Plethora of Potential Platforms: a review of current and future trends in expression system choice in the biopharmaceutical industry

by
Sarah Louise Foley

This project is submitted in part fulfillment of the HETAC requirements for the award of Master of Science in Biopharmaceutical Science
May, 2018

Title: Plethora of Potential Platforms: a review of current and future trends in expression system choice in the biopharmaceutical industry

Name: Sarah Louise Foley

ID Number: [redacted]

Supervisor: Kate Cotter

Declaration:

"I hereby declare that this project is entirely my own work and that it has not been submitted for any other academic award, or part thereof, at this or any other education establishment".

Sarah Foley

Abstract

In the thirty years since the first biopharmaceuticals were approved for use in humans, the choice of expression system for new biopharmaceuticals has remained largely unchanged and is still mainly determined by the necessity of human-like post translation modifications, such as glycosylation, for the required method of action of the therapeutic protein. Currently, those biopharmaceuticals requiring post translational modifications are overwhelming expressed in mammalian systems, while those that do not are expressed in microbial systems. Chinese Hamster Ovary and *Escherichia coli* are the overwhelming front-runners, respectively, with over 85% of all biopharmaceuticals between them. While alternative systems such as methylotrophic yeast, human, transgenic animals, insects and plants have all made the leap to commercial manufacture, the inherent conservatism of the industry and regulatory authorities, and the inefficient or non-human glycosylation patterns have so far precluded alternative expression systems from displacing the dominant platforms outside a few niche areas. Expanding knowledge of glycosylation pathways and new genetic engineering tools, especially the site-specific CRISPR/Cas9 nucleases, offer a path forward for a wider choice of expression platforms to meet the ever-expanding needs of the biopharmaceutical industry.

Acknowledgements

The author wishes to thank her project supervisor Kate Cotter for her valued input and patience throughout the project, her Pfizer colleagues for their support and encouragement, and her housemates for their generous sacrifice of the use of the kitchen table and for multiple cups of tea.

Contents

Declaration i
Abstract ii
Acknowledgements iii
Chapter 1: Introduction 3
Chapter 2: Literature Review 6
  Post Translation Modifications 6
  Current Status of the Biopharmaceutical Industry 10
  Mammalian (Non-human) Expression Systems 11
  Prokaryotic Expression Systems 21
  Lower Eukaryotic Expression Systems 28
  Human Expression Systems 33
  Transgenic Animal Expression Systems 38
  Insect Expression Systems 42
  Plant Expression Systems 48
  Cell-free Expression Systems 55
  CRISPR/Cas9 56
Chapter 3: Overall Conclusions and Recommendations 62
Bibliography 69

Chapter 1: Introduction

In 1982, Genentech's Humulin, recombinant insulin produced in *Escherichia coli* (E. coli) cells, became the first recombinant protein approved for therapeutic use in humans. It became the world's first biopharmaceutical, a term coined in the early 1980s to differentiate these new recombinant products, produced by modern molecular biological methods, from more traditional biological products, extracted from native sources (Walsh 2003).

While the potential for proteins naturally produced in the body to treat various diseases had been long appreciated, efforts to take advantage of this potential were often stymied by the small quantities produced in nature and the laborious, expensive, and potentially unsafe methods of harvesting them. The advent of recombinant (combining DNA from two different organisms) technology has helped to overcome such difficulties and has had a positive effect on the production of clinically useful proteins in multiple ways (Walsh 2003).

First, it eliminates the problem of source availability: clinically important proteins, such as interferons and interleukins, are produced in only minute quantities in the body. Second, it reduces potential safety issues: extraction of proteins from native sources has in the past led to the unwitting transmission of infections such as blood-borne diseases such as hepatitis C or human immunodeficiency virus (HIV). Third, it eliminates having to harvest from unreliable, distasteful or dangerous sources such as urine (fertility-related hormones) or from milking venomous snakes (anti-coagulants). Finally, it permits the engineering of the native protein to produce new proteins with an improved clinical benefit, such as improved half-life in the body, or more selective targeting (Walsh 2003).

The *sine qua non* of recombinant technology and the beginning of any biopharmaceutical is the availability of an expression system to express the protein of interest. An expression system is the combination of the expression vector (the method by which the DNA from one species is transferred to another), the gene for the protein of interest, and the cellular host for the expression vector (Biology Online Dictionary 2018).

Unlike traditional small molecular weight chemical drugs, which are synthesized by highly controlled and reproducible chemical reactions, biopharmaceuticals are manufactured in inherently variable biological systems, with inherently variable materials and processes, leading to inherently variable drug products. While this heterogeneity is influenced by many factors, such as culture conditions, purification and final storage conditions, the first and most influential is the choice of expression system: the cell that will ultimately become the mini-factory for the biopharmaceutical.

A typical commercial biopharmaceutical manufacturing process begins with a single selected clone from a newly created engineered cell line. This single cell is cultured to increase its number and the resulting culture is aliquoted into small amounts and frozen to form what is known as the master cell bank. From a single thawed vial of the master cell bank, the cell line is cultured and aliquoted and frozen again to form what is known as the working cell bank. It is from a single thawed vial of the working cell bank that a single batch of biopharmaceutical is produced. The starter culture is used to inoculate a series of larger seed culture volumes, which are in turn used to inoculate the production-scale bioreactor, where the protein of interest will be produced (Walsh 2003).

Given that each biopharmaceutical thus starts with a single genetically engineered cell, the choice of host organism in the expression system is a critical decision in the development of the biopharmaceutical as the host organism's genetic background provides many of the key features of the expression system likely to impact on its commercial manufacture. These features include growth characteristics such as the specific growth rate, the maximum achievable cell density, the nutritional and hence media requirements, the robustness at the cellular and genetic level, the control of protein assembly and degradation, the secretion capacity, the potential amount of endotoxin produced (only a factor in Gram negative bacterial systems), and finally, the ability to perform post-translational modifications (Meyer & Schmidhalter 2012):

For commercial manufacture, the ideal host organism for an expression system will have a high specific growth rate (the higher the specific growth rate, the less time to achieve the desired mass of cells in the production bioreactor), be able to achieve high densities in culture (more cells

equals more product produced per liter of culture) and have simple nutritional needs so as to minimize the cost of media preparations. It will be robust at the cellular level to withstand a wide range of temperature, pH, osmolality, dissolved oxygen, and shear stress conditions and at the genetic level to be able to undergo multiple subculture divisions without losing the exogenous DNA for the protein of interest. It should secrete the final product into the medium and produce little or no endotoxin, simplifying harvest and downstream purification methods. It should have the ability to localize or chaperone the protein of interest away from exogenous proteases or cellular conditions that would cause the degradation of the protein prior to being harvested. Finally, it should have the ability to perform all required post-translational modifications necessary for the desired clinical function of the protein of interest (Meyer & Schmidhalter 2012).

Unfortunately, there is currently no universal expression system able to meet all the ideal requirements for all types of therapeutic proteins; usually, there is a trade off between cost, efficiency, safety, and the physiological limits of the host organism. Ultimately, regulatory requirements and the desired clinical function of the therapeutic protein will determine the optimal choice of host organism for the expression system.

Chapter 2: Literature Review

Post Translation Modifications

Of all the ideal requirements of an expression system for the production of biopharmaceuticals for human use, the requirement that has had the most impact on the final product and hence on the choice of expression system used is the ability of the host organism to perform post-translation modifications (Walsh 2014).

Co-translational (CTM) or post-translational modifications (PTM) refer to covalent modifications of a protein either during or after synthesis from the RNA transcript. There are hundreds of known CTM/PTMs: although the human genome contains only approximately 21,000 genes, there are estimated to be around 1-2 million different proteins in the human proteome due to multiple protein variants from the same gene differentiated by these modifications (Jefferis 2016).

These modifications can affect the structure, function, stability, and safety of a protein: conformational changes to the protein can distort binding sites and functional domains and expose residues that would normally be hidden inside the tertiary or quaternary structure of the protein, promoting aggregation, and potentially triggering immune reactions. Even naturally occurring PTMs on endogenous proteins, especially in the context of disease states, can result in immune reactions, thought to be triggered by the exposure of the patient to new epitopes, causing the generation of novel antibody specificities, triggering an autoimmune reaction. Therefore, for any recombinant protein, any PTMs that differ relative to the native protein have the potential to cause an immune reaction and should be evaluated (Kuriakose, Chirmule & Nair 2016).

Given all the potential implications of post translation modifications to the function and safety of a biopharmaceutical, it is clear why regulatory agencies require that a manufacturer be able to characterize these PTMs to allow the determination of appropriate specifications for 'identity, purity, potency, quantity and safety' (European Medicines Agency 1999).

In the context of currently available therapeutic proteins, the most commonly encountered post translation modifications are carboxylation and hydroxylation, amidation and sulfation, disulfide bond formation and proteolytic processing, and finally and most commonly, glycosylation (Walsh 2010).

Carboxylation involves the conversion of target glutamate residues in the protein, by carboxylase enzymes, into gamma-carboxyglutamate. Hydroxylation involves the conversion of target aspartate residues in the protein, by hydroxylase enzymes, to beta-hydroxyaspartate. These PTMs are important in the binding of calcium ions and are often key and in some cases, essential, to the proper function of the protein. They are especially common in certain blood clotting factors, such as factors VII, IX and X, which are important in the treatment of hemophilia (Walsh 2010).

Sulfation involves the attachment of a sulphate group (SO^{3-}) by sulfotransferase enzymes to the oxygen atom of target tyrosine, serine or threonine residues in the protein and is especially associated with secretory and membrane proteins of which blood coagulation factors are another commonly associated biopharmaceutical class. Sulfation is mainly thought to affect protein-protein interactions but its absence does not always completely reduce protein activity or interactions. While over 90% of native human factor IX molecules are sulfated, this is true of only 15% of approved recombinant factor IX, with apparently little effect on clinical function. However, hirudin, a powerful anticoagulant derived from leeches, displays a ten fold stronger affinity for thrombin when it is sulfated versus when it is not (Walsh 2010). Sulfation also plays a role in the interaction of factor VIII with its plasma carrier protein, von Willebrand's factor: people who inherit a factor VIII mutation involving the swap of a tyrosine residue to a phenylalanine residue, which results in lower sulfation, often show symptoms of mild hemophilia (Jefferis 2016).

Amidation involves the replacement of the C-terminal carboxyl group of a protein with an amide ($CONH_2$) group. While the exact biological function of this PTM has yet to be discovered, it has been known to have an effect on the stability and function of at least one approved biopharmaceutical: recombinant salmon calcitonin. Calcitonin is a hormone that regulates blood serum calcium and phosphate levels and the salmon-derived version of the protein has been

found to be one hundred times more potent than its human homologue. It is used in the treatment of Paget's disease (where bone is formed and reformed faster than normal), and hypercalcaemia caused by malignant cancers (Walsh 2010).

Disulphide bonds are an oxidation reaction between two sulphur-containing thiol groups of cysteine residues. They are an important part of maintaining the tertiary and, in the case of polypeptide proteins like antibodies, quaternary structure of the protein. The standard structure of an human IgG protein is two 'light' chain and two 'heavy' chain peptides. These are joined together with twelve intrachain and four interchain disulphide bonds, with the interchain bonds being clustered at the 'hinge' region of the heavy chains (Jefferis 2016).

Both prokaryotic and eukaryotic cells maintain a reduced state in the cytoplasm so disulphide bonds are not typical in cytoplasmic proteins. Eukaryotic cells maintain the oxidative environment necessary for disulphide bond formation inside the endoplasmic reticulum, an organelle that is not present in prokaryotic cells. Formation of disulphide bonds in certain prokaryotic cells takes place in the periplasmic space between the inner cell membrane and the outer membrane of the cell wall in Gram-negative bacteria (Walsh 2010).

