So you Need a Protein - A Guide to the Production of Recombinant Proteins

Ramamoorthi Jayaraj[1] and Peter M. Smooker[*,2]

[1]*Skin Pathogens Research Laboratory, Menzies School of Health Research, Casuarina, NT, Australia*

[2]*School of Applied Sciences, RMIT University, Bundoora, Australia*

Abstract: The field of biotechnology owes a great deal to the ability to produce recombinant proteins, which can be made in far greater abundance than many native proteins, and are more easily quality controlled. There is a great need for individual proteins to be produced for research purposes. This review is aimed at researchers who are not experienced at protein expression, but find that they have a need to produce a recombinant protein. We detail the major expression systems that will be commonly used in the laboratory situation- bacterial, yeast and insect cell culture. The application of each, and the relative advantages/disadvantages are discussed.

INTRODUCTION

Recombinant proteins can be used for diagnostic tools, vaccines, therapeutic proteins or functional enzymes. Well over 150 human and veterinary based biopharmaceuticals have received approval and the majority of these approved represent recombinant proteins [1] with many being recombinant antibodies. While the latter class have limited use in the veterinary sector, non-antibody recombinant proteins have a large role to play.

In the veterinary sector, recombinant protein vaccines are being tested extensively and used to control infectious diseases, but are also used to modulate hormones to aid in livestock gender selection [2]. "Home-grown" diagnostic kits utilising recombinant proteins are often used for research purposes to assess immune responses to vaccination or to monitor infection.

The main driver for the use of recombinant proteins in the veterinary sector is the economic loss due to a variety of viral, bacterial and parasitological infections of livestock. This has been the impetus for research into methods to both diagnose and prevent these infections. Veterinary diagnostics has traditionally relied on empirical observation of symptoms and/or faecal matter, however in order to detect disease earlier and therefore at lower pathogen levels more sensitive methods are required. The field of diagnosis has been revolutionised by the advent of the Polymerase Chain Reaction (PCR), however not all pathogens are amenable to such analysis. Sensitive serological methods for detecting immune responses induced by pathogens, or the pathogens themselves, often require the production of recombinant protein. Similarly, much of the research conducted into new veterinary vaccines also involves the production of recombinant proteins. Although vaccination has been a major success in both the medical and veterinary fields, there are obviously numerous infectious diseases that cannot be prevented using traditional attenuated or killed vaccines. Recombinant proteins can play a role in the identification of suitable antigens, using a screening methodology, and thus protein expression and purification is a key process in the systematic and comprehensive analysis of vaccine candidates [3]. Finally, recombinant proteins are a preferred source of antigen to use in a vaccine, as the quality can be controlled, and the availability is usually higher than it is with native protein.

A major consideration in the development of recombinant proteins is the choice of expression system, i.e. where is the protein to be expressed? The three major main systems for the expression of recombinant proteins are *Escherichia coli*, yeast and baculovirus mediated protein expression, although several other systems are available, including other bacteria (e.g. *Lactococcus lactis*), and mammalian cell expression. Which one is chosen depends primarily on the nature of the heterologous protein to be expressed. In this review, we consider these three most commonly used expression systems and compare their merits and weaknesses. For illustrative purposes we will mainly consider the expression of proteases, as these present particular challenges depending on the application required. This review is not intended as a step-by-step guide to protein production, as these will be found with manufacturer's literature of the system that is employed.

The Methodology

The methodology for all expression systems is fundamentally similar. The basic requirements are a coding sequence for the heterologous protein, a vector (usually a plasmid) into which the coding sequence is cloned behind a promoter that is active in the chosen host, and a suitable host that will then express the heterologous protein. Irrespective of the final host (and of course the encoding sequences), the cloning of the coding sequence into plasmid and subsequent amplification is performed in *E. coli*.

Part of the cloning strategy is that the coding sequence to be expressed is cloned in- frame with a second peptide or protein, such that the proteins are expressed as a contiguous, single unit. These entities are termed affinity tags, which are

*Address correspondence to this author at the School of Applied Sciences, RMIT University, Australia; Tel: (613) 9925 7129; Fax: (613) 99257110; E-mail: peter.smooker@rmit.edu.au

extremely valuable tools for the facilitation of protein purification and allow the purification of any tagged recombinant protein, irrespective of the intrinsic biochemical properties of the protein [4]. These affinity tags will bind to a partner ligand, usually immobilised on a resin, such that the protein containing the tag is isolated from all other (host) proteins that do not bind. A good example is a simple run of six histidine residues, which will co-ordinate divalent cations. Here immobilised metal affinity chromatography (IMAC) is used to purify a His-tagged protein. There are a variety of different tags that are used, from the small His-tag to larger tags such as glutathione *S*- transferase (GST, approx 26 kDa) which can have a marked effect on the solubility of the target protein [5]. Generally, affinity chromatography will result in recombinant protein of acceptable purity. If higher purity is required, high performance liquid chromatography (HPLC) or size exclusion chromatography (SEC) can be used to improve the purity of the expressed protein [6, 7].

