Selecting appropriate hosts for recombinant proteins production: Review article

Golam Reza Sharifi-Sirchi 1, Mokhtar Jalali-Javaran 2

Associate Professor Department of Biotechnology and Plant Breeding 1, University of Hormozgan, Bandar Abbas, Iran. Associate Professor Department of Biotechnology and Plant Breeding 2, Tarbiat Modares University, Tehran, Iran.

(Received 11 Mar, 2015 Accepted 10 Mar, 2016)

Abstract

In recent years, the number of recombinant proteins used for therapeutic applications and industry has increased dramatically. Recombinant proteins are produced in many host organisms (microbial, insect, plant and mammalian cells). There are many factors to consider when choosing the optimal system for protein expression and purification including the mass, purity or solubility of the recombinant protein, its number of disulphide bonds, the types of post-translational modifications desired, the destination of the expressed protein, the cost and simplicity of the recombinant protein production system and finally regulatory aspects in the case of pharmaceutical production. If a soluble, non-toxic protein needs to be produced, bacteria are the first choice. However, if a complex post-translational modification is required, eukaryotic cells have to be the expression host. E. coli, Saccharomyces cerevisiae and mammalian cell lines containing Baby Hamster Kidney (BHK), Chinese Hamster Ovary (CHO) and hybridoma cells are main host systems for production of most antibodies and proteins for diagnostic and therapeutic use. There is not general expression host system to act optimally for all recombinant proteins. Therefore, several expression host system must be experimentally evaluated for each recombinant protein.

Key words: Expression System, Host, Recombinant Proteins

Introduction:

In recent years, the number of recombinant proteins used for therapeutic applications has increased dramatically. Many of these applications involve complex glycoproteins and antibodies with relatively high production needs. These demands have driven the development of a variety of improvements in protein expression technology, particularly involving mammalian, insect and plant cell and microbial culture systems (1). The firsts confirmed biopharmaceuticals in the end of twentieth century were recombinant proteins with unaltered sequences or murine antibodies. In these days, engineered products have taken possession of the International markets. These products are for example hormones like insulin or growth hormones, growth factors like erythropoietin, interleukin-based products, blood clotting factors (e.g. factor VIII) and interferon but also humanized antibodies or fragments thereof, vaccines and therapeutic enzymes (2,3). The purpose of this paper is to guide the investigator in the decision-making process for choosing an appropriate expression system regarding target protein.
Important factors for selecting expression system

Process evaluation of recombinant protein production involves several variable including host organism, growth conditions, mediums and inducers which should be considered by multivariate analysis (4). Factors to consider when selecting an expression system include the mass, purity and or solubility of the recombinant protein, its number of disulphide bonds, type of posttranslational modifications, the destination of the expressed protein and/or cost and simplicity of recombinant protein production system. Post-translation modifications encompass correct disulphide formation in the endoplasmic reticulum (ER), proteolytic cleavage of an inactive precursor, glycosylation, or e.g. phosphorylation, acetylation, sulfation and fatty acid addition or prenylation (5). In the following paragraphs, some of these factors will be considered.

Cytoplasmic Proteins

For a cytoplasmic protein, the optimal choice of an expression system depends on the protein mass and the number of disulphide bonds in the protein. For proteins between 10 and 50 kDa and containing few disulphide bonds, E. coli is a good option for soluble protein expression (6). For larger proteins or those with many disulphide bonds, if soluble expression is desired then usually either baculovirus or yeast is the preferred choice.

Although, cellular metabolism of the host can be influenced by catalytic properties of some engineered protein, but universally, expression of recombinant proteins is well tolerated by the host cell even though it induces a "metabolic burden". A frequent problem when expressing (Eucaryotic) proteins in E. coli cells is that the protein is forming insoluble aggregates called “inclusion bodies”. The inclusion body formation and metabolic burden are two nearly distinct main factors to determine soluble recombinant protein production ability of host cell. There are a number of available protocol changes that help in redirecting proteins into the soluble cytoplasmic fraction from inclusion bodies in of E. coli. On the whole, these protocol changes can be created into procedures where the expression strategy is modified to obtain soluble expression or procedures led to protein refolding from inclusion bodies (7) including: changes in pH, reduction of cultivation temperature, substrate depletion and alteration in dissolved oxygen concentration. Also, the use of specific E. coli expression strains (e.g. trxB and/or trxB/gor strains) can help to reduce inclusion body formation (8). An elegant way to increase solubility of the recombinant protein is to co-express its interaction partner which can cause proper folding. Finally, use of alternative strategies like use of systems based on artificial oil bodies and in vivo rescuing systems based on E. coli ribosomes have been shown to tackle inclusion body problems (9). If the recombinant protein can be tagged, it ease its purification by affinity chromatography. The maltose binding protein (MBP)-tag is known to help increase solubility (10). Moreover, screening for and selection of more soluble variants by directed evolution is also another option.