Glycosylation, the covalent attachment of carbohydrate molecules (glycans) to the surface of a protein, is by far the most common and complex of all PTMs associated with biopharmaceuticals, with almost half of all licensed (Kuriakose, Chirmule & Nair 2016) and 11 of the top 20 best selling products (Zhang et al. 2016) exhibiting this PTM. Glycosylation can impact a wide variety of protein attributes including yield, folding and trafficking within the cell, bioactivity, solubility, stability against proteolytic degradation, immunogenicity and clearance rate from circulation (Walsh 2010).

The complexity of glycosylation derives from the fact that, unlike the amino acid sequence of a protein, it is not template driven. It is instead driven by enzymes and can affected by many factors including the enzymes present in the cell line in which it is expressed, epigenetic factors and the external environment during the development of the protein. Furthermore, the

architecture of the carbohydrate molecules is not linear; the structures can branch off in many different ways, increasing the complexity possible (Zhang et al. 2016).

Given all the potential implications for effects to the function and safety of a drug, it is clear why regulatory agencies expect companies to fully characterize any PTMs present on their products. While the vast amount of work involved seems daunting, the benefits of a well characterized PTMs (besides the obvious above-mentioned function and safety) includes the ability to engineer or refine the products and/or processes to improve the clinical effectiveness or safety profile. The earlier in the development timeline this characterization can be done the better, as it permits early design of the protein, choice of expression system, and clone selection which will have the greatest effect on post translation modifications (Kaltashov et al. 2012).

For example, recombinant erythropoietin lacking the N-glycosylation of the native protein was found to have a higher activity in vitro, but a significantly reduced activity in vivo, due to more rapid clearance in the body as a result of the missing glycans. By engineering the recombinant protein to have additional occupied N-glycosylation consensus sites compared to the native form (five instead of three) the half-life of the product in the blood was increased (Walsh 2010). Another example: due to the known tendency of methionine residues to oxidize, antibody sequences with this residue in their complementarity-determining regions are often excluded as drug candidates (Jefferis 2016).

Current Status of the Biopharmaceutical Industry

The current biopharmaceutical industry is a $200 billion dollar a year industry, and is growing rapidly, a bright spot in what had been some bleak years for more traditional drug pipelines (Wright 2015). As of 2014, there were 246 approved products in the United States (US) and European Union (EU) markets, 166 with distinct active ingredients (Walsh 2014). The most common class of biotherapeutic is monoclonal antibodies (mAbs), mainly targeting inflammatory and/or autoimmune conditions, followed by recombinant insulins. Other common indications include cancer, hemophilia, metabolic disorders, and vaccines against infectious diseases (Walsh 2014).

Two thirds of this $200 billion, representing 45% of all approved products, is generated from products manufactured in genetically engineering non-human mammalian cell lines, with the undisputed 'workhorse of the industry' being the Chinese Hamster Ovary (CHO) expression system: in 2017, six of the top ten best selling drugs (including both traditional pharmaceuticals and biopharmaceuticals) were recombinant proteins produced in CHO cells (Brown et al. 2017).

Microbial expression systems, still dominated by the first ever approved biopharmaceutical expression system, *E. coli*, account for second most common expression system with 40% of all recombinant products, while yeast, human, transgenic animals, insect and plants (in decreasing order of number of approvals) make up the remaining 15% (Walsh 2014) (Xu & Zhang 2014). A comparison of the advantages and disadvantages of each expression system, with reference to the above-mentioned ideal attributes, will provide explanation for the current status of each platform in the industry.

Mammalian (Non-human) Expression Systems

The year 1987 marked the approval of the first biopharmaceutical, tissue plasminogen activator, produced in a mammalian expression system: Chinese Hamster Ovary (CHO) cells (Walsh 2003). Since then, the platform has gone on to become the predominant expression system of choice: from 1982 to 2014, 35.5% of approved protein therapeutics were made using CHO cells (Brown et al. 2017)

Chinese hamster (*Cricetulus griseus),* a family of rodents native to the deserts of Mongolia and China, were first used in laboratory research in the 1920s for the typing of pneumonia bacteria. As natural carriers of the parasite *Leishmania*, they also became a popular tool in epidemiological research, and potentially other less savoury applications: literally smuggled into the United States in 1948 by Drs. C.H. Hu and R. Watson, the pair were later accused of war crimes by the Germ Warfare Commission of China, who in the escalating Cold War tensions of the era, feared that they would be used as agents of biological warfare by infecting them with diseases such cholera or plague and then parachuting them over Manchuria (Jayapal et al. 2007).

Cells were first isolated from a hamster ovary and established in culture by Dr. Theodore Puck in 1957. They soon proved themselves to be very resilient to *in vitro* cultivation conditions, with a relatively fast generation time. In addition, because of their low chromosome number (normal diploid cells have 22 chromosomes) and the fact that gene inactivation makes them functionally hemizygous for many genes, CHO cells were often used in experiments to isolate and characterize mammalian cell mutants. This work led to the discovery of a variety of different auxotrophs (inability to synthesize a compound needed for growth) such as dihydrofolate reductase (DHFR) that would 'facilitate their migration from the laboratory bench to industrial reactors' (Jayapal et al. 2007).

There are now several established lineages of CHO cells used in the production of biopharmaceuticals: CHO-K1 (DHFR+/+), DUKX-B11 (DHFR-), CHO-DG44 (DHFR-/-), CHO-K1SV (GS-KO) etc. The first biopharmaceutical was manufactured in the DUKX-B11 line, which was originally isolated from the CHO-K1 line; the DUKX-B11 line carried a deleted

DHFR locus and a mutation on the other, which made them unable to reduce folate to tetrahydrofolate, an essential precursor for the synthesis of purine, thymidylate, glycine and serine. When cells from the line are transfected with a vector containing the DHFR gene with the gene of interest, selection of those cells that have taken up the vector can be carried out by growing the cells in medium deficient in hypoxanthine and thymidine (Brown et al. 2017).

Another commonly used CHO cell line, CHO-DG44, was derived from a different mutant (CHO-Mtx-RIII) than DUKX-B11 yet using the same approach. However in this case, the new line carried a deletion of the DHFR mutation on both loci. More recently, another cell line has been derived from the CHO-K1 line that utilizes a different metabolic selection/amplification system: the glutamine synthetase (GS)/methionine sulfoximine (MSX). These cells have been engineered via zinc finger nuclease (ZFN) technology to have a knockout of the GS gene on both loci. Cells are transfected with a vector containing the gene of interest and the GS gene and grown in media deficient in glutamine with added GS-inhibitor, MSX. (Brown et al. 2017).

The CHO genome of the ancestral CHO-K1 cell line was first published in 2011 (Xun et al. 2011). Genomic studies since then of the different CHO cell lines have demonstrated a large amount of heterogeneity between them: of the original 22, the CHO-K1 and DG44 populations contained a modal chromosome number of 20 and 13 differed as a result of various deletions and rearrangements. This genetic instability, while underpinning their usefulness in cell line development (possibility of variants with unusual phenotypic traits that may aid their growth in large scale culture such as the ability to grow in suspension, high densities, high protein productivity etc.) also means that they are prone to genetic drift, which can result in production instability or altered performance over repeated subculture (Brown et al. 2017).

In both the DHFR/MTX and GS/MSX amplification technology systems, the integration of the vector into the CHO cell genome is random, resulting in a highly heterogeneous clone population, with varying levels of recombinant protein production and genetic stability; in addition, each clone will react differently to different media and culture condition. This is due to the fact that the chromosomal surroundings of the vector will exert a strong influence on the

promoter, which in turn will affect the transcription rate of the gene of interest. (Lai, Yen & Ng 2013).

Increasing concentrations of the selection drug are used to result in gene amplification, resulting in greater amounts of recombinant protein being produced but also increasing the resulting heterogeneity of the population. This is because when mammalian cells are exposed to an environment or stressor that required an increase in the translation of a normal protein in order to survive, a benefit or advantage is conferred on cells that have an increased copy number of the gene in question. The increase in copy number as a result of selective pressure is known as gene amplification and involves large regions of chromosome, not only selected genes (Kingston et al. 1993).

The process is highly inefficient: high producing clones occur rarely and therefore thousands of clones have to be screened to obtain a set of production clones. Repeat dilutions and screenings are performed until a population derived from a single clone is produced. Such work is very time, labour, and capital intensive: the creation of a new cell line for recombinant protein production using the traditional DHFR amplification system can take six to twelve months (Lai, Yen & Ng 2013). Large efforts in research have gone into making this process more efficient.

For faster screening of clones, several automated systems have been developed to speed up some aspects of the clone screening process, such as fluorescence-activated cell sorting (FACS), and the ClonePix and Cell Xpress systems (Noh, Sathyamurthy & Lee 2013).

FACS systems involve the binding of antibodies conjugated to fluorescent molecules to the secreted protein of interest on the cell surface. Cells enter the sorter and a laser beam causes the bound fluorescent molecules to emit light that is detected by an optical detector. The machine places a charge on the droplets containing cells exhibiting a high level of fluorescence to sort them into collection tubes or well plates (Lai, Yen & Ng 2013).

However, as the signal is dependent on the protein of interest remaining associated with the protein of interest, the system is more suited to screening for clones that do not secrete the

recombinant protein, which is not ideal as secretion of the protein of interest into the medium is preferred in large scale manufacturing for ease of harvesting. There are different strategies developed to optimize the FACS system to search for high producing clones that secrete the recombinant protein. In the gel microdrop secretion assay, the cell is coated in an agarose matrix coated with biotin. Primary antibody is added to the drop to bind the secreted protein. Secondary biotinylated antibody captures the primary antibody, and the complex binds to a streptavidin bridge immobilized on the agarose matrix for scanning by the FACS. The cold capture method relies on the hypothesis that the amount of protein that temporarily associates with the cell surface must be correlated to the total amount of protein being secreted by the cell. Cells are sorted at very cold temperatures of 0-4 degrees Celsius to help keep the secreted protein from diffusing away from the cell. Finally, an indirect screening method uses the selection marker as the basis for screening: fluorescein isothiocyanate-labeled methotrexate binds to the intracellular DHFR selection marker (Lai, Yen & Ng 2013).

The ClonePix automated colony picker system involves cultivating single cells in a semi-solid media, which allows the formation of individual colonies and helps to trap the secreted protein in close proximity to it. Labeled antibody, already present in the media, binds to the secreted protein and precipitates around the colony in a fluorescent halo. The system is capable of screening 10,000 clones within an hour, and is sensitive enough to isolate clones making up only 0.003% of the population (Lai, Yen & Ng 2013).

The Cell Xpress technology combines live cell imaging and laser-mediated manipulation to identify and purify the highest producer in a sample well. Multi-colour imaging is obtained through various fluorescence detection agents and the sample wells are coated with a capture matrix to trap the expressed recombinant protein (e.g. protein G is used for antibodies). Software analyzes the images, and through comparison of the different fluorescence levels, ranks the cells in the sample well. The laser is then directed to induce photomechanical lysis of low ranking clones until finally the highest producing cell is left in the well for subsequent transfer for further expansion (Lai, Yen & Ng 2013).

The quality, in addition to the quantity of recombinant protein produced must also be assessed during clone selection, such as protein mass, sequence, presence or absence of post translation modifications and glycosylation patterns. This is especially important in the development of biosimilars, where costly preclinical and clinical development stages can be avoided if the candidate protein can be shown to be highly similar to the reference product. While methods based on chromatographic or electrophoretic separation have been traditionally used, liquid chromatography coupled to mass spectrometry technologies now permit rapid, high-throughput analyses (Noh, Sathyamurthy & Lee 2013).