All commonly used expression systems utilise expression vectors that encode a tag, such that the coding sequence to be expressed is cloned in-frame with the tag. Usually the initiation codon is supplied by the vector sequence, the tag-encoding sequences follow, and then the sequences encoding the heterologous protein are inserted at a cloning site. Often there is an encoded protease cleavage site just before the cloning site- this allows the removal of the vector-encoded sequences (ie: the tag) from the purified protein. Expression is driven from a promoter located 5' to the initiation codon. Purification procedures, once a lysate is made from cultured cells, are basically the same irrespective of the expression system. A schematic representation of the relevant vector sequences, and the cloning site for the sequences encoding the protein, is shown in Fig. (**1**). Resultant cleavage of the protein to release the N-terminal regions is also shown.

E. coli Expression Systems

The first choice of an expression system for the production of recombinant proteins for many investigators is *E. coli* [8]. The *E. coli* expression system has many advantages compared with other expression systems such as easy growth conditions, rapid biomass accumulation, and a simple scale-up process [9]. This prokaryotic organism is often used for the industrial production of therapeutic or commercial based proteins [9]. It is the system of choice for the expression of relatively small, soluble, un-modified proteins, and produces over half of the recombinant proteins produced by pharmaceutical companies.

The scale of expression can vary from several millilitres for pilot studies up to thousands of litres in industrial fermentations. Typically, once cloning of insert into expression vector has been performed, pilot (small scale purification) is used as an screen of expressing clones [3]. Clones expressing proteins can be identified by SDS-PAGE and Western blotting with antisera, either directed at the recombinant itself, or the tag (typically a monoclonal antibody that recognises the His-tag or another tag). An expressing clone is identified and used in 50-250 mL culture volumes for further optimisation, refinement of expression conditions and optimisation of purification [8].

In large scale expression, two litres of bacterial culture (Luria Bertani broth+ kanamycin + ampicillin) will yield approximately 50 to 80 g of wet weight of cells. If the protein aggregates into inclusion bodies, approximately 100 to 300 mg of recombinant protein is available (2% to 5% of total proteins) [7, 8, 10]. If the protein is soluble (usually desirable) yields will be somewhat less. A higher level of protein expression can be attained by use of enriched media (e.g. terrific broth), which can reduce volumetric requirements by increasing cell densities [3, 11].

However, there are limitations to bacterial expression. It has been estimated that only 50% of bacterial proteins and 10% of eukaryotic proteins can be expressed in *E. coli* in a soluble form [12, 13]. This shows that there are many situations when *E. coli* may not be the appropriate system. Furthermore, the percentage of expressing soluble proteins decreases when the protein molecular weight is above 60 kDa. The co-expression of molecular chaperones or heat shock proteins with target recombinant proteins can enhance the solubility and can help to attain native (i.e. as in the non-recombinant) conformation [10, 14].

A particular problem in expressing proteins in the *E. coli* cytoplasm is the formation of disulfide bonds, which does not occur in the reducing environment of the bacterial cytoplasm. This results in the mis-folding of such proteins. This can be alleviated by targeting the expressed proteins to the periplasm, but often it is more convenient to use a eukaryotic host.

