991; 8: 423-88.
21. Cereghino GP, Cereghino JL, Ilgen C, Cregg JM. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr Opin Biotechnol* 2002; 13: 329-32.
22. Gemmill TR, Trimble RB. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim Biophys Acta* 1999; 1426: 227-37.
23. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. Roles for Glycosylation in the Immune system. *Science* 2001; 291: 2370-6.
24. Helenius A, Aebi M. Intracellular functions of N-linked glycans. *Science* 2001; 291: 2364-9.
25. Davis TR, Wickham TJ, McKenna KA, Granados RR, Shuler ML, Wood HA. Comparative recombinant protein production of eight insect cell lines. *In vitro Cell Dev Biol* 1993; 29A: 388-90.
26. Ikonomou L, Schneider YJ, Agathos SN. Insect cell culture for industrial production of recombinant proteins. *Appl Microbiol Biotechnol* 2003; 62: 1-20.
27. Joosten CE, Park TH, Shuler ML. Effect of silkworm hemolymph on N-linked glycosylation in two *Trichoplusia ni* insect cell lines. *Biotechnol Bioeng* 2003; 83: 695-705.
28. Invitrogen www.invitrogen.com
29. Benting J, Lecat S, Zacchetti D, Simons K. Protein expression in *Drosophila Schneider* cells. *Anal Biochem* 2000; 278: 59-68.
30. Park JH, Kim HY, Han KH, Chung IS. Optimization of transfection conditions for expression of green fluorescent protein in *Drosophila melanogaster* S2 cells. *Enzyme Microb Technol* 1999; 25: 558-63.
31. Smutzer G. Cell-free transcription and translation. *Scientist* 2001; 15: 22.
32. Spirin AS, Baranov VI, Ryabova LA, Ovodov SY, Alakhov YB. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* 1988; 242: 1162-4.
33. Sawasaki T, Hasegawa Y, Tsuchimochi M, Kasahara Y, Endo Y. Construction of an efficient expression vector for coupled transcription/translation in a wheat germ cell-free system. *Nucleic Acids Symp Ser* 2000; 44: 9-10.
34. Roche www.roche-applied-science.com
35. Betton JM. Rapid translation system (RTS): a promising alternative for recombinant protein production. *Curr Protein Pept Sci* 2003; 4: 73-80.
36. Yusibov V, Hooper DC, Spitsin SV, Fleysh N, Kean RB, Mikheeva T, Deka D, Karasev A, Cox S, Randall J, Koprowski H. Expression in plants and immunogenicity of plant virus-based experimental rabies vaccine. *Vaccine* 2002; 20: 3155-64.
37. Samyn-Petit B, Wajda Dubos JP, Chirat F, Coddeville B, Demaizieres G, Farrer S, Slomianny MC, Theisen M, Delannoy P. Comparative analysis of the site-specific N-glycosylation of human lactoferrin produced in maize and tobacco plants. *Eur J Biochem* 2003; 270: 3235-42.
38. Ohya K, Matsumura T, Ohashi K, Onuma M, Sugimoto C. Expression of two subtypes of human IFN-alpha in transgenic potato plants. *J Interferon Cytokine Res* 2001; 8: 595-602.
39. van Berkel PH, Welling MM, Geerts M, van Veen HA, Ravensbergen B, Salaheddine M, Pauwels EK, Pieper F, Nuijens JH, Nibbering PH. Large scale production of recombinant human lactoferrin in the milk of transgenic cows. *Nat Biotechnol* 2002; 5: 484-7.
40. Gilbert AB. Egg albumen and its formation. Physiology and Biochemistry of the Domestic Fowl (Bell, D.J., Free, B.M. eds), 1984; pp. 1291-1329, Academic Press.
41. Ivarie R. Avian transgenesis: progress towards the promise. *Trends Biotechnol* 2003; 21: 14-9.
42. Tranter HS, Board RB. The antimicrobial defense of avian eggs: biological perspective and chemical basis. *J Appl Biochem* 1982; 4: 295-338.
43. Harvey AJ, Speksnijder G, Baugh LR, Morris JA, Ivarie R. Expression of exogenous protein in the egg white of transgenic chickens. *Nat Biotechnol* 2002; 20: 396-9.

44. Harvey AJ, Speksnijder G, Baugh LR, Morris JA, Ivarie R. Consistent production of transgenic chickens using replication-deficient retroviral vectors and high-throughput screening procedures. *Poult Sci* 2002; 81: 202-12.
45. Love J, Gribbin C, Mather C, Sang H. Transgenic birds by DNA microinjection. *Biotechnology* 1994; 12: 60-3.
46. Lampe DJ, et al. Factors affecting transposition of the Himar1 mariner transposon in vitro. *Genetics* 1998; 149: 179-87.
47. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 2003; 60: 523-33.
48. Carrington JC, Dougherty WG. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc Natl Acad Sci USA* 1988; 85: 3391-5.
49. Dougherty WG, Carrington JC, Cary SM, Parks TD. Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J* 1988; 7: 1281.7.
50. Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ. Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* 1989; 172: 302-10.
51. Nagai K, Thogersen HC. Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*. *Nature* 1984; 309: 810-2.
52. Chang JY, Alkan SS, Hilschmann N, Braun DG. Thrombin specificity. Selective cleavage of antibody light chains at the joints of variable with joining regions and joining with constant regions. *Eur J Biochem* 1985; 151: 225-30.
53. Haun RS, Moss J. Ligation-independent cloning of glutathione fusion genes for expression in *Escherichia coli*. *Gene* 1992; 112: 37-43.
54. Xu M-Q, Paulus H, Chong S. Fusions to self-splicing inteins for protein purification. *Methods Enzymol* 2000; 236: 376-418.
55. Gevaert K, Vandekerckhove J. Protein identification methods in proteomics. *Electrophoresis* 2000, 21: 1145-54.
56. Chalmers MJ, Gaskell SJ. Advances in mass spectrometry for proteome analysis. *Curr Opin Biotechnol* 2000; 11: 384-90.
57. Yates 3rd JR. Mass spectroscopy. From genomics to proteomics. *Trends Genet* 2000; 16: 5-8.
58. McRee DE. Practical protein crystallography. San Diego: Academic Press, c1993.
59. Stout GH, Jensen LH. X-ray structure determination: a practical guide. New York: Wiley, c1989.

Reimpresos:

Dr. Hugo A. Barrera Saldaña
Facultad de Medicina, Departamento de Bioquímica
Universidad Autónoma de Nuevo León
Av. Madero y Dr. E. Aguirre Pequeño s/n
Col. Mitras Centro,
64460 Monterrey, N.L.
Tel.: (81) 8329 41 73 y 74. Fax (81) 8333 77 47
Correo electrónico: hbarrera@fm.uanl.mx

Recibido el 31 de agosto de 2004.
Aceptado el 11 de abril de 2005.