There have also been different expression vector engineering strategies to help the insertion of the gene of interest be a more site-directed and predictable. For example, the gene for the protein of interest and the selection marker gene can be transfected into cells on separate expression vectors but the reliability of protein of interest expression with the selection maker can be low; this is improved by placing both genes on the same expression vector but multiple promoters on the same vector can result in translational interference (Lai, Yen & Ng 2013).

One strategy used overcome this limitation is to place an RNA element called an internal ribosome entry site (IRES) between the gene for the protein of interest and the selection marker gene on the expression vector. An IRES site, typically found upstream of viral genes, allows for the initiation of translation by the ribosome independent of the usually required 5'cap recognition. This strategy has proven especially useful in the expression of heterodimic proteins such as monoclonal antibodies, where a balance expression of heavy and light chains is required (Lai, Yen & Ng 2013).

When selection pressure is high, i.e. a high concentration of either MTX or MSX, depending on the selection system used, any surviving clones will likely have high transcription level of the selection marker gene. However the high concentration of drug slows the growth rate of the cells, which can make an already long process even longer. A strategy to improve selection stringency at lower doses of selection drug is to impair the function of, or 'attenuate', the selection marker. This can be achieved in two ways. The first is to introduce a mutation into the selection marker gene itself. The second is to reduce its translation. Successful strategies to reduce translation

include codon de-optimization of the selection marker gene, use of weaker promoters, or to use an attenuated IRES (Lai, Yen & Ng 2013).

Another expression vector engineering strategy is to include elements in the expression vector that, wherever the expression vector ends up being inserted into the genome, help to maintain the integrated DNA in a transcriptionally active form. Matrix Attachment Regions (MARS) are sequences that act as attachment points along the DNA to help chromatin stay anchored to the nuclear matrix during the interphase stage of mitosis. They are also associated with the hyperacetylation of histones and the demethylation of DNA, thus permitting the transcriptional proteins to have access to the DNA. Finally, they serve as binding sites for transcription factors. Ubiquitous Chromatin Opening Elements (UCOE) have also been successfully used; these elements help to prevent the formation of heterochromatin, which is more tightly packaged in the nucleus, from euchromatin, thus keeping the DNA in the 'open' conformation necessary for translation. Including these elements in the expression vector helps to mitigate integration position effects, such as integration into a normally less transcriptionally active region of the genome (Lai, Yen & Ng 2013).

Ideally, the integration of the expression vector would be directed to specific areas of the genome, such as naturally transcriptionally active areas to promote expression or to avoid the disruption of necessary host cell genes. Site-specific recombinase enzymes, first discovered in bacteria (Cre) and yeast (Flp) cleave DNA at a specific sequence, usually 34-38 base pairs long, and then promote repair via, depending on the orientation of the recognition sites, insertion, deletion or inversion of a donor DNA fragment (Wang et al. 2011). These enzymes are used in the expression vector engineering strategy known as recombinase-mediated cassette exchange (RMCE), where first, a reporter gene flanked by the specific recombinase recognition sites is randomly integrated into the host cell line, using traditional random integration, a lengthy process which is one major drawback of the system. Clones are screened for high expression of the reporter gene and a single insertion site, thereby selecting for integration into a location in the host genome that has a naturally high rate of transcription (Lai, Yen & Ng 2013). Once the parental host cell line is established, the expression vector containing the gene of interest flanked by the recognition sequences is co-transfected with the recombinase, leading to homologous

recombination between the recognition sequences in the parental host cell genome and the expression vector, inserting the gene of interest into the specified target site on the genome (Noh, Sathyamurthy & Lee 2013).

Targeting vectors and recombinases have also been used in combination with an entirely separate artificial chromosome, the ACE system. The artificial chromosome is designed with multiple integration acceptor sites and natural centromeres and telomeres to enable replication without integration into the host genome. This reduces the chances that a random integration will cause a chromosomal aberration and increases the homogeneity of transformed clones. First the ACE is incorporated into the host cell. Next, the host cell is transfected with the ACE targeting vector and ACE integrase enzyme. The promoterless selection marker in the targeting vector is activated only when incorporated into the ACE: the integration sites are designed to be just downstream of a SV40 promotoer, hence only cells that have undergone correct recombination of the vector into the ACE, not just those that have taken up the targeting vector, will survive the selection pressure. The gene of interest is incorporated via the integrase into the integration sites on the ACE and because there are multiple insertion sites, consecutive transfections can be performed to achieve a high gene copy number without gene amplification (Lai, Yen & Ng 2013).

Additional site-specific integration strategies have benefited from the discovery of site-specific meganucleases such as zinc finger nuclease (ZFN) and transcriptional activator-like effector nucleases (TALEN), which are proteins that can be engineered to recognize and cleave specific DNA sequences. ZFNs work by fusing different types and numbers of binding domains that target a 3 base pair sequence of the target DNA with a nuclease domain. By combining different binding domains, a specific site, from 18 to 36 base pairs, on the genome can be targeted. The TALEN binding domain is an array of repeating units, each of which targets a specific DNA nucleotide, depending on the amino acid residue present at two specific, highly conserved positions. The endonuclease domain of the FokI restriction enzyme is often used in ZFN and TALEN systems since it does not have its own specific cleavage site and requires dimerization for its cleavage domain to function. This means that a pair of ZFN or TALEN binding domains targeting adjacent sequences can position two FokI domains in close proximity to each other, allowing the cleavage of the DNA at the targeted sequence (Lai, Yen & Ng 2013). TALENs are

more recently developed than ZFN and have the advantage of being less toxic and more specific in its binding (Noh, Sathyamurthy & Lee 2013).

The ZFN-FokI system was recently used create a line of CHO cells with three gene knockouts: dihydrofolate reductase (DHFR), glutamine synthase (GS) and alpha-1,6-fucosyltransferase8 (FUT8). The DHFR and GS knockouts allow for use of the DHFR/MTX or GS/MSX selections systems. The FUT8 gene allows for the transfer of fucose carbohydrate residues onto glycans; a CHO line with a FUT8 knockout allows for the production of monoclonal antibodies that lack the fucose residue on their Fc regions, a PTM modification that has been linked to increased antibody dependent cell mediated cytotoxicity function, which depending on the method of action of the therapeutic monoclonal antibody can result in higher treatment efficacy (Lai, Yen & Ng 2013).

Despite the long process necessary to achieve the creation of a new high producing cell line, there are many advantages to using CHO cells that have helped assure their current dominance in biopharmaceutical production. First is their ability to grow to high densities in suspension versus adherent culture. This permits manufacturers to build larger bioreactors, able to produce kilograms to tons of product. Second, they are able to adapt to synthetic environments, such as serum-free media. Besides their increased cost effectiveness, such chemically defined, serum-free media formulations are becoming the standard for regulators, as it reduces or eliminates concerns with potential quality and safety issues (viral, prion and mycoplasma contamination as well as animal spongiform encephalopathy transmission) regarding raw materials of animal origin (Brown et al. 2017).

Another safety advantage is the fact that CHO cells are also inherently incompatible with human pathogens, making them a very safe expression system: a study in 1989 found that of 44 different human pathogenic viruses tested, including measles, herpes, HIV, influenza and polio, the majority do not replicate in CHO cells (Jayapal et al. 2007) and completion of the CHO-K1 genome confirmed that this is due to the fact that many viral entry genes, while present in the CHO genome, are not expressed (Xun et al. 2011). More than thirty years of experience and over one hundred different approvals with no serious safety incidents have built up a lot of confidence

in the safety and reliability of the platform and therefore new regulatory approvals using the expression system are usually rapid (Brown et al. 2017).

Relative to other expression systems, mammalian systems have two major drawbacks: first, they have a much slower doubling time, 24 hours versus 20 minutes for bacteria or 2 hours for yeast (Brown et al. 2017), and second, they produce protein yields that are 10 to 100 times smaller; for the first generation of biopharmaceuticals produced in mammalian expression systems, yields of 100 mg/L were considered reasonable (Jayapal et al. 2007). In thirty years of use however, there have been many efforts to improve this. Improvements in clone selection, media and culture condition optimization, bioreactor design, and longer cultures times made possible by the use of fed batch and perfusion processes mean that yields of 10-15 g/L are possible today (Brown et al. 2017).

However, despite all the above advantages and disadvantages, the ideal expression system attribute that has been the main reason for the dominance of CHO and other non-human mammalian expression systems can be explained by their ability to perform the higher level post translational modifications, such as glycosylation and disulfide bond formation, that are required for the main classes of therapeutic proteins, that are not possible or greatly altered in microbial and other lower eukaryotic platforms (Brown et al. 2017).

While not all biopharmaceuticals are glycosylated (insulins, vaccines) the most common and most profitable type of biopharmaceutical are therapeutic glycoproteins such as antibodies, erythropoietin, and blood factors, precluding their manufacture in expression systems where glycosylation cannot be performed (Walsh 2014).

CHO cells are particularly well suited to producing therapeutic glycoproteins with human-like glycosylation patterns: sequencing of the ancestral CHO-K1 cell line showed that of the 300 genes associated with glycosylation in humans, 99% were present in the CHO cell genome, although only 53% of those were expressed (Xun et al. 2011). The noted types of glycoslylation that CHO cells are unable to perform are alpha-2,6-sialylation and alpha-1,3/4 fucosylation, as they lack the glycosyltransferase enzymes that perform these highly specific linkages (Lalonde &

Durocher 2017). CHO cells, as do most non-human mammalian species, also produce glycans that do not occur in humans, N-glycolneuraminic acid (Neu5Gc) and galactose-alpha-1,3-galactose (alpha-Gal) which can potentially trigger immune responses in humans (Ghaderi et al. 2012); however CHO cells produce very low levels (less than 0.2% and less than 2%, respectively) of these glycans compared to murine (mouse) mammalian expression systems, another reason for their predominance in the industry (Lalonde & Durocher 2017).

Alternative non-human mammalian hosts to the predominant CHO cells include baby hamster kidney (BHK) and mouse myeloma cells (NS0, Sp2/0). BHK cells, mostly used for the production of veterinary vaccines such as foot and mouth disease and rabies, are currently used in the manufacture of only two marketed recombinant glycoproteins, both blood clotting factors: NovoSeven (factor VIIa) and Kogenate/Kovaltry (factor VIII). Blood clotting factors are notoriously hard therapeutic proteins to manufacture, as they are large and require a large amount of PTMs to properly function, such as abundant glycosylation and sulfation. Sulfation being one of the PTMs, along with fucosylation and GalNac transferases, found to be significantly reduced in CHO cells (Xun et al. 2011), explains the use of the alternative expression system. Murine myeloma cells are derived from mouse tumour cells that no longer produce their original immunoglobulins. They do however retain the cellular machinery required to produce immunoglobulin and this is where they are mainly use in the industry: the production of monoclonal antibodies, such as Cetuximab (Erbitux) and Palivizumab (Synagis) (Lalonde & Durocher 2017).