Another limitation is the inability of *E. coli* to glycosylate proteins. While perhaps not a problem for most veterinary applications, many therapeutic proteins are glycosylated. With the discovery of N-linked glycosylation in some bacterial species*,* much effort is being directed into the creation of *E. coli* strains that can add glycans to protein [15]. Also, for use such as vaccination or therapeutic protein, the bioactivity (if applicable) and the amount of residual LPS present in the preparation should be assayed.
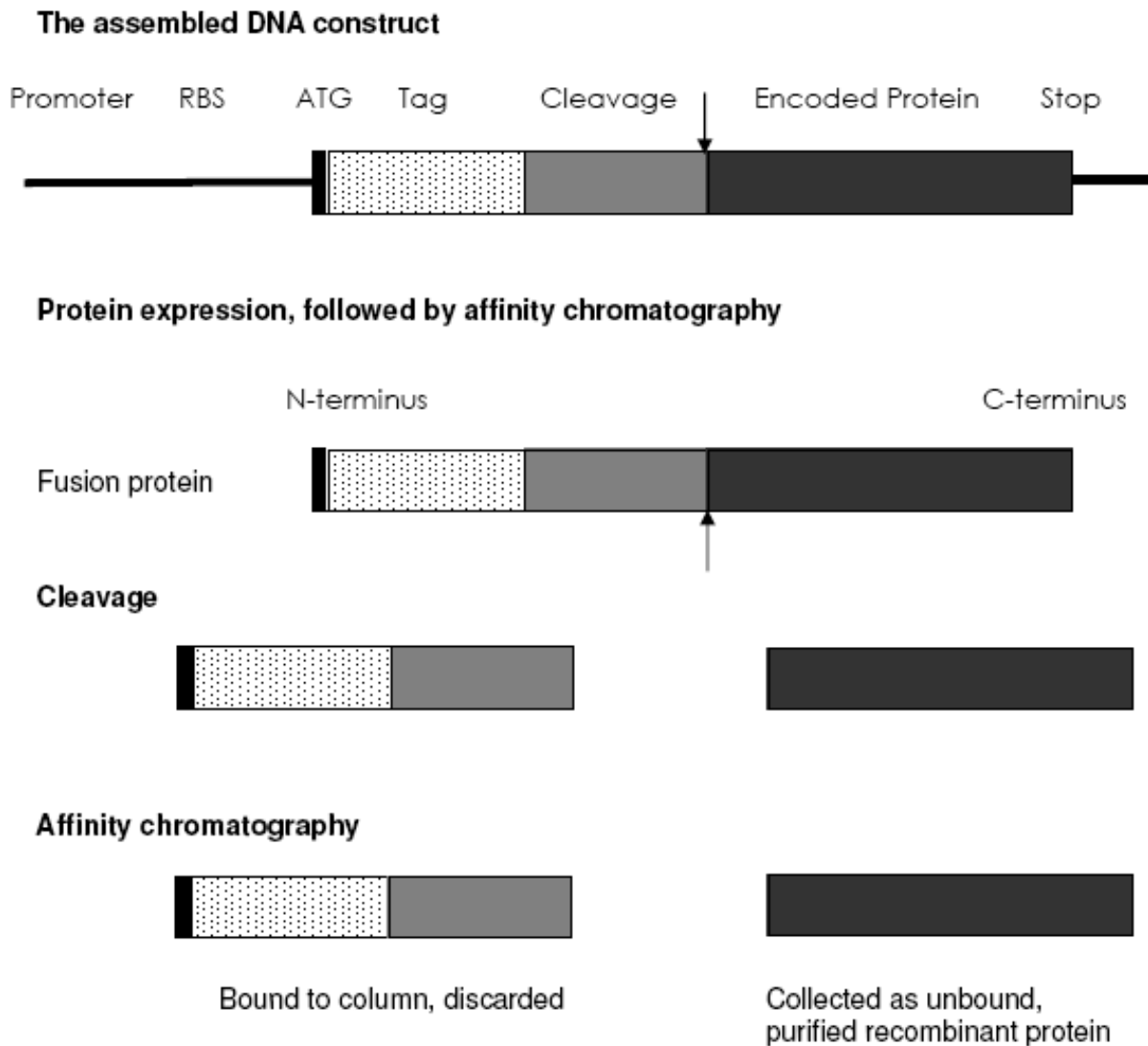
If a eukaryotic expression system is required, there are several alternatives. The simplest is yeast, and the two most usually employed are *Saccharomyces cerevisiae* (Baker's yeast), and the methylotropic yeast *Pichia pastoris*.

Yeast Mediated Protein Expression

The yeast expression system has been successfully used for several decades for the production of heterologous proteins of various origins, either human for therapeutic proteins or derived from pathogens for use as vaccines. Yeast are an important host used for research purposes to produce proteins that cannot be expressed in an active form in *E. coli*. In a similar way to prokaryote expression in *E. coli*, yeast has advantages such as ease of microbial growth and cultivation on inexpensive growth media, while adding the ability to perform many post-translational modifications such as O- or N- linked glycosylation, phosphorylation, disulphide bridge formation, proteolytic processing and folding in a eukaryotic system [15-17]. However, there are several complex post-translational modifications such as prolyl hydroxylation and amidation that yeasts cannot perform.