Secreted Proteins

Secretion of recombinant protein has several superiorities, including: more correct protein folding chance like disulphide-bond rearrangement which is catalyzed by two preplasmic disulphide bond isomerases; DsbC and DsbD, easiness of purification, decrease of N-terminal Met extension and protease attack.

To improve production of secreted complex proteins, periplasmic chaperones and proteases can be manipulated (11). At least three different types of protein secretion systems (type I, type II, and type III) in Gram-negative bacteria have been identified which type II being the most widely employed. The type II system use the Sec pathway to export a premature protein (with signal sequence) into the periplasmic space and then, process to a mature protein. Extracellular secretion of recombinant proteins is promoted in harvested E. coli cells by several methods. These include the use of biochemical, chloroform shock, lysozyme treatment and physical methods like osmotic shock, freezing and thawing (11). Fusing the recombinant product (which can no secret normally) to a protein expressed on the outer membrane (e.g. OmpF) or to a carrier protein that is normally secreted into the medium (e.g. hemolysin) are another strategies to extracellular protein production. For example, human β-endorphin could be secreted into the culture medium when fused to OmpF (12).
Secretory production and protein folding can be improved by use of co-expression of periplasmic chaperones like FkpA, Skp, SurA, and disulphide-bond formation (DSB) family proteins. Any of the expression hosts can be used to produce secreted proteins. However, E. coli may be suboptimal for expressing secreted eukaryotic proteins but this is highly dependent on the downstream application (12,13).

**Membrane Proteins**

Membrane proteins are encoded by about 20-30% of all proteins (14). Membrane proteins represent an extremely challenging class of proteins to express in large quantities. Difficulty of high-level expression of recombinant protein in E. coli is due to toxic effects exerted by hydrophobic protein domains of the host. Also, presence of large amounts of membrane protein could disturb membrane function and/or constitutive activity of the recombinant membrane proteins can promote intracellular signaling or impair expression host growth by the function of the recombinant membrane protein itself (e.g. ion channel). On the other side, fusing of full-length membrane proteins is highly desirable since their transmembrane segments (TMS) often contain important structural information leading proper sub cellular sorting, oligomerization, and folding of recombinant protein (15). Membrane proteins expression in bacteria is low and may direct better integration into the inner membrane (target region for membrane proteins biological activity). As high-level expression of membrane proteins is the rate-limiting factor, use of weak promoter, low copy number plasmid, low temperature and combination of them during expression is most beneficial. Moreover, for correct folding and function of the recombinant membrane proteins, they might need post-translational modifications, like N-glycosylation, disulphide-bond formation for which a specific lipid composition of the host membrane and molecular chaperones would be required. Often, production of the soluble, hydrophilic portion will suffice. In this condition, the membrane spanning domain can be removed and the desired soluble portion can be expressed (16). There are no clear guidelines on choosing the best system to express intact membrane proteins. Researchers have reported modest success with expressing G protein-coupled receptors using the baculovirus and yeast methods. Membrane protein synthesis by cell-free expression does not appear to be restricted by origin, size or topology of the target, and its global application is therefore a highly valuable characteristic (17).

**Toxic Protein**

In biological systems, a particular protein need to be functional and beneficial. Over-expression of recombinant foreign protein into host cell generally results in different temporal, spatial and quantitative expression than its expression in native host. Also, the function of the expressed recombinant protein often is not needed by the host cell. In some cases, the recombinant protein therefore can cause host cell death. This phenomenon is nominated as protein toxicity and responsible recombinant proteins are called as toxic proteins. Recombinant DNA fragments are normally not toxic to the host cell except they have repetitive sequences with high affinity to some transcription factors. Therefore, it might sequester those away from their regular cellular location. Also, a recombinant DNA or RNA may be toxic to the host cells. On the whole, It is estimated that host toxicity is caused by recombinant protein (80%), recombinant RNA (15%) and recombinant DNA (less than 1%).