Prokaryotic Expression Systems

E. coli bacteria are named after the German pediatrician Theodor Escherich, who called them 'bacterium coli communale'. He could have had no idea then how common they would become: for more than thirty years they have been to microbial expression systems what CHO cells are to mammalian expression systems: the undisputed workhorse of the biopharmaceutical industry (Brown et al. 2017). While CHO cells may dominate in number of products, microbial systems dominate in terms of pure protein produced: of 26.4 metric tonnes of protein produced in 2010, 68% (17.9 metric tonnes) was produced in microbial expression systems. Amongst bacterial expression systems, *E. coli* is so dominant that only two other species of bacteria are currently in use for the manufacture of biotherapeutics, both also Gram-negative species. *Vibrio cholera* is used for the manufacture of the cholera toxin subunit B for the product Dukoral, an oral cholera vaccine. *Bordetella pertussis* is used for the manufacture of recombinant pertussis toxin for the product Triacelluvax, an acellular Diptheria-Tetanus-Pertussis (DTP) vaccine (Walsh 2014).

One reason for the dominance of *E. coli* amongst microbial expression systems is that it and its phages have long been model organisms in the study of molecular biology, especially those aspects related to understanding gene function and regulation. The isolation and purification of restriction enzymes by Werner Arver in 1968 represents one such milestone that helped to rapidly expand the molecular biology toolbox, permitting the first biopharmaceutical product to appear in 1982, a remarkably fast development considering the normal lengthy approval procedure for therapeutics; the long history of study and multitude of fundamental discoveries made using *E. coli* also helped with regulatory and safety concerns (Meyer & Schmidhalter 2012).

Microbial systems have many advantages that should make them the undisputed champion expression system, even over mammalian cells. They are easy to genetically manipulate, they grow very fast (doubling time of 20 minutes), in high quantities, in very inexpensive simple media, using long-established and robust fermentation technology. In fact, if the desired therapeutic protein is small and does not require human-like glycosylation for its clinical function, then microbial systems are the clear choice for cost-effective manufacture, a fact that is borne out in the industry by considering the types of therapeutic proteins manufactured in

microbial systems: mainly insulins, vaccine components and smaller cytokines like interferons (Walsh 2014).

However, because prokaryotic systems lack the internal organelles and enzymes to perform post translational modifications such as disulphide bond formation and glycosylation, this has precluded their use in the manufacture of the most common type of therapeutic protein: monoclonal antibodies (Walsh 2014).

Antibodies, of which the most commonly used therapeutic class are immunoglobulin G (IgG), are large, multi-domain glycoproteins, naturally produced in the body by B-type immune cells. A mature IgG molecule consists of four different peptides, two 'light' chains, consisting of a variable and conserved region each, and two 'heavy' chains, consisting of a variable and three conserved regions each. The entire light chain and the heavy chain variable and first conserved region makes up what is known as the antigen binding fragment (Fab) while the remaining conserved domains of the heavy chain make up what is known as the crystallizable fragment. The Fab domain will vary between different antibodies and are what give the molecule its specific binding ability to its target antigen. The Fc region is conserved within the class of IgG antibodies and while it does not interact with the antigen, it is important for stabilizing the structure of the antibody, and for inducing the effector functions of the antibody such as activation of the complement system and antibody dependent cell mediated cytotoxicity. All IgG antibodies have a glycan attached to an asparagine residue at position 297 of the heavy chains. About 20% of IgG have additional glycan attachment sites molecules. (Lee & Jeong 2015).

Correct glycosylation is important for many features of the final protein. In gonadotropic hormones it is linked to correct protein folding and secretion. It has been shown to be essential in the targeting and trafficking of erythropoietin in the cell. It can also affect the binding of the protein to its intended receptor or ligand. For example, replacement enzymes for lysosomal disorders require the glycans to have exposed mannose residues to facilitate uptake into lysosomes. Exposed mannose receptors can however make a protein more likely to be cleared from the body, as they are targeted by mannose receptors in liver cells. Therefore, glycans with terminal sialic residues are important for the half-life of the product in the body. Finally,

glycosylation has been known to affect the stability of the protein: removal of the glycan from Cerezyme, one of the aforementioned replacement enzymes for lysosomal disorders, causes the protein to aggregate and precipitate (Walsh 2010).

Any microbial system hoping to challenge the dominant mammalian platforms in the production of glycosylated biopharmaceuticals should aim to overcome the disadvantages of existing mammalian platforms which are the naturally occurring heterogeneity of glycosylation patterns produced and the currently limited possibilities to tailor glycosylation patterns to improve performance of the biopharmaceutical *in vivo*. Challenges to creating tailored homogenous and human like glycosylation in microbial systems also include the need to ensure that any genetic changes introduced do not affect the productivity and robustness of the platform, to also consider the less elucidated role of O-glycosylation and its possible immunogenic potential in humans, to achieve the same or improved glycosylation and secretion efficiency, to eliminate or reduce unwanted proteolytic activity, to have the ability to produce hetero-monomeric proteins and disulphide bonds, and the ability to minimize N-terminal variability (Meyer & Schmidhalter 2012).

The discovery of prokaryotic N-linked glycosylation genes in the Archea domain led to the exciting opportunity for 'glycoengineering' *E. coli* by introducing the ability to perform glycosylation into more established expression systems and in 2002, the successful transfer of N-glycosylation machinery from *Campylobacter jejuni* into *E. coli* was performed. However, there have been issues that have prevented this from being the major breakthrough it should be. First, the glycan produced in *C. jejuni* is immunogenic to humans. Second, the glycan is linked to the asparagine residue through an unusual sugar residue, although efforts to adapt this to the more usual N-acetylglucosamin-asparagine linkage have also been successful (Meyer & Schmidhalter 2012).

By using the *PglB* oligosaccharyltransferase (OST) from *C. jejuni* in conjunction with four different glycosyltransferases from *S. cerevisiae,* proteins with $Man_3GlcNac$ glycans, the conserved base from which complex human glycans are extended, were produced in *E. coli.* Unfortunately, despite optimization, only 1% of the target protein was glycosylated and those

that were glycosylated showed a heterologous mix of structures. Other experiments have managed to improve homogeneity of the glycan but efficiency is still too low for practical use in commercial systems: only 3% of the target protein was glycosylated. Thus, while the basic proof of concept for the production of N-linked glycosylation in microbial systems has been established, there are still additional functions, like terminal sialylation, that still need to be established in prokaryotic systems and the low glycosylation efficiencies at these early stages shows that much more optimization is required. It is also not known if the 'flippase' enzyme responsible for flipping the growing glycan from the cytoplasm into the periplasmic space can handle the larger human type glycans or whether further glycan development will have to occur in the periplasmic space. Finally, the prokaryotic OST is more specific that its eukaryotic homologes in its consensus sequence, which limits the range of proteins that can be fully glycosylated (Chinyere Anyaogu & Hasbro Mortensen 2015).

One disadvantage of using Gram-negative bacteria, such as *E. coli,* is the fact that the lipopolysaccharides that make up the main portion of their outer cell membranes, known as endotoxins, induces a strong immune reaction when injected into patients. While some recombinant proteins are secreted into the extracellular space, the vast majority of recombinant proteins produced in E. coli are retained in the cytoplasm or periplasmic space, requiring that the cells be disrupted to release their contents as the first stage of downstream purification. This has unavoidable complication of releasing large amounts of endotoxin that must be cleared in later downstream processing steps (Meyer & Schmidhalter 2012). Another disadvantage of recombinant proteins being retained in the cytoplasm or periplasmic space is that they can aggregate and form inclusion bodies, which necessitates downstream processes to solubilize and/or refold the recombinant protein. By contrast, Gram-positive bacteria do not have an outer membrane and are able to secrete recombinant protein into the media (Meyer & Schmidhalter 2012).

To date however, Gram-positive bacteria, while extensively used for the industrial production of amino acids for more than fifty years, have surprisingly not been used in the manufacture of therapeutic proteins for human use. *Corynebacterium glutamicum* is one alternative candidate being seriously considered. It is non-pathogenic and has 'Generally Recognized As Safe' status.

It has the ability to secrete properly folded and functional proteins directly into the culture medium, and secretes minimal amounts of its own proteases and host cell proteins, making it an ideal expression system for protease-sensitive products. However, compared to *E. coli,* it currently has a much lower transformation efficiency and there are only a few expression vectors available although there are ongoing research efforts to overcome these disadvantages. A commercialized expression system, the CORYNEX system, has been developed by the Japanese company Ajinomoto and a number of different recombinant proteins, including human epidermal growth factor and antibody fragments have been successfully produced in *C. glutamicum* (Liu et al. 2015). The first approved biopharmaceutical in a Gram-positive expression system may not be far off as in 2016, the San Diego based Ajinomoto Althea, Ajinomoto's contract manufacturing subsidiary, announced the start of Phase I clinical trials for their first biopharmaceutical candidate (Ajinomoto Althea 2016).

Although the commercial production of fully assembled and glycosylated antibodies has not been performed in microbial systems, efforts to create therapeutics using various antibody fragments, such as single chain variable fragments (scFv), consisting of only the heavy and light chain of the variable region, and antigen binding fragments (Fab), discussed above, have been more successful as these fragments, unlike the conserved Fc region, are not glycosylated. There are currently two approved biopharmaceuticals consisting of Fabs produced in *E. coli*: Lucentis (ranibizumab), for the treatment of wet macular degeneration, and ReoPro (Abciximab), a platelet aggregation inhibitor (Lee & Jeong 2015). Besides the advantage of being able to be manufactured in cheaper microbial systems, these smaller antibody fragments can offer the advantage of being able to reach tissues and targets that full-sized antibodies cannot. However, potential drawbacks include the lack of Fc effector functions and reduced half-life in the blood precisely due to their smaller size, lack of glycosylation and lack of Fc-based antibody recycling (Walsh 2014).

In microbial expression systems it is often desired to target a recombinant proteins to the periplasmic space instead of letting them accumulate in the cytoplasm, as the periplasm provides the oxidative environment usually necessary for correct protein folding, the formation of disulphide bonds, and other post translational modifications. In order to target recombinant

proteins to the periplasmic space, the gene of interest is engineered to include an N-terminal 'signal peptide' that directs the immature protein to the Sec translocon, the most common secretory pathway in *E. coli*. Unfolded immature protein accumulating in the cytoplasm before translocation is still at risk for formation of aggregates and inclusion bodies. To reduce the time the unfolded protein spends in the cytoplasm, an alternative pathway, the signal-recognition particle (SRP)-dependent pathway can be used. In the SRP-dependent pathway, the ribosome is linked via the signal-recognition particle to the Sec translocon so that translation of the peptide occurs simultaneously with the translocation into the periplasmic space (Lee & Jeong 2015).

Additional strategies to achieve more efficient secretion and prevent aggregation in the cytoplasm include the co-expression of periplasmic foldases and molecular chaperone proteins. Chaperone proteins help proteins fold correctly into their native conformation, which is important both for protein function, but also reduces the likelihood of aggregation due to association of usually hidden hydrophobic amino acid residues. The chaperone proteins bind to these short sequences of exposed consecutive hydrophobic amino acids, which are usually flanked by basic residues. Chaperones can be divided into three classes depending on their function. The folding chaperones DnaK and GroEL actively drive conformational changes in their substrate proteins, using ATP as an energy source. Holding chaperones IbpA and IbpB associate with partially folded proteins until the folding chaperones are available to bind. Finally, disaggregating chaperone ClpB promotes the solubilization of already aggregated proteins (Berlec & Strukelj 2013).

The formation of disulphide bonds is important in the structure of antibodies and antibody fragments, yet the prokaryotic cytoplasm is a reduced environment, which is not suitable for their formation. Translocation to the periplasm, with its oxidizing environment, is required for disulphide bond formation. Once in the periplasm, bond formation is catalyzed by Dsb proteins, a series of thiol-disulphide-oxidoreductases. DsbA is a soluble periplasmic protein that oxidizes cysteine residues in target proteins and is later recycled itself by the periplasmic membrane protein DsbB. Incorrectly formed disulphide bonds can be reversed and rearranged by the disulphide bond isomerase DsbC (Berlec & Strukelj 2013).