Saccharomyces cerevisiae

Saccharomyces cerevisiae has a long history of use in the food and beverage industries, and has been the workhorse for

The assembled DNA construct

Promoter   RBS   ATG   Tag   Cleavage │   Encoded Protein   Stop

Protein expression, followed by affinity chromatography

N-terminus   C-terminus

Fusion protein

Cleavage

Affinity chromatography

Bound to column, discarded

Collected as unbound,
purified recombinant protein

Fig. (1). Schematic diagram of the construct used to express recombinant protein, and purification of the expressed protein. This construct is conceptually similar regardless of the host that it is in, or if it is maintained on a plasmid (*E. coli, S. cerevisiae*) or integrated into the genome (*Pichia*, Baculovirus). Lines represent non-coding sequences, rectangle are coding sequences. RBS = Ribosome binding site. ATG = initiation codon, Tag = (e.g.) sequence encoding 6 x His, GST, etc. Cleavage = sequence encoding recognition sequence for a protease, Encoded protein= the recombinant protein that is to be expressed, Stop = stop codon. The arrow indicates both the site at which the coding sequence is inserted in-frame with the preceding sequences (ie: cloning site), and the point at which the protease will cleave the resultant recombinant protein. Once the recombinant protein is expressed and purified and proteolytically cleaved, the N- terminal sequences (tag, protease recognition sequence) can be removed by affinity chromatography, leaving the desired recombinant as a purified protein. Not drawn to scale- usually the coding sequence for the protein will be considerably larger.

gene expression for several decades. Expression can be driven from strong constitutive promoters such as alcohol dehydrogenase (ADHI) or enolase (ENO) or from inducible promoters (PHO, CUP1, GAL1 and G10). The main advantage of this yeast is to facilitate the direction of foreign protein into the secretory pathway, by fusion of the mature form of the desired recombinant protein to the signal sequence of yeast mating factor. Protein folding, formation of disulphide bridges and glycosylation occurs during secretion [15, 16]. We have used this system extensively for the secretion of parasite proteases, as have others [18-22].

One undesirable attribute of *S. cerevisiae* is the potential to hyperglycosylate proteins, which may hinder the antigenic properties or function of the protein of interest by masking

key epitopes or functional sites. In the other direction, the hyper-antigenic nature of terminal α, 1, 3 glycon linkages added to expressed proteins makes them unsuitable for therapeutic use [16]. For the purposes of generating a protein for vaccination purposes, modifications have been made to the encoding DNA, such that the amino acids that are the target for glycosylation are substituted. For example, we have made three codon substitutions in a liver fluke cathepsin B clone, substituting Ala for Asn, to remove three potential N-linked glycosylation sites (recognition sequence N-X-S/T) from a yeast-expressed protein [20]. However, if the natural state of the protein is to be glycosylated, such substitutions may abrogate natural immunogenicity and/or activity.

In these cases, alternative expression systems should be sought.

Finally, *S. cerevisiae* generally expresses recombinant proteins at a relatively low level. Other species of yeast, of which one is detailed below, can lead to significantly higher yields.

Pichia pastoris

Pichia pastoris retains all the advantages of expression in *S. cerevisiae*, but yields higher levels of protein [17]. This methanotropic yeast can drive expression from the alcohol oxidase I (AOX I) promoter, which is one of the strongest regulatory promoters known. There is the capacity to fuse expression plasmids at specific sites in the *P. pastoris* genome, either as single or multiple copy. Furthermore, the yeast is easy to grow to very high density in fermenters, ensuring high levels of recombinant protein production.

The AOX I promoter is tightly repressed by glucose and induction requires methanol as the carbon source. Although this gives maximum induction of recombinant protein expression, the usage of methanol is considered hazardous (as a product of petroleum) which may not be advisable for the production of therapeutic proteins [16, 23]. The constitutive glyceraldehyde 3 phosphate dehydrogenase (GAP) promoter is available to drive expression on glucose and glycerol substrates, and glutathione dependent formaldehyde dehydrogenase (FLDI) is another promoter that can be induced by methylamine (a non-toxic nitrogen source) [24].

Similar to *S. cerevisiae,* one of the major disadvantages of *P. pastoris* is in the potential for hyperglycosylation. Recent developments have created strains that can partially overcome this limitation [25, 26], however an alternative is to use insect cells for expression as described below.