Recombinant proteins (which are toxic to the expression host) can be challenging to produce, but this obstacle can usually be overcome. If the recombinant protein is toxic, it is often useful to determine whether the problem is host cell specific. If so, then the protein can be expressed in a more compatible expression host. Some technologies were used to overcome toxic protein such as: induction conditions (induction time, induction temperature and concentration induction), presence of inhibitor in the medium, high concentration of antibiotic in medium, plating inoculation, use of lac I strains, low copy number plasmid, tightly regulated vector, weak promoter, strong terminator upstream of inducible promoter, fusion expression, individual domains, different cellular compartment, fusion to insoluble protein. Another option is to use a tightly regulated, inducible expression system such as those available for E. coli and P. pastoris. For example, several elaborate inducible expression systems have been developed for E. coli (18). To
avoid the toxicity of recombinant DNA, the recombinant DNA fragment should be cloned in the site flanked by some repressors such as vectors with multiple operators.

In addition, some established systems for expressing recombinant proteins, such as transgenic animals, are technically too challenging, time consuming and prohibitively expensive to be a viable option for the average laboratory.

**Selecting the right production host**

It need to select the right host for obtaining the suitable final product. Thus, if a soluble, non-toxic protein needs to be produced, bacteria are the first choice. However, if a complex post-translational modification is required, eukaryotic cells have to be the expression host. As the host development times are significantly different, choosing the right expression host will reduce the time to market. The starting point for any cloning and expression effort is *E. coli* since it has easy and rapid cultivation system and dedicate a variety of collection of expression system. However, there is not any common expression host system which can be optimized for all recombinant proteins.

*E. coli*, *Saccharomyces cerevisiae* and mammalian cell lines containing Baby Hamster Kidney (BHK), Chinese Hamster Ovary (CHO) and hybridoma cells are main host systems for production of most antibodies and proteins for diagnostic and therapeutic use. Just in three years 2003-2006, 31 new antibody and recombinant protein products were confirmed which *E. coli* and mammalian cell were host system for nine and 17 of these products respectively (2). Mammalian cells are the most used host system for monoclonal antibodies production.

**Escherichia coli Bacteria**

*E. coli* has several advantages than other host system including homogeneity of the recombinant protein, low production cost, and short generation time and some disadvantages like tendency to form inclusion body, inefficient in disulphide bonds formation, lack of ability to perform post-translational modifications and secretary mechanism (19,20). Many recombinant hormones (e.g. insulin) and interferons which do not require post translational modifications are produced commercially in *E. coli* (2). The growth media for *E. coli* are inexpensive and there are relatively straightforward methods to scale-up bioproduction.

In *E. coli*, recombinant proteins are normally either directed to the cytoplasm or to the periplasm. Proteins directed to the cytoplasm can be the most efficiently expressed ones, giving yields of up to 30% of the biomass. The rate of translation and folding in *E. coli* is almost 10-fold higher than that observed in eukaryotic cells, and this presumably contributes to the inclusion body formation of eukaryotic proteins (21). Several methods have been described for maximizing the formation of soluble, properly folded proteins in the cytoplasm and minimizing inclusion body formation. The most straightforward method involves lowering the temperature to 15–30 °C during the expression period (22). Presumably, the reduced temperature slows the rate of transcription, translation and folding, thereby allowing proper folding (22). In addition, lower temperature has been shown to decrease heat shock protease activity which could otherwise contribute to degradation of the recombinant protein. *E. coli* is normally inefficient in promoting the correct formation of disulphide bonds when recombinant proteins are expressed in the cytoplasm; normally disulphide bond formation occurs only in the periplasm where it is catalyzed by the Dsb system (23). Consequently, if disulphide bond formation is needed, the recombinant protein should be directed to the periplasm via a cleavable signal peptide (e.g., pelB). However, a major disadvantage of periplasmic expression is the significant reduction in yield. Because of remarkable ability to enhance the solubility and promote the proper folding of its fusion partners, *Escherichia coli* maltose-binding protein (MBP) has emerged as an attractive vehicle for the production of recombinant proteins (24,25).

**Yeasts**

Yeasts have many of the advantageous features of *E. coli* such as a short doubling time and a readily manipulated genome, but also has the additional benefits of a eukaryote that includes improved folding and most posttranslational modifications. A number of approved biopharmaceuticals such as growth hormone, vaccines (recombinant hepatitis B vaccines) and most notably insulin products
obtained from engineered *S. cerevisiae*. *Pichia pastoris* is an attractive and beneficial alternative host, because of high cell density growth, post translation modification, lower degree of glycosylation and easiness of manipulation (26).

Yeast have faster growth rates and higher amounts of the protein products than mammalian cell host. Additional advantage of host system is the secretion of protein products into the growth medium, which usually makes the downstream processing easier (26). However, post-translational modification of yeast differs from mammalian cells like unwanted glycosylation which is a possible disadvantage of yeasts. Although, degree of glycosylation is depended on yeast strain and medium (27). The length of time to assess recombinant gene expression with the *P. pastoris* method is approximately 3-4 weeks which includes the transformation of yeast, screening the transformants for integration and an expression time course. An appealing feature of *P. pastoris* is the extremely high cell densities achievable under appropriate culture conditions (28). Using inexpensive medium, the *P. pastoris* culture can reach 120 g/l of dry cell weight density. *P. pastoris* has been used to obtain both intracellular and secreted recombinant proteins. Like other eukaryotes, it efficiently generates disulphide bonds and has successfully been used to express proteins containing many disulphide bonds.