Another strategy for the creation of disulphide bonds in prokaryotic cells is to engineer the oxidative environment of the cytoplasm. The cytoplasm has advantages over the periplasmic space because of its greater volume and because proteins do not have to be translocated across the periplasmic membrane; deletion of thioredoxin reductase and glutathione reductase genes in *E. coli* resulted in strains with an oxidative cytoplasm and were used to successfully produce functional antibody fragments (Lee & Jeong 2015).

While higher eukaryotic families such as mammals or insects tend to have relatively conserved physiological and metabolic profiles, conversely, microbial families show a wide diversity, offering opportunities to discover novel expression platforms with improved and/or novel characteristics. For example, protein aggregation can be a problem in the production any recombinant protein, but certain classes, such as membrane proteins, are more likely to aggregate when grown in large volumes in the cell. This aggregation is mainly driven by the interaction between exposed hydrophobic residues on the protein. Thermodynamic studies have shown that entropic forces are the main force in the interaction and therefore could be reduced at lower temperatures. Psychrophilic bacteria, able to grow at 4 degrees Celsius or even lower, could offer a solution to protein aggregation. *Pseudoalteromonas haloplanktis*, first isolated from Antarctic coastal seawater, is a representative example (Corchero et al. 2013).

It has already been used to advantage in the production of certain 'difficult proteins' such as human nerve growth factor, which requires a complicated post translational modification of three non-consecutive disulphide bonds and homodimerization. When grown in *E. coli* cells, the protein accumulates into inclusion bodies; when grown in *P. haloplanktis*, the protein is fully soluble, translocated to the periplasmic space, and accumulates in almost fully dimeric form. In fact, insoluble protein aggregates have never been found in recombinant *P. haloplanktis* (Corchero et al. 2013).

As the range of potential therapeutic proteins expands, not all of which are possible to grow in *E. coli*, and as genetic and glycoengineering efforts in *E. coli* begin to run up against the physiological limits of the expression system, expansion of the microbial expression system repertoire is long overdue.

Lower Eukaryotic Expression Systems

Bringing up third place in the biopharmaceutical expressions system race are eukaryotic microorganisms, in particular two species of yeast, *Saccharomyces cerevisiae* and *Pichia pastoris*, in order of frequency of use (Walsh 2014). The first biopharmaceutical created using yeast cells (*S. cerevisiae)* was approved for use in human in 1984, a recombinant Hepatitis B (HBV) antigen called Recombivax (Corchero et al. 2013). The first recombinant insulin in yeast came in 1987, marketed by Novo (now Novo Nordisk) to replace the company's previous product, an insulin enzymatically derived from porcine insulin (Nielsen 2013). Together therapeutic proteins produced in both types of yeast account for one sixth of all biopharmaceuticals produced for use in humans. They continue to play a particularly dominant role in the production of recombinant insulin and insulin analogs, in addition to human serum albumins, and vaccines (Walsh 2014).

Yeast are unicellular organisms like bacteria, and are among the simplest of eukaryotes. However, unlike bacteria, they have an internal cellular organization that is quite similar to higher eukaryotes; their genome is organized in a nucleus, and they have organelles such as mitochondria, the endoplasmic reticulum, Golgi apparatus, and secretory vesicles (Berlec & Strukelj 2013).

S. cerevisiae, also known as Baker's yeast, has been used in biotechnology by humans for centuries, most commonly in baking and brewing applications. As such, it has a designation of Generally Recognized As Safe (GRAS), which facilitates regulatory approvals. It was the first eukaryotic organisms to have its genome sequenced and there have been many investigations and manipulations of its physiology via molecular biological methods (Corchero et al. 2013). The strain has some of the beneficial characteristics of prokaryotic systems such as the ability to grow fast in relatively cheap and simple media, and it is easily genetically manipulated due to its preferential use of homologous DNA repair. It combines this with the advantages of mammalian systems such as the ability to perform proteolytic processing, folding, and disulfide bond formation. However, it is unable to grow to very high densities, has limited protein secretion, and exhibits excessive and irregular glycosylation patterns (Berlec & Strukelj 2013).

Optimization efforts in *S. cerevisiae* have focused on improving the secretory efficiency of the strain. Control of cultivation parameters such as pH and temperature has helped to increase cell densities and reduce protein degradation and strain engineering has focused on modifying, at the genetic level, crucial steps in the protein secretion pathway. Strategies include use of endoplasmic reticulum targeting sequences in the recombinant protein, the over-expression of endoplasmic reticulum chaperone and/or Golgi body trafficking proteins, and the creation of knock-out strains deficient in endogenous proteases. Unfortunately, most efforts have found that these targets need to be quite specific, depending on the particular protein being expressed (Berlec & Strukelj 2013)

P. pastoris, a methylotrophic (able to use methanol as its sole carbon source) yeast, is the second most commonly used strain for biopharmaceutical production. Like *S. cerevisiae*, it combines the benefits of prokaryotic and higher eukaryotic systems with a few extra advantages: unlike *S. cerevisiae*, it is able to grow to high cell densities and to secrete the protein of interest with a low level of excretion of its own host cell proteins, facilitating downstream processing (Berlec & Strukelj 2013). The strain also performs less extensive glycosylation than *S. cerevisiae* (Rashmanova et al. 2018). Methylotrophic yeast secrete protein more effectively because of the structure of their Golgi body, where proteins are prepped for secretion: their Golgi bodies are stacked adjacent to the endoplasmic reticulum (Matthews et al. 2017).

The ability to naturally perform some glycosylation is one area where eukaryotic expression systems have an advantage over prokaryotic expression systems. However, in yeast, the final glycan pattern produced is much simpler than that seen in mammalian and human expression systems, consisting of many terminal mannose residues. The initial glycan structure, formed in the endoplasmic reticulum, is the same in both yeast and humans: it arrives at the Golgi body as a $Man_8GlcNAc_2$ glycan. In humans, this structure is first trimmed to a Man_3,$GlcNAc_2$ before the addition of other types of carbohydrate residues, usually terminating in a sialic acid. In yeast however, the $Man_8GlcNAc_2$ intermediate glycan undergoes the addition of up to thirty additional mannose residues (Berlec & Strukelj 2013). This high mannose content on the exposed terminal

ends of the glycan causes proteins to be cleared from the blood stream in patients at a much faster rate and can also be a cause of immunogenic reactions (Ghaderi et al. 2012).

Enzymatic glycosylation, where the existing glycan is cleaved from the protein and the desired final glycan is transglycosylated onto the protein using a mutant enzyme-glycan conjugate is one method to overcome the issue of unwanted native glycan structures; however it is costly and does not scale well (Matthews et al. 2017)

Thus, another common optimization strategy has focused on 'humanizing' yeast glycosylation patterns and in the strain P. pastoris, efforts have been successful enough to engineer a fully humanized strain, capable of producing sialylated recombinant erythropoietin (Hamilton & Gerngross 2007). An expected first step was the elimination of the alpha-1,6 mannosyltransferase enzyme, responsible for the hypermannosylation that is characteristic of glycosylation pattern of the species. A temperature sensitive mutant of *S. cerevisiae* was noted to have less mannose elongation, and a mutation in the OCH1 gene was found to be responsible; elimination of the OCH1 and MN1 produced a strain that produced only $Man_8GlcNAc_2$ glycan, the optimal starting point for producing human-like glycosylation (Hamilton & Gerngross 2007).

The next breakthrough came when a library of fusion proteins consisting of alpha-1,2 mannosidase catalytic domains fused onto various yeast endoplasmic reticulum and Golgi localization signals were introduced into an OCH1 deficient strain of *P. pastoris,* resulting in strains capable of producing mostly $Man_5GlcNAc_2$. These strains were then transformed with a library of N-acetylglucosaminetransferase I also fused to yeast ER and Golgi localization signals. The resulting strain was capable of producing $GlcNacMan_5GlcNAc_2$ glycans. A repeat of the combinatorial library approach with N-acetylglucosaminetransferase II yielded a strain capable of producing proteins with $GlcNac_2Man_3GlcNAc_2$ glycans, the first instance of complex glycoprotein in yeast. Subsequently, the library approach was used with a triple fusion protein of beta 1,4-galactosyltransferase, UDP-galactose epimerase and yeast Golgi localization signal to yield the first synthesis of galactosylated protein in yeast (Hamilton & Gerngross 2007).

Sialylation, the transfer of sialic acid onto the galactose residues is often the last step in glycosylation in human glycoproteins. It was viewed as the final and greatest challenge in the humanization of yeast glycans, as it involved four independent enzymatic functions not known to be present in any wild type yeast. First, the enzymes necessary to convert endogenous UDP-GlcNac into CMP-sialic acid were introduced into the strain. Combined with expression of a CMP-sialic transporter and an alpha-2,6 sialyltransferase-localization signal fusion, the first production of complex sialylated glycoprotein in yeast was complete (Hamilton & Gerngross 2007).

While yeast are currently the dominant platform in lower eukaryotic expression systems efforts are ongoing to establish expression systems in many other species, in particular filamentous fungi, microalgae and protozoa.

The filamentous fungi species *Trichoderma reesei* is a soil based microbe already widely used in food applications, and thus has the beneficial Generally Recognized As Safe (GRAS) designation. It is able to use cellulose as its carbon source, and so can be grown on lower cost media sources such as agricultural or industrial waste streams. It also naturally produces glycans that are a suitable precursor to mammalian glycosylation, which should ease glycoengineering efforts to 'humanize' its glycosylation pathway. Finally, filamentous fungi are very efficient secretors of recombinant proteins, without also being high secretors of host cell proteins. While one drawback is the relative fraction of endogenous proteases secreted with the protein of interest, knock-out strains lacking these proteases have been shown to increase titers without also affecting cell growth (Matthews et al. 2017) (Corchero et al. 2013).

The microalgae species *Chlamydomona reinhardtii* also has a GRAS designation due to its widespread use in both human and animal feed production. As a photoautotroph, that is, able to perform photosynthesis, microalgae have some of the simplest media requirements of any expression system, able to grown on only light, water and basic micronutrients (Corchero et al. 2013) (Matthews et al. 2017).

The protozoan species *Tetrahymena thermophile*, is a ciliate microorganism found in fresh water sources, and is already available as a commercial expression system (Matthews et al. 2017). Cilian AG have demonstrated expression of human enzymes, vaccines, and antibodies with the CIPEX system and are currently developing both a flu vaccine and a digestive enzyme preparation for the treatment of exocrine pancreatic insufficiency, common in patient's with cystic fibrosis. The preparation is based on endogenous ciliate enzymes, in contrast to traditional preparations, which are harvested from animal slaughter waste (Cilian 2018).

Alternative eukaryotic systems have implications on downstream processes, which are usually the bottleneck in modern processes as yields increase. A cleaner harvest feed leads to reduced purification costs. Studies show that fungi and protozoa secrete a much lower amount of host cell proteins than their CHO counterparts. Additionally, the purity of the protein of interest is much higher for yeast and protozoa compared to raw CHO supernatant. Because the harvest pool is less complex for these alternative eukaryotic expression systems, it means that a wider variety of downstream purification technologies may be used, such as multimodal resins, or membrane chromatography. In addition, certain species of eukaryotes can be induced to flocculate out of solution, potentially leading to harvest pools clear enough to purify the protein of interest through crystallization, a process not feasible with current CHO-based processes (Matthews et al. 2017).

