TeaserProduction of recombinant proteins is essential for drug development and discovery, but can often be problematic; why do these processes fail and how can these problems be overcome?

Recombinant protein production in bacterial hosts

Tim W. Overton

Bioengineering, School of Chemical Engineering, The University of Birmingham, Birmingham, B15 2TT, UK

Tim W. Overton is a lecturer in the School of Chemical Engineering at The University of Birmingham. A biochemist and molecular microbiologist by training, his research focuses on bioprocess optimisation using molecular microbiology and flow cytometry approaches. Currently, he is investigating recombinant protein production and biofilm-mediated biotransformation processes and the elimination of bacteria from varied feedstreams. He is the director of the MSc programme in Biochemical Engineering, which trains science and engineering graduates to become the next generation of bioprocess engineers, with a focus on biopharmaceutical engineering.

The production of recombinant proteins is crucial for both the development of new protein drugs and the structural determination of drug targets. As such, recombinant protein production has a major role in drug development. Bacterial hosts are commonly used for the production of recombinant proteins, accounting for approximately 30% of current biopharmaceuticals on the market. In this review, I introduce fundamental concepts in recombinant protein production in bacteria, from drug development to production scales. Recombinant protein production processes can often fail, but how can this failure be minimised to rapidly deliver maximum yields of high-quality protein and so accelerate drug discovery?

Proteins and peptides are the mainstay of the biopharmaceutical sector; over 200 protein drug products are currently on the market [1,2], and more are currently undergoing preclinical and clinical trials. However, proteins are complex molecules in terms of their structure and function and, unlike many pharmaceuticals, cannot be synthesised chemically. Therefore, proteins are manufactured in biological processes, usually inside host cells (although a growing number of cell-free expression technologies are available). These proteins, synthesised in a host cell frequently of a different species to their origin, are termed 'recombinant proteins' (see Glossary) because the DNA encoding them has been recombined or engineered.

Recombinant proteins are required at different stages of the drug discovery process and in different quantities. Initial drug development studies frequently involve the structural determination of proteins that are drug targets, for example human membrane proteins [3]; such structural studies are often required in the development of protein and small molecule drugs. These studies typically require small quantities of recombinant protein (on the milligram scale). Further in the development process, larger quantities of protein drug are required for preclinical and clinical trials, synthesised under current good manufacturing practice (cGMP) conditions [4]. Systems must be in place to enable the synthesis of recombinant proteins in this range of scales so that drug development can proceed. Large drug companies undertake much of this work in-house, whereas other companies might contract out recombinant protein production to contract manufacturers.

E-mail address: t.w.overton@bham.ac.uk.