Baculovirus-Mediated Insect Protein Expression

The application of insect cells for the production of proteins for commercial use has been expanding [27]. The system is based on the ability of recombinant baculovirus to infect insect cells, and the foreign protein that is encoded by the baculovirus is expressed by the cellular machinery. The expression system was first developed for the expression of human IFN-β in 1983. Baculovirus- meditated expression systems possess a number of advantages: high expression levels driven by the strong polyhedrin promoter (see below), ability of the insect cells to perform a variety of post-translational modifications, a high capacity for multiple genes or large inserts, and is bio-safe and easy to handle [27, 28]. Extremely high yields mean that the baculovirus expression system is an ideal tool for industrial protein production. Numerous recombinant proteins produced in insect cells have been tested as prophylactic vaccine candidates against an array of infectious agents such as HIV, hepatitis C virus, enterovirus, severe acute respiratory syndrome corona virus (SARS-CoV), dengue virus, rota virus, respiratory syncytial virus, human and avian influenza virus and a number of protozoan parasites. A number of commercially available vaccines such as Bajovac CSF E2[Tm] (Classical swine fever, Bayer AG), Ingelvac (porcine circo virus type 2, Boehringer Ingelheim) and FluBioK [Tm] (HA vaccine, Protein Science Corp) are produced in this expression system [11, 28].

The baculovirus genome contains a gene encoding an abundant protein, polyhedrin. This protein accumulates in infected insect cells (*Spodoptera frugiperda* Sf9, *Trichoplusiana* and cell lines from *Drosophila* and *Bombyx mori* have been used) at the end of the infectious cycle, and is then released into the culture medium. The polyhedrin protein is not essential for viral reproduction, and therefore the gene can be replaced with a heterologous coding sequence which is then expressed at high levels [29]. The system is amenable to suspension culture which makes it possible to generate a large amount of protein with relatively ease [11].

In order to produce recombinant protein, the heterologous coding sequence is cloned into a transfer vector which targets these sequences to the polyhedrin gene in the viral genome. After co-transfection of the transfer vector and viral genome into insect cells, the heterologous sequence inserts at the appropriate site, behind the late viral strong polyhedrin promoter. Compared to bacterial and yeast expression, it is a relatively time consuming process to generate clones for protein expression and many of the systems commercially are designed for more efficient production of recombinant viruses, which is a rate-limiting step in using this system.

However, one of the major advantages of baculovirus expression is that proteins expressed in insect cells are more "mammalian-like" than those expressed in bacteria or yeast. This is particularly the case where glycosylation is required, as the insect cells carry out N-linked glycosylation in a similar way to mammalian cells, although the attached glycans are often not as complex. However, this overcomes the problem of hyperglycosylation observed in yeast expression systems. Again, much effort has gone into modifying the baculovirus expression system to produce glycoproteins with mammalian-like glycans attached.

If mammalian proteins require post-translational modifications, and are required with exactly the same properties as is expressed in the host, mammalian tissue culture may be required. However, as this is extremely unlikely for the majority of proteins for veterinary application, this will not be discussed.

Choosing a System

As indicated earlier, which expression system is used is usually dictated by two things: the ultimate use of the product, and the nature of the individual protein to be expressed. If the protein being expressed requires any form of post-translational modifications (e.g. disulfide-bond formation, attachment of glycans etc) then bacterial expression will usually not be optimal, and a eukaryotic system is the best choice. Whether this is yeast (*S. cerevisiae* or *P. pastoris*), baculovirus or even mammalian cell culture depends on the modifications required, and how "native" the protein needs to be. For efficient secretion of disulfide-bonded proteins, yeast is sufficient, while expression in insect cells is superior to yeast for glycosylation.

As an example, we and others have expressed a variety of cathepsin proteases from the liver flukes *Fasciola hepatica* and *F. gigantica*. These proteases are secreted from the parasite as pro-enzymes, and contain a number of disulfide bonds. For this reason, they are difficult to express in *E. coli* in an active form, as disulfide bonds cannot be formed in the

reducing environment of the bacterial cytoplasm (as mentioned there are strategies available to enable such expression by targeting to the periplasm of *E. coli*). Therefore, active proteases have been expressed in either yeast [18-20, 30] or in insect cells [21]. As stated, the end-use of the recombinant protein must also be considered. For purposes such as determining activity, yeast expression to yield active proteases is used. The advantage of using recombinant proteins for this purpose is that it is easy to alter the protein by introducing a mutation in the coding sequence, and the resulting activity change in the protein can be assayed [30]. We have also used yeast to generate protein for vaccination [20], although this may not be necessary for some purposes. For example, Kesik and colleagues expressed cathepsin protease

in *E. coli*, and purified the resultant inclusion bodies which were used for vaccination. Although the protein present in inclusion bodies will be largely mis-folded, B and T cell epitopes will still induce immune responses, which in this case were protective.