**Baculovirus/Insect Cells**

Baculovirus-mediated expression in insect cells offer a beneficial recombinant protein production system. The most commonly used baculovirus which manipulated for recombinant protein production, *Autographa californica* virus, is a lytic, large (130 kb), double-stranded DNA virus. *Autographa californica* virus is routinely amplified in derived insect cell lines from the fall armyworm *Spodoptera frugiperda* (SF 9, SF 21), and recombinant protein expression is completed either in the aforementioned lines or in a line derived from the cabbage looper *Trichoplusia ni* (High-Five) (29). Insect cells host system have several advantages like disulphide bonds formation, performing most post translational modification of mammalian cell like production of glycoproteins with the complex N-linked glycans, ease of culture, higher expression levels and high tolerance to osmolality and by-product concentrations (30).

The first vaccine (Cervarix® by GlaxoSmithKline, UK) which was against certain types of the human papillomavirus the causative agent of cervical cancer, produced in insect cells using the baculovirus system was approved within the EU, in Australia and in Philippines.

**Mamalian Cells**

Large biomolecules with specific glycosylation and posttranslational modification should be produced in mammalian cell culture. Mammalian expression systems with secretary mechanism and serum free media (for the CHO and HEK293 cell lines) simplify the purification of recombinant protein and bring down manufacturing cost. Hybridoma cells, CHO, murine myeloma cell and murine myeloma cell lines are the most used cell lines for mammalian cell cultivations in biopharmaceutical productions (31). Now, high cell line with improvements of the titers are expected to have productivity more than 10 – 15 g/l (33).

**Plants**

Plants are beneficial expression systems which present unique advantages than conventional host systems (bacteria, mammal and yeast cells). Some of their advantages are production of animal pathogens (such as prion, virions and etc.) free yield, lower manufacturing cost, easy scale-up and purification technology, appropriate folding and glycosylation, easiness of culture, harvest and processing. Also, they have this ability to direct proteins to the environments which reduce protein degradation and increase stability (33-36). Plants have homologous chaperon to mammalian cells which enable them to efficient control of protein assembly. N-glycosylation and other post-translational modifications can be performed on recombinant proteins by the plant secretory pathway (37,38).

The first recombinant plant-derived pharmaceutical protein, human serum albumin, initially produced in transgenic tobacco and potato plants. Recombinant protein obtained by stable nuclear transformation, stable plastid transformation and transient transformation (39,40). Currently, molecular farming products presented into four
broad areas of parental therapeutics, pharmaceutical intermediates, industrial proteins (e.g., enzymes), and monoclonal antibodies (MAbs) and antigens for edible vaccines (38,39).

**Animals**

Blood, milk, egg white, seminal plasma, urine and silk gland, insect larvae haemolymph are theoretically possible hosts to produce recombinant proteins. Silk gland is a promising system in particular cases (40-44). But milk from farm mammals and chicken egg white systems have being implemented. Human antithrombin III, obtained from farm transgenic mammals milk, received the agreement of EMEA (European Agency for the Evaluation of Medicinal Products) to release to the market in 2006 (45). Rabbits, pigs, sheep, goats and cows are implemented to produce recombinant proteins in their milk. Also, two monoclonal antibodies and human interferon-b1a have been recovered from chicken egg white (45).

**Conclusion:**

Antibodies and recombinant proteins for diagnostic or pharmaceutical use can be produced in many different host systems (microbial, insect, mammalian cells, yeast and transgenic plant and animal). Each system has both advantages and disadvantages for expressing recombinant proteins. High quality and quantity of product and choice of system is largely protein dependent. A careful evaluation of the characteristics of the recombinant protein along with the downstream application must be considered when selecting an expression method. Since some of circumstances in the time of appropriate expression system choice are not obvious, several expression host system must be evaluated. Also, it is important that for each host system, variables like: medium components, growth condition and host population, be analyzed by multivariate statistical designs.

**Acknowledgment:**

Author like to convey thanks to the Ministry of Science and technology of Iran for providing the financial means and laboratory facilities.

**References:**


24. Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 1999;8(8):1668-1674.


Appropriate Hosts for Proteins

Golam Reza Sharifi-Sirchi, et al


222