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2013.11.008

GLOSSARY

Affinity chromatography separation of proteins based on binding affinity. For example, MBP has high binding affinity to amylose; therefore, amylose can be used to attract MBP and selectively purify it from other proteins.
Bacteriophage a virus that infects bacteria.
Bioreactor a tank, usually made of stainless steel or glass, that is used for growth of organisms. Bioreactors are usually designed to maintain culture conditions within certain parameters (e.g. temperature and pH control), mix the contents and supply the cells with oxygen and nutrients. Bioreactors also enable containment, ensuring that the cells are not contaminated by external factors or vice versa.
Current good manufacturing practice (cGMP) the set of guidelines that governs the safe production of pharmaceuticals for human use, as defined by regional drug regulatory bodies, such as the FDA (http://www.fda.gov) and Medicines and Healthcare Products Regulatory Agency (MHRA [4]).
Fab fragment a truncated antibody fragment comprising one light chain (V_L and C_L domains) and one truncated heavy chain (V_H and C_H1 domains), disulfide-bonded together. The Fab fragment contains the antigen-binding portion of the antibody.
Glycosylation oligosaccharide moieties are frequently attached to proteins of eukaryotic origin. This process occurs in the endoplasmic reticulum and is often essential for correct protein function. Antibodies usually require glycosylation for correct function. Lack of glycosylation, or incorrect glycosylation (the incorrect sugar molecules being attached to the protein), leads to loss of protein function and frequently rapid clearance of protein drugs from the body.
Gram negative/positive classification of bacteria depending upon cell wall structure. Gram-negative bacteria have two membranes separated by a periplasm, which contains a thin layer of peptidoglycan, a structural polysaccharide. Gram-positive bacteria (e.g. *Bacillus* spp) have cell walls comprising a single membrane and a thicker layer of peptidoglycan.
Isopropyl β-D-1-thiogalactopyranoside (IPTG) inducer molecule for the Lac and pET expression systems. IPTG binds to the Lac repressor LacI, whose natural ligand is the disaccharide lactose, resulting in repression of DNA binding. IPTG is used in recombinant protein production as, unlike lactose, it is non-metabolisable and, therefore, is not degraded by bacteria over time.
Metabolic burden the stress caused by requirements for energy generated by metabolism (in the form of ATP) and metabolic intermediates within the cell.
Plasmid a circular DNA fragment, usually 1–100 kb in size, that is replicated independently of the host cell chromosome.
Post-translational modification biochemical modifications to a protein that occur following translation. These include glycosylation, phosphorylation, and acylation [5].
Promoter the region of DNA upstream of a gene or set of genes that specifies when transcription occurs.
RNA polymerase a multisubunit enzyme complex that catalyses the process of transcription. RNA polymerase recognises and binds to a promoter region upstream of a gene and then generates the RNA molecule corresponding to the DNA sequence of the gene.
RNA polymerase sigma S subunit (RpoS) a master regulator of the general stress response in *E. coli*.
Single-chain variable fragment (scFv) comprises the variable domains of the light (V_L) and heavy (V_H) chains of an antibody, linked together by a flexible peptide linker. It contains the antigen-binding portion of the antibody.
Site-directed mutagenesis specific changes to DNA nucleotide sequences that are made to alter the peptide sequence of a protein, the promoter sequence or codon usage.
Transformation the process by which bacteria are induced to take up plasmid DNA. When containing the plasmid, the bacteria are termed 'transformants'.

The first decision that must be made when making a recombinant protein is choice of the host system. Before choosing a host, the chemical properties of the desired recombinant protein are investigated; if the desired protein is glycosylated or otherwise extensively post-translationally modified [5], then a eukaryotic expression system is usually chosen, because such modifications are frequently essential for correct protein function and bacteria are currently unable to incorporate such modifications. However, disulfide bond formation (DSB) is possible in bacteria [6]. The development of eukaryotic-like post-translational modification in bacterial hosts is a current area of research, recently reviewed by Nothaft and Szymanski [7]. Some possible eukaryotic expression systems are outlined in Box 1.

For proteins that are not required to be synthesised in a glycosylated or extensively post-translationally modified form, bacteria are an excellent expression system because of their relative sim-

BOX 1

Eukaryotic hosts: their advantages and drawbacks.

A central problem with bacteria as hosts for recombinant protein production is their inability to post-translationally modify proteins in the way that human cells can, for example glycosylation (attachment of antennae of specific sugar epitopes to proteins) [5]. One reason for this is their different cellular structure; bacteria lack the endoplasmic reticulum and Golgi apparatus, the sites of post-translational modification. If a recombinant protein is required to be glycosylated or otherwise modified (e.g. phosphorylated or lipidated), then a eukaryotic host is usually used. Many glycosylated and otherwise post-translationally modified protein biotherapeutics are generated in cultured mammalian cell lines such as CHO cells (recently reviewed in [69]). Manufacture in these hosts is more expensive and complicated than bacterial processes, owing to the higher cost of culture media, low cell tolerance for changes in reaction conditions and slow growth rates. Typical protein drug products of CHO cells include monoclonal antibody therapies for treating cancer and erythropoietin [1].
A halfway house between mammalian cells and bacteria remains yeast systems, typified by the bakers' and brewers' yeast *Saccharomyces cerevisiae* and the methylotrophic (methanol-utilising) *Pichia pastoris* (reviewed in [70]). Given that yeasts are eukaryotic, they are able to generate some post-translational modifications such as glycosylation; current research is focusing on engineering yeast glycosylation patterns to mimic human cells. Yeast are also simple and quick to grow and can generate high yields of recombinant proteins. Yeast-generated protein drugs on the market include vaccines and insulin [1,14]. Ultimately, the choice of host cell system is usually a compromise between the ease and cost of growth and the overall yield and function of the generated recombinant protein.