There is also the possibility to purify inclusion body protein, and refold *in vitro* to yield active protein. Chow and colleagues have devised a set of rules that can be applied to many proteins, and this is available as a web server [31]. Using a refolding method, we have compared the *in vitro* enzymatic activity of cathepsin L1g expressed in either yeast of *E. coli*. The protease expressed in *E. coli* was isolated from inclusion bodies, and then refolded using step-wise urea gradient chromatography, dilution and dialysis. Yeast
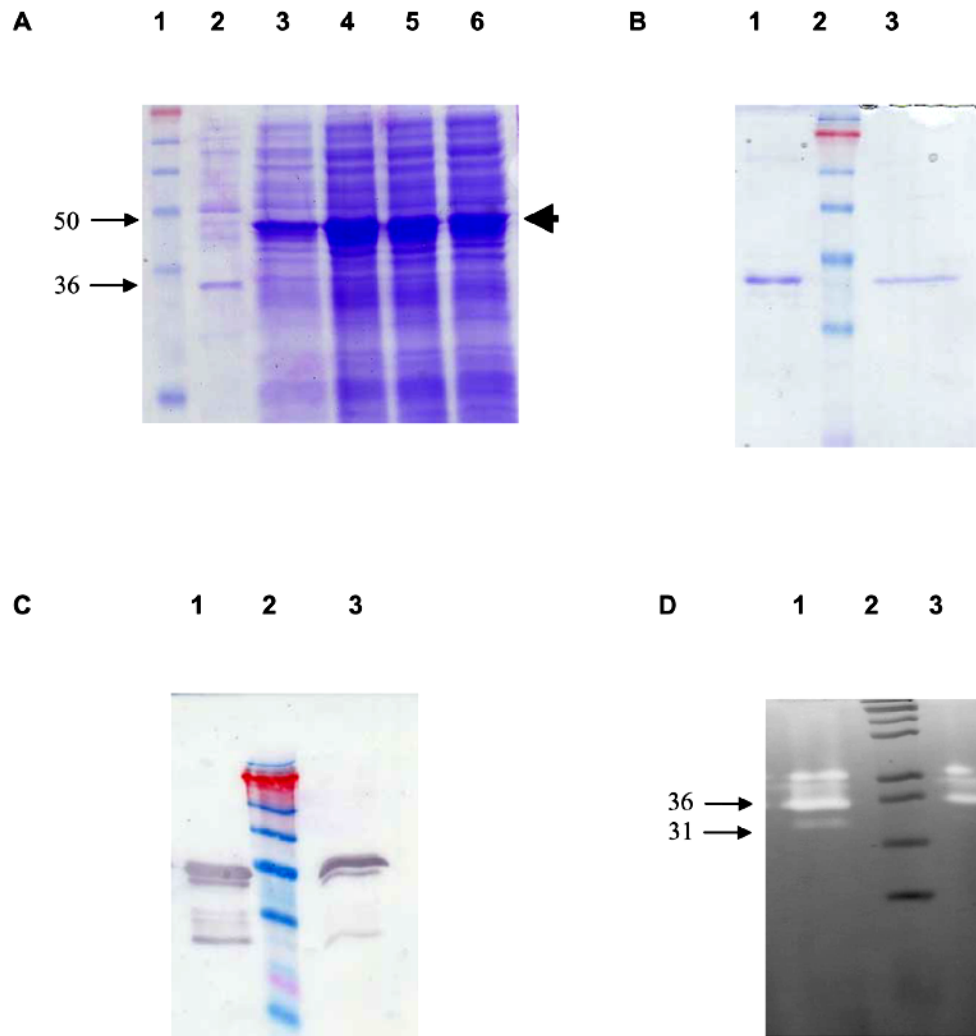
Fig. (2). SDS-PAGE analysis of expressed recombinant cathepsin L proteins. (**A**) *E. coli* cultures. After induction, 1 mL aliquots of culture were taken at hourly intervals. Cells were collected by centrifugation, resuspended in phosphate buffer, loading dye added and the samples electrophoresed. Lane1, marker; lane 2, cell lysate before IPTG induction; lanes 3- 6, lysates from 1 to 4 hrs post-induction. Substantial recombinant protein is expressed after induction, indicated by the bold arrow. (**B**) Purified cathepsin L. Protease was expressed in *E. coli,* purified by His-tag affinity chromatography, and re-folded, or expressed in yeast and purified by affinity chromatography. Lane 1, refolded cathepsin L; lane 2, marker; lane 3, yeast-expressed cat L. (**C**) Western blot analysis using mouse anti-cathepsin L sera. Lanes as for (**B**). Two reactive bands corresponding to pro-protein and mature protein can be observed. (**D**) SDS-Gelatin PAGE protease activity assay. Lanes as for (**B**). Active protease will digest the gelatine present in the gel, resulting in a cleared region after staining the gel with Coommassie blue. Proteolytic activity can be observed in both the *E. coli* and yeast-expressed samples. Indicative sizes for markers are given in kDa. Panels 1-3 used the same markers.