In addition, because mammalian viruses cannot propagate in these systems, the risk of adventitious agents posing a risk to humans is reduced in these systems. This means that viral clearance requirements and costs could be reduced. This has already been agreed with European regulators for the production of Jetrea (ocriplasmin), manufactured in *P. pastoris.* Since, no raw materials of animal or human origin are used in the manufacturing process, and *P. pastoris* has an established safety profile (ten products approved for use in humans), it was agreed that there was a reduced risk of contamination with viral adventitious agents: similar strategies could be employed for therapeutics made in alternative eukaryotic hosts (Matthews et al. 2017).

Human Expression Systems

If human-identical proteins are the ideal end goal of most biopharmaceutical processes, an obvious question is why not use actual human cells as the host cell for the expression system? The traditional answer has been safety and the conservative nature of both regulatory bodies and the industry itself.

It has been argued (Rose et al. 2005) that some of these concerns may be misguided or overblown; infection with human pathogens is likely to result in an easily detectable pathogenic effect in human cells, whereas the agent may be dormant in rodent cells. Furthermore, the risk of transmission of prion-based diseases for any new cell line is already strictly screened and managed through documentation requirements and downstream viral inactivation and clearance protocols already exist that may provide viral clearance greater than that currently employed for current CHO-based processes (Dumont et al. 2016).

Despite this, anticipating regulatory hurdles due to a lack of a species barrier to help mitigate the risk of contamination with adventitious and endogenous human pathogens, most manufacturers choose the trade off of using previously approved non-human mammalian expression systems, as these still have the ability to synthesize proteins that are very similar to those naturally occurring in humans (Swiech, Picanco-Castro & Covas 2012).

Yet similar is not identical; even naturally occurring PTMs on endogenous proteins, especially in the context of disease states, can result in immune reactions, thought to be triggered by the exposure of the patient to new epitopes, causing the generation of novel antibody specificities, triggering an autoimmune reaction. Therefore, for any recombinant protein, any PTMs that differ relative to the native protein have the potential to cause an immune reaction and should be evaluated (Kuriakose, Chirmule & Nair 2016).

It is known that non-human mammalian expression systems attach monosaccharide structures at the terminal, and thus most exposed, position on glycans that are not found in natural human glycans: these are Gal 1-3 Gal (alpha-Gal) N-glyconeuraminic acid (Neu5Gc). All humans

express circulating antibodies, in varying amounts, to these glycan structures, meaning there is an increased risk of an immunogenic reaction when patients are exposed to biopharmaceuticals that carry these non-human glycan epitopes. Titers of anti-alpha-Gal antibodies can reach as much as 1% of total circulating antibodies, while, in certain individuals, levels of anti-Neu5Gc antibodies can surpass those of antibodies directed against blood groups, another type of anti-glycan antibody, with its well known clinical implications (Ghaderi et al. 2012). There have been documented cases of severe anaphylactic reactions to these non-human epitopes. A well documented example is the monoclonal antibody Cetuximab (Erbitux), manufactured in murine myeloma Sp2/0 cells and indicated for the treatment of certain cancers; patients in parts of the United States who appeared to have higher pre-existing anti-alpha Gal antibodies were more likely to develop sensitivity reactions (Butler & Spearman 2014).

In human glycoproteins, the monosaccharide in the terminal, outermost position of the glycan is usually occupied with a class of monosaccharide called sialic acids. There are over 43 different versions of sialic acid, all with a conserved nine-carbon backbone and carboxylic group at the C1 position. Varying groups at positions 4,5,7,8 and 9 and the possibility of different α-ketosidic linkages at the C2 position generates this diversity. The transfer of these sialic acids onto the growing glycan chain is performed in the Golgi body by enzymes called sialytransferases, of which more than twenty have been discovered. The most common sialic acid is N-acetylneuraminc acid (Neu5Ac), which serves as the precursor for the biosynthesis of most other forms. In mammalian cells, the two most common sialic acids found at cell surfaces are Neu5Ac and Neu5Gc, which is formed by the hydroxylation of Neu5Ac by the enzyme CMP-N-acetylneuraminic acid hydroxylase (Cmah). The *Cmah* gene, while highly conserved from sea urchins to primates, was disrupted in humans approximately three million years ago due to a frameshift mutation; this mutation, present in all humans, was discovered by two separate teams in 1998 and represents the first known genetic difference between human and chimpanzees that could be directly linked to a change in phenotype (Ghaderi et al. 2012).

The low doses of early recombinant products meant that patients were exposed to only small levels of Neu5Gc from biopharmaceuticals. However, biotherapeutic treatments today are given in high milligram quantities per dose over long periods of time. This frequent exposure may

increase cases of immunological reactions against the biopharmaceutical. Indeed, studies show that recombinant coagulation factors have a higher incidence of causing the development of 'inhibitor' antibodies (antibodies against the drug itself) than from coagulation factors derived from plasma (Swiech, Picanco-Castro & Covas 2012).

The existence of antibodies against alpha-Gal has already been shown to induce the production of additional antibodies against the protein backbone of a glycoprotein displaying such epitopes. Mice deficient in alpha-1-3galactosyltransferase, the enzyme that transfers alpha-Gal monosaccharides onto glycans, were injected with alpha-Gal conjugated bovine serum albumin (BSA). The enzyme deficient mice generated new anti-BSA antibodies while wild-type mice did not. Furthermore, enzyme deficient mice with pre-existing anti-alpha-Gal antibodies showed an increased T-cell response to the BSA protein moiety itself. While yet to be demonstrated for the Neu5Gc epitope, it is thought that its presence on biopharmaceuticals produced in non-human mammalian expression systems may help to explain the development of inhibitor or 'anti-drug' antibodies that have been seen in various patients (Ghaderi et al. 2012).

Non-human mammalian expression systems also have limitations with other essential PTMs: γ-carboxylation and β-hydroxylation are known to be important in the correct function of proteins that help to regulate blood coagulation. Thought to mediate the binding of calcium, this PTM is carried out by specific carboxylase and hydroxylase enzymes, which convert target glutamine residues to γ-carboxyglutamate, aspartate residues to β-hydroxyaspartate, or asparagine residues to β-hydroxyasparigine. While some recombinant blood coagulation factors display less or no carboxylation/hydroxylation compared to endogenous proteins without sacrificing clinical efficacy, protein C, required the full complement of its nine carboxylation and single hydroxylation sites to be modified to be fully functional. As CHO cells were unable to carry out these PTMs at a cost efficient level, the use of a human expression cell line was required (Jefferis 2016).

To date, only two human cell lines have been developed for the commercial production of recombinant proteins; they are HEK 293 and HT-1080.

HEK 293 cells are easily grown in suspension serum-free culture, reproduce rapidly, and are efficient at protein production. Derived from human embryonic kidney cells transformed with adenovirus type 5 fragments almost thirty five years ago, the first approved biotherapeutic produced using the cell line was only approved in 2001, the above-mentioned protein C, marketed by Eli Lilly under the trade name Xigris (drotrecogin alfa) and indicated for the treatment of patients with sepsis (Swiech, Picanco-Castro & Covas 2012). The product was withdrawn in 2011, after a placebo-controlled trial showed no improvement in mortality, with an increased risk of bleeding (Dumont et al. 2016).

There have been four additional products approved for use using HEK 293 cell, three of which are blood coagulation factors. Approved in 2014, Nuwiq (human-cl rhFVIII) for the treatment of hemophilia type A, has a similar glycosylation profile to human-plasma derived factor VIII and shows no alpha-Gal or Neu5Gc epitopes. Also approved in 2014, Eloctate (rFVIIIFc) and Alprolix (rFIXFc) are recombinant fusion proteins for patients with hemophilia A and B, respectively. Both involve the fusion of the respective human coagulation factor fused to the Fc portion of immunoglobulin G1. All require PTMs more efficiently performed in human versus CHO cells: FIX requires carboxylation of its first twelve glutamic acid residues to permit its binding to phospholipid membranes, while FVIII has six tyrosine sulfation sites essential to its binding to von Willebrand factor (Dumont et al. 2016).

Treatment with recombinant factors VII and IX produced in non-human mammalian systems has been successful for many years; however the development of inhibitor antibodies in patients with hemophilia treated with replacement clotting factors still occurs in approximately 30% of patients with severe type A and 5% of those with severe type B with the type of causative mutation in the F8 or F9 gene shown to play a role in the likelihood of developing the inhibitor antibodies. The impact of the presence or lack of PTMs on inhibitor development is still unknown, but a promising sign is the lack of inhibitor development in any patient during the clinical trials of both fusion products (Dumont et al. 2016).

HT-1080 cells, derived from an ‘epithelial-like’ fibrosarcoma, have been used by Shire, using the company’s proprietary gene-activation technology to produce four commercial therapeutic

proteins: Dynepo (epoetin delta), Elaprase (iduronate-2-sulfatase), Replagal (α-glactosidase), and VPRIV (velagluciferase afla) (Dumont et al. 2016).

While still awaiting its first commercial approval, the PERC6 cell, developed by Crucell and DSM Biologics represents one of the more advanced of the available human cell lines. Currently, a variety of companies have licenced the cell line, and fourteen products are currently in Phase I and II clinical trials. Initially developed for the production of pharmaceutical grade recombinant human adenoviral vectors, (used in the manufacture of vaccine and gene therapies) the line was established from human embryonic retinoblasts. There are several features of this cell line that make it an attractive expression system. First, it can grow readily in suspension or adherent culture at very high cell densities (up to 10^8 cells/mL) in serum-free and animal-component-free media. It requires no gene amplification to achieve high protein production levels, saving time in development and improving the stability of the final clone. The development of the cell line has been extensively documented and has already met the stringent US and EU regulatory requirements, giving it an excellent safety profile (Swiech, Picanco-Castro & Covas 2012).

Transgenic Animal Expression Systems

Within six years of the approval of the first recombinant human therapeutic in 1982, the production of the first recombinant human protein expressed in a transgenic animal was achieved with the production of human factor IX in the milk of transgenic sheep. However, the first approved therapeutic for human use derived from a transgenic animal did not occur until 2006 in the European Union and 2009 in the United States. Called Atryn, and manufactured by rEVO Biologics Inc, it is a recombinant antithrombin produced in the milk of transgenic goats. A second biopharmaceutical was approved in 2012 in the European Union and 2014 in the United States, this time produced in the milk of transgenic rabbits: a recombinant human C1 esterase inhibitor protein manufactured by Pharming under the trade name Ruconest (Bertolini et al. 2016).

The main advantages of transgenic animals as an expression system combines the most desired trifecta of expression platform qualities: low cost of production, high productivity, and high quality of the synthesized protein. For example, the production cost of an animal platform is only one tenth that spent on a platform bioreactor for a mammalian cell platform: estimates are that while a 100,000L bioreactor commonly used in CHO cell platforms costs $400 million and takes 3-5 years to become operational, an animal transgenic platform, including the associated purification unit cost $80 million in a much shorter operational timeframe. Expression levels in milk from transgenic animals with yields in double digit g/L levels are achievable and the glycosylation pattern, depending on the host animal is very similar to humans (Bertolini et al. 2016).

For an example of the low cost/large scale possibilities, consider that one transgenic goat can produce 500-800 litres of milk a year. Once a founder female is established, the herd can be expanded by traditional breeding. A herd consisting of two hundred goats could thus be capable of producing metric tonnes of product. Monoclonal antibodies, including versions of currently available commercial products such as Adalimumab (Humira), Trastuzumab (Herceptin), and Cetuximab (Erbitux), expressed in goat milk show glycosylation patterns that provide equivalent and even enhanced antibody function: the Cetuximab produced in goat milk was shown to lack

1,3-alpha-galactose, a sugar moiety not found in normal human glycoproteins and thus potentially immunogenic, that is found in the commercial product (Bertolini et al. 2016).