and *E. coli* derived protease was analysed using Western blotting and gelatin-substrate SDS-PAGE, and both were shown to exhibit immunoreactivity and proteolytic activity (Fig. **2**).

It is also apparent that not all eukaryotic expression systems are equal. Cathepsin L was expressed in both yeast and insect cells, and then tested for vaccination efficacy in a challenge model [21]. Only the baculovirus-mediated protein expressed in insect cells was able to induce protection, while the yeast expressed protein was ineffective. Part of the reason for this may be that while both systems produced glycosylated protein, the yeast- expressed protein was hyperglycosylated which may mask key epitopes on the surface of the protein. As stated above, the ability to make defined mutations in the encoding sequences can alleviate this. We have expressed a liver fluke cathepsin B in yeast after eliminating each of the three N-linked glycosylation sites present [20]. This prevents hyperglycosylation of the protein.

The ultimate use of the protein will also determine the expression system used. For example, if a protein is being used as a diagnostic target for humoral immunity, it may not matter if the protein is not in native conformation, as although some conformational epitopes may be lost, there will exist linear epitopes that can be recognised. We have previously expressed a cathepsin L from liver fluke in bacteria, and successfully used this (purified from inclusion bodies) in an ELISA to measure humoral responses to vaccination and infection (Smooker and Spithill, unpublished data). Similarly, producing protein for use in T-cell assays (e.g. ELISPOTS) does not necessarily require native protein.

CONCLUSION

The use of recombinant proteins in the veterinary sector is likely to increase. For example, as disease monitoring becomes a more widespread undertaking, diagnostic kits to perform these tasks will be required. Similarly, it is likely that many new vaccines will not be based on traditional killed or attenuated pathogens, but will be subunit vaccines produced as recombinant proteins.

Generally speaking, the default system will be bacterial expression, as this is the cheapest and easiest to set up. If the protein is bacterial, and a soluble protein, it will usually be expressed well in such systems, as will a large number of soluble eukaryotic proteins. However, some knowledge of the structure and ultimate use of your protein product can point to the system required, and save time and effort using a system not suited to your protein.

In deciding on a system to use, questions that can be asked include:

1.   Is the protein small, soluble and usually present in the cytoplasm? First choice *E. coli.*

2.   Is the protein normally extracellular, and contains disulphide bonds? If yes, consider a eukaryotic system (yeast or insect cells).

3.   If (2), does my protein need to be in native conformation? If no, then *E. coli* may be sufficient.

4.   Does my protein require glycosylation? If yes, use yeast or insect cells.

5.   Is the protein prone to hyperglycosylation? If yes, use *Pichia* or insect cells in preference to *S. cerevisiae*.

6.   Is my protein usually glycosylated, but this is not required for the application? If yes, express in *E. coli* if possible, or mutagenise amino acids in the glycosylation recognition sequence and express in eukaryotic systems.

REFERENCES

[1]   Walsh G. Biopharmaceuticals: recent approvals and likely directions. Trends Biotechnol 2005; 23(11): 553-8.

[2]   Babiuk LA. Vaccination: a management tool in veterinary medicine. Vet J 2002; 164(3): 188-201.

[3]   Lesley SA. High-throughput proteomics: protein expression and purification in the postgenomic world. Protein Expr Purif 2001; 22(2): 159-64.

[4]   Kimple ME, Sondek J. Overview of affinity tags for protein purification. Curr Protoc Protein Sci 2004; Chapter 9: Unit 9.9.

[5]   Cabrita LD, Dai W, Bottomley SP. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. BMC Biotechnol 2006; 6: 12.

[6]   Tyers M, Mann M. From genomics to proteomics. Nature 2003; 422(6928): 193-7.

[7]   Willis MS, Hogan JK, Prabhakar P, *et al*. Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions. Protein Sci 2005; 14(7): 1818-26.

[8]   Wingfield PT. Overview of the purification of recombinant proteins produced in Escherichia coli. Curr Protoc Protein Sci 2003; Chapter 6: Unit 6.1.

[9]   Baneyx F. Keeping up with protein folding. Microb Cell Fact 2004; 3(1): 6.

[10] Wingfield PT, Palmer I, Liang SM. Folding and purification of insoluble (inclusion body) proteins from Escherichia coli. Curr Protoc Protein Sci 2001; Chapter 6: Unit 6.5.

[11] Hunt I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. Protein Expr Purif 2005; 40(1): 1-22.

[12] Braun P, LaBaer J. High throughput protein production for functional proteomics. Trends Biotechnol 2003; 21(9): 383-8.

[13] Graslund S, Nordlund P, Weigelt J, *et al*. Protein production and purification. Nat Methods 2008; 5(2): 135-46.

[14] Tolia NH, Joshua-Tor L. Strategies for protein coexpression in *Escherichia coli*. Nat Methods 2006; 3(1): 55-64.

[15] Wacker M, Linton D, Hitchen PG, *et al*. N-linked glycosylation in Campylobacter jejuni and its functional transfer into *E. coli*. Science 2002; 298(5599): 1790-3.

[16] Cereghino GP, Cregg JM. Applications of yeast in biotechnology: protein production and genetic analysis. Curr Opin Biotechnol 1999; 10(5): 422-7.

[17] Sudbery PE. The expression of recombinant proteins in yeasts. Curr Opin Biotechnol 1996; 7(5): 517-24.

[18] Dowd AJ, Smith AM, McGonigle S, Dalton JP. Purification and characterisation of a second cathepsin L proteinase secreted by the parasitic trematode Fasciola hepatica. Eur J Biochem 1994; 223(1): 91-8.

[19] Jayaraj R, Piedrafita D, Dynon K, Grams R, Spithill TW, Smooker PM. Vaccination against fasciolosis by a multivalent vaccine of stage-specific antigens. Vet Parasitol 2009; 160(3-4): 230-6.

[20] Law RH, Smooker PM, Irving JA, *et al*. Cloning and expression of the major secreted cathepsin B-like protein from juvenile Fasciola hepatica and analysis of immunogenicity following liver fluke infection. Infect Immun 2003; 71(12): 6921-32.

[21] Reszka N, Cornelissen JB, Harmsen MM, *et al*. Fasciola hepatica procathepsin L3 protein expressed by a baculovirus recombinant can partly protect rats against fasciolosis. Vaccine 2005; 23(23): 2987-93.

[22] Smooker PM, Kennedy NJ, Steeper KR, Christopoulos H, Spithill TW. Fasciola: kinetics and quality of humoral responses to fatty acid binding protein and cathepsin l following delivery as DNA vaccines in mice. Exp Parasitol 2001; 97(3): 154-60.

[23] Gerngross TU. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol 2004; 22(11): 1409-14.

[24] Sunga AJ, Cregg JM. The Pichia pastoris formaldehyde dehydrogenase gene (FLD1) as a marker for selection of multicopy expression strains of P. pastoris. Gene 2004; 330: 39-47.

[25] Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N. Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. Nat Protoc 2009; 4(1): 58-70.

[26] Vervecken W, Callewaert N, Kaigorodov V, Geysens S, Contreras R. Modification of the N-glycosylation pathway to produce homogeneous, human-like glycans using GlycoSwitch plasmids. Methods Mol Biol 2007; 389: 119-38.

[27] Hu YC. Baculovirus as a highly efficient expression vector in insect and mammalian cells. Acta Pharmacol Sin 2005; 26(4): 405-16.

[28] Hu YC, Yao K, Wu TY. Baculovirus as an expression and/or delivery vehicle for vaccine antigens. Expert Rev Vaccines 2008; 7(3): 363-71.

[29] Kesik M, Jedlina-Panasiuk L, Kozak-Cieszczyk M, Plucienniczak A, Wedrychowicz H. Enteral vaccination of rats against Fasciola hepatica using recombinant cysteine proteinase (cathepsin L1). Vaccine 2007; 25(18): 3619-28.

[30] Smooker PM, Whisstock JC, Irving JA, Siyaguna S, Spithill TW, Pike RN. A single amino acid substitution affects substrate specificity in cysteine proteinases from Fasciola hepatica. Protein Sci 2000; 9(12): 2567-72.

[31] Chow MK, Amin AA, Fulton KF, *et al*. The REFOLD database: a tool for the optimization of protein expression and refolding. Nucleic Acids Res 2006; 34(Database issue): D207-12.