Scale up in transgenic animals versus traditional mammalian expression systems is also an advantage. In bioreactor-based platforms, increasing bioreactor size can change a wide range of product characteristics and thus requires expensive intensive development characterization. With transgenic animals, each animal, or 'production unit', is independent and the production characteristics of the product will not change. Additionally, it is possible to produce animals that have two or more transgene, raising the possibility of producing multiple biopharmaceuticals from a single animal (Bertolini et al. 2016).

Given these advantages, what explains the large delay between the first established use and the first approve biopharmaceutical in the platform? The answer is the challenge and inefficiency of the incorporation of the transgene into the germ line of the animal (Bertolini et al. 2016).

For the expression of heterologous proteins in the milk of transgenic animals, the most developed and mature of the transgenic animal expression platforms, the process of producing transgenic animals involves introducing the gene of interest combined with a milk-specific promoter, via injection through a fine glass needle into embryos, which are then transferred into animals for gestation. First generation 'founders' are identified after birth via biopsy and then bred when mature to confirm that the transgene has been incorporated. For female founders, their milk can be tested after delivery of their own young, while for male founders, they must be bred to produce daughters, which must then also be bred themselves (Rudolph 1999).

Because integration of the transgene into the genome of the embryo is random and rare, production of transgenic founder animals is obviously time consuming and inefficient. The development of cloning by nuclear transfer using somatic cells (SCNT), which involves the transfer of the entire nucleus from a donor female founder somatic cell into donor oocytes to create a larger herd in a single breeding cycle, has helped to increase this efficiency somewhat and make characterization of the founding cell line more precise (Rudolph 1999).

Besides collection in milk, recombinant proteins have been expressed in blood, urine, seminal plasma and eggs. The production of large amounts of functional recombinant human hemoglobin in pigs has been possible since 1994, however the use of the model has been discontinued due to difficulties in purifying the human protein from the very similar porcine version, a problem that could be disqualifying issue in other models where native and human version of a desired therapeutic protein are very similar (Bertolini et al. 2016).

Large amounts of human polyclonal antibodies are expected to be extracted after immunization from the blood cows that have their own immunoglobulin genes replaced with human versions and chicken models have also been used for production of antibodies where mice systems failed to produce the appropriate immune response. However, unlike milk, blood as an expression system medium cannot store high levels of recombinant protein, the harvesting of the material is invasive and harmful to the animal and the production of very biologically active proteins may affect the health of the animal (Bertolini et al. 2016).

Urine has the advantage of being able to be produced in both male and female founder animals at a much earlier development stage than lactation, and is easy and non-invasive to harvest. The storage of large amounts of recombinant protein in urine is also less harmful to the health of the animal; however, unfortunately the amount of protein secreted into the bladder is usually low (Bertolini et al. 2016).

Seminal plasma is abundant in some male animals, especially pigs, and is easily harvested without affecting the health of the animal. However there is still a gap in the knowledge of how complex proteins are matured and secreted in semen and there is a potential impact on fertility, and hence maintenance of the transgenic line (Bertolini et al. 2016).

Eggs are an attractive source for high level recombinant protein production: a single hen can lay up to 330 eggs/year and the average egg white contains approximately 4 grams of protein. Transgenic chickens have short generation times and can have prolific rates of reproduction via artificial insemination (Farzaneh et al. 2017). Alexion produces commercially two products produced in transgenic chicken eggs both for rate genetic illnesses giving them favourable

assessments by regulatory agencies for use in humans. One is a recombinant human lysosomal acid lipase (LAL) used to treat LAL-deficiency, which received approval by the FDA in 2015 and marketed under the trade name Kanuma (Farzaneh et al. 2017). The second is SBC-103, a recombinant human alpha-N-acetyl-glucosaminidase for the treatment of mucopolysaccharidosis IIIB. SBC-103 in particular received Fast Track designation by the FDA in 2015 (Bertolini et al. 2016).

Insect Expression Systems

The *in vitro* cultivation of insect cells was first established in 1962 by Thomas Grace (Davidson 2012). While the primary motivation for the development of stable insect lines was the study of insect physiology and the production of baculoviruses for the control of insect pests, the development in the 1980s of the ability to genetically modify these baculoviruses into expression vectors for recombinant proteins, the expression system known as insect culture with baculovirus expression vector (IC-BEVS) brought insect cell culture into the mainstream of biotechnology (Drugmand, Schneider & Agathos 2012).

In the IC-BEV system, there are two distinct phases. First, the insect cells (IC) are cultivated until the desired cell density is reached. Second, the cells are infected with the baculovirus expression vector (BEV) containing the gene for the protein of interest. Infected cells stop growing (uninfected cells continue to proliferate) and produce large volumes of the protein of interest before the infection causes the cell to lyse, whereupon the product is then harvested (Drugmand, Schneider & Agathos 2012).

Highly flexible and versatile, the IC-BEV system can express proteins from any species and can also be used to express more than one gene product per vector. Protein titers of up to 300 mg/mL have been reported (Drugmand, Schneider & Agathos 2012) and estimates of production costs are approximately $50-200/g (Huang & McDonald 2009). Furthermore, as baculoviruses do not infect human cells, most adventitious insect viruses do not infect mammalian cells, and insect cells do not contain any human oncogenes, the system has safety advantages over other expression systems (Drugmand, Schneider & Agathos 2012).

Another advantage of the IC-BEV system is that the genetic engineering is focused on the vector, not the host cell line itself. This means that a single cell line can be used to generate multiple proteins, a ‘plug and play’ system, eliminating the need to characterize and seek regulatory approval for a new cell line for each new protein. It also means that a single facility can be constructed to produce multiple products, with rapid turnover between products (Cox 2012).

The system is particularly well suited for the cost effective development of vaccines where outbreaks are hard to predict and require a rapid response such as influenza. Despite considerable effort by the WHO to promote tech transfer and develop vaccine production facilities in the developing world, there is no financial incentive to maintain the capacity to produce flu vaccines in the absence of an active epidemic. The adaptable nature of the IC-BEV platform means that facilities could be rapidly changed over from making more other vaccines in the event of an epidemic (Cox 2012).

The world's first approved biopharmaceutical product derived from insect cells was produced in the IC-BEV system in 2009: Cervarix, manufactured by GlaxoSmithKline, is a vaccine against two oncogenic strains on the human papilloma virus (GlaxoSmithKline 2018). The system was also used in the manufacture of Provenge, a unique treatment for prostate cancer, manufactured by Dendreon and approved in 2010. It is the world's first personalized autologous cell immunotherapy. To make the treatment, special immune cells called antigen-presenting cells are extracted from the patient's blood and matured in culture in the presence of various recombinant prostate tumour antigens (which are made in the IC-BEV system), before being injected back into the patient (Dendreon 2018).

An insect cell expression system was also a part of the production of the world's first approved gene therapy in 2012. Glybera, (alipogene tiparvovec), manufactured by UniQure, was approved by the European Commission for the treatment of lipoprotein lipase deficiency (LPLD), a rare genetic condition characterized by high circulating levels of triglyceride-rich lipoproteins, causing severe pancreatitis. Treatment had previously consisted of a diet severely restricted in fats. Most patients found sticking to such a diet difficult to comply with and for some patients the diet did not resolve their frequent and painful attacks of pancreatitis. While the commission had initially declined the approval due to a lack of evidence of long-term efficacy of the treatment, it reassessed its opinion specifically in the case of these patients, finally granting a marketing approval under exceptional circumstances within that cohort (Walsh 2014). Unfortunately, those exceptional circumstances proved extremely exceptional: only one patient has been treated since approval (Regalado 2016) and in April 2017, UniQure announced that they would not seek

renewal of the marketing authorization of the drug when it was due to expire in October 2017 (Warner 2017).

As insect and mammalian cells both originate from multi cellular organisms, they will have similar growth requirements; however insect cells have a few advantages over their mammalian counterparts. First, they grow at lower temperatures, and can tolerate greater levels of amino acids and glucose without switching to overflow metabolism, where cells use the far less efficient fermentation instead of the more efficient respiration to generate energy despite the presence of oxygen. Also, the slightly acidic optimal growth pH of 6.2-6.9 means that an inexpensive phosphate buffer system is sufficient, eliminating the need for the carbon dioxide supply necessitated in the commonly used open bicarbonate system to maintain the neutral pH required by mammalian cells (Zitzmann et al. 2017).

Despite these advantages, the use of the IC-BEV system in the manufacture of the current main class of biotherapeutic, large glycoproteins such as monoclonal antibodies, has been limited by one major issue: while insect cells are capable of performing most post-translational modifications including cleavage, and proper folding, the glycosylation patterns produced are significantly different enough to cause clinically significant differences from products produced in mammalian systems.

Specifically, due to a lack of terminal glycosyltransferase activity, the N-glycosylation pattern on glycoproteins produced in insect cells is much simpler, resulting in N-glycans that terminate in shorter pauci-mannose type structures instead of the terminally sialylated structures produced by mammalian cells (Chang et al. 2003). The consequence is made explicit by (Toth et al. 2014):

"This difference between insect and mammalian N-glycosylation effectively precludes any opportunity to use BEVS to produce therapeutic glycoproteins because the circulatory half lives, and therefore, biological activities of these products are known to be significantly lower in the absence of terminal sialic acids".

As well as affecting biological activity, the presence of non-mammalian glycosylation is a known immunogenic trigger, further precluding the use of IC-BEVS for most therapeutic glycoproteins. However, in the context of vaccines or other therapies where the engagement of the immune system is vital to the method of action of the product, which has been the case for all currently approved biotherapeutics derived from insect cells, this ‘unwanted’ feature in fact becomes an advantage. In order to stimulate a strong immune response, it is important that the antigen be presented to the immune system in a highly immunogenic form. The uptake of antigen proteins by the antigen presenting cells (dendritic) of the immune system is mediated by ‘scavenger’ receptors on its outer surface. One such receptor is the mannose receptor that binds to the pathogen-associated glycans rich in terminal mannose residues. This uptake by the mannose receptors has been shown to enhance the ability of dendritic cells to present the antigen to T cells in both mouse and human immune systems. Therefore, because recombinant vaccine antigens produced in insect cells will be naturally abundant in terminal mannose residues on their glycans, they will offer potentially improved binding to the mannose receptors on antigen presenting dendritic cells compared to antigens produced in mammalian systems (Betting et al. 2009).

Further research into how proteins are processed in the insect endoplasmic reticulum and Golgi is needed to better understand how they differ from mammalian processes and how to possibly ‘humanize’ the proteins expressed in insects. There have been varied but related strategies to overcome the lack of sialylation at the research level, mainly involving “glycoengineering” the necessary glycosyltransferases and N-glycan precursor-producing genes into the insect cell lines. Such strategies have proved effective, but not 100% efficient, limiting the progression to commercial scale production (Chang et al. 2003) (Toth et al. 2014).

Additionally, as stated by (Geisler & Jarvis 2012), all such strategies require media supplementation with an expensive sialic acid precursor, making the production costs on scale up prohibitive. They have come up with a solution involving the insertion of a bacterial gene usually involved in sialic acid breakdown. Reasoning that insect cells ‘have the product but not the substrate’ for this enzyme, they correctly reasoned that it might function in reverse, solving this costly issue for research labs.

Media formulation is yet another area where insect cell culture still does not have an edge over mammalians systems. Like mammalian cells, insect cells do best in media supplemented with serum derived from animals. However, with the risk of adventitious agent contamination from the use of animal derived products becoming more of a focus of regulatory agencies, there is a push towards the use of chemically defined media. There are commercially available insect culture media available, but they are usually cell-line specific, proprietary, and expensive (Drugmand, Schneider & Agathos 2012).

Another potential drawback is that the lytic nature of the IC-BEV system means that the protein product can be exposed to proteases and due to the transient nature of the infection, protein secretion time is shorter than with mammalian cells expressing constitutively expressed protein. The two-step (growth and then infection) production process also complicates perfusion and continuous batch modes of operating and the lysed cell and viral fragments present challenges to downstream processing (Drugmand, Schneider & Agathos 2012).

Stable insect cell lines (with the heterologous gene integrated in the host genome) do exist, are known to work well in continuous culture mode and can sometimes produce higher titers than the BEV system. However, these systems are more expensive than the BEV system and as such, do not offer a financial incentive for industry to use them over the more familiar mammalian systems (Drugmand, Schneider & Agathos 2012).

Although still only at a research stage, insect cells are being used in the developing technology of using cells themselves as drug vehicles for novel cancer and gene therapies. In a 2015 paper, Roy et al. describe the use of insect cells as a delivery vehicle for therapeutic oncolytic viruses: proteins with known ability to regulate growth in vitro lose potency when administered systemically due to poor stability in the blood and/or rapid clearance from the circulatory system. When incorporated into a vector such as a conditionally replicating oncolytic virus, the target genes can be delivered systemically to the tumors where they replicate, along with the therapeutic agent directly at the site of the disease. However, the viruses themselves are prone to clearance by the host immune system and so a cellular carrier is needed to prevent this.

Clinically, the use of mammalian cells as carriers is limited due to multiple issues, including that they are harder to isolate, are prone to off-target homing signals, and have limited viral productivity due to the fact that oncolytic viruses are by nature cytotoxic to mammalian cells. Lacking mammalian homing receptors and able to sustain a stable infection of oncolytic virus, insect cells present a promising alternative (Roy et al. 2015).

Plant Expression Systems

With plant expression platforms, molecular farming or 'pharming', as it is more charmingly called, can take on multiple forms, such as the cultivation of whole plants, transient expression using viral infection, in vitro culture of plant tissues and plant cell culture. While whole plants offer theoretically unlimited scale up potential, the processing, regulatory, and environmental hurdles mean that plant cell culture, with its analogy to microbial and mammalian cell culture, has more immediate potential as a method of producing plant-made pharmaceuticals (Xu & Zhang 2014).

Two of the most obvious benefits of plant cell culture are cost and safety. The media used in plant cell culture are usually very simple sugar-based media and higher plants do not require light to grow. Lower plants, such as mosses, require light but use only water and inorganic salts to grow (Xu & Zhang 2014). As media costs make up a large portion of the total cost of mammalian cell culture, the savings can be considerable. The fact that human pathogens are unable to propagate in theses cultures offers a huge safety benefit and, besides the obvious quality and regulatory benefits, can also translate into reduced cost of goods, as there is less need for costly viral prevention and filtration protocols or from product losses due to contamination. An estimate of production costs puts that of plant-expressed proteins at $50-100/g versus $1000-10,000/g, a massive difference (Huang & McDonald 2009).

Both of these benefits have played a part in the success of the worlds first plant-based therapeutic for use in humans, Elelyso, made by Protalix and licensed by Pfizer, which was approved by the FDA in 2012. Made in carrot root cells, taligluciferase alfa is a recombinant active form of the human lysosomal enzyme β-glucocerebroside used to treat Gaucher's disease, a rare genetic lysosomal disease (Protalix 2018).

Currently, the only treatment is replacement enzyme therapy and since 1994, Cerezyme, an analog of the human enzyme produced in CHO cells and made by Genzyme, has been the standard of care (Xu & Zhang 2014). In 2009, due to a vesivirus contamination of one of its bioreactors, Genzyme was forced to halt production of Cerezyme. The delay, expected initially to

be of short duration, dragged on as other manufacturing issues appeared and the resulting drug shortage and subsequent crisis in the Gaucher patient community has soured their relations with the company (Pollack 2010). Into the breach steps the new Protalix biosimilar, with reduced risk of viral infections due to its plant-produced product.

In addition to its better safety profile, Protalix's product offers another cost savings over the Genzyme product directly related to its plant-based production, one that would normally be considered a drawback. Proper glycosylation of the enzyme is key to its bioavailability and function. As the glycosylation patterns between plants and mammals are different, this can have implications, for good or bad depending on the application. In the case of the enzyme involved in Gaucher's disease treatment, when grown in mammalian cells, as the Genzyme product is, the enzyme requires a 'trimming' of the glycan molecules in order to expose the required terminal mannose residues needed for uptake of the enzyme into the lysosome. When grown in plant cells, this trimming step is not required, resulting in a significant cost reduction allowing competitive pricing: Elelyso is 25% cheaper than Cerezyme (Xu & Zhang 2014).

On top of all the safety and cost benefits, the Protalix product also works better than the Genzyme product as a result of these glycosylation differences. Likely as a result of the inefficient trimming step, the mannose pattern on the mammalian derived product tends to be more heterogeneous than that of the plant derived product and clinical trials have shown that the circulatory half-life of Elelyso is significantly longer than Cerezyme, leading some to refer to these biosimilars as 'biobetters' (Xu et al. 2012).

Protalix's success with Elelyso has paved the way for other plant-produced products to enter the market. Current Protalix products in the pipeline include α-galactosidase for the treatment of Fabry's disease, an oral form of anti tumor necrotic factor (a biosimilar to the popular Enbrel by their licensing partner, Pfizer), an inhaled form of deoxyribonuclease I for the treatment of cystic fibrosis, and an oral form of their star Elelyso (Protalix 2018).

Another platform showing great promise uses moss as the plant of choice. Due to a high level of homologous recombination in its nuclear DNA, the moss platform is highly amenable to genetic

engineering or ‘glycoengineering’ to alter glycosylation patterns, more so than other organisms, including higher level plants (Decker & Reski 2007). This allows the potential optimization of glycosylation patterns, even beyond those naturally occurring, allowing the performance of already existing treatments to be improved.

For example, the clinical efficacy of IgG antibodies is mediated by an effect known as antibody dependent cellular cytotoxicity (ADCC). There is also work to be done in elucidating the clinical necessity of complex and mixed glycoforms, since research has shown that they are not always necessary. ADCC is mediated by the binding of IgG antibodies to cellular receptors. This binding is improved when IgG antibodies lack human core fucose residues at N-glycan sites.

Recombinant antibodies lacking these residues were generated in glycoengineered moss and showed a 40-fold improvement over IgG produced in mammalian CHO cells (Decker & Reski 2007). Greenovation, the company making best use of these useful properties of moss cells, states that this has promising implications for oncological applications (Greenovation 2018).

Based in Germany, Greenovation currently has four products in its pipeline, including a biosimilar to Protalix’s treatment for Gaucher’s disease, a treatment for Fabry’s and Pompe disease (both genetic lysosomal disorders), and treatments for atypical hemolytic uremia syndrome (an autoimmune disorder). The competitive edge these ‘biobetters’ aim to take advantage of is the optimized N-glycosylation possible in moss cells. As with Protalix’s product, more homogenous N-glycosylation improves the uptake of the products versus their competitors. For the Gaucher’s disease treatment, Greenovation claims up to 100% of terminal N-glycans will be terminated in the necessary mannose, resulting in improved uptake into target cells (Greenovation 2018).

Besides the above-mentioned N-glycosylation, O-glycosylation is another important but less studied type of glycosylation that has implications for folding, solubility, stability, biological function and antigenicity of a protein. In mucin-like proteins, O-glycosylation has been shown to improve the water-binding capabilities of these proteins. As mucins are known for being resistant to degradation by proteolytic enzymes, O-glycosylation has been proposed as an alternative to

PEGylation, which is time consuming, results in increase protein heterogeneity (increasing purification costs) and can negatively impact the biological activity of a protein (Webster & Thomas 2012). However caution must be taken to ensure that the benefit is not counterbalanced by an increased risk of immunogenicity in humans (Gomord & Faye 2004).

Another advantage of the plant-based platform is that plant cells have the potential to become both drug producer and drug vehicle for oral medications. Plant cells contain a rigid cell wall comprised mainly of cellulose, which cannot be broken down by human stomach enzymes but are instead broken down by microbes in the intestines. This means that a protein-based drug inside a plant cell would be protected from the stomach acids that would normally preclude an oral delivery (Xu & Zhang 2014).

'Edible vaccines' in which an antigenic protein is encapsulated in plant cells is regarded as "a cost-effective, easy to store, easy to administer and socioculturally readily acceptable vaccine delivery system, especially for developing countries" (Xu & Zhang 2014). Of seventeen neglected tropical diseases, as classified by the WHO, five already have plant-based models proposed. An example of success is a study showing that mice fed corn expressing a rabies viral G protein were 100% protected from a lethal dose of rabies. While studies to confirm protection in humans are needed, it represents a promising alternative to expensive parenteral rabies vaccines (Rosales-Mendoza et al. 2012).

Protalix has an edible form of its successful Gaucher's disease treatment in development and is performing Phase II trials. Results show that a daily dose of bioencapsulated recombinant glucocerebrosidase resulted in a sustained increase in the enzyme's presence in the blood, leading to an expected steady state level of active enzyme similar to that of healthy adults (Protalix 2018). The savings and thus competitive edge incurred from avoiding expensive purification and aseptic processing for parenteral administration would be substantial, not to mention the benefits to patients regarding ease of delivery. There are still hurdles related to assuring dose manageability and uniformity that must be addressed before this technology becomes widespread but this is a promising area for development (Xu & Zhang 2014).

While plant specific glycosylation patterns can be an advantage depending on the mode of action of the therapeutic, the fact remains that for certain classes of therapeutics, the differences between plant and mammalian patterns discourages the use of the plant-based platform (Gomord & Faye 2004). To that end, research into 'humanizing' glycosylation patterns in plant-based systems is still a major focus of current research. These efforts have included both reducing the plant-specific glycosylation by knocking out the plant genes responsible, and mimicking human glycosylation by introducing human glycotransferases (Xu, Ge & Dolan 2011) (Gomord et al. 2010). Like the efforts in insect cells, these have been partially successful but are not efficient enough yet to justify replacing the current mammalian systems on a commercial scale.

Another issue that has been plaguing plant cell culture is its relatively low protein yields which could undermine any costs benefits incurred from cheaper media. To be considered commercially viable, yields of 10 mg/ml are generally regarded as the threshold expression range and though yields currently range between 0.1 mg/ml and 247 mg/ml, it is thought that a five to ten-fold boost in average yields is required to push plant cell culture into the desired profit margin (Xu, Ge & Dolan 2011).

With regards to bioreactor design, while lessons from microbial and mammalian systems can be learned, distinct properties of plant cells, such as their large cell size, tendency to aggregate, and sensitivity to shear stress can impose limitations on scale-up. Sensitivity to shear stress deserves a particular mention as it can vary greatly between plant types and even within plant type depending on the age of the culture or the cultivation schedule: low shear stress designs are ideal (Xu, Ge & Dolan 2011).

Similar to the improvements in mammalian cell culture yields, work on this issue is being done via numerous approaches, which can be classified mainly into molecular and process development methods.

Molecular strategies include enhancing gene transcription through the use of either constitutive or inducible promoters, improving the translation efficiency by engineering of the 5' and 3' capping of transgenes, minimizing post translational degradation by targeting the expressed proteins to

the endoplasmic reticulum, or by the fusion of expressed proteins with ‘chaperone’ tags (Xu, Ge & Dolan 2011).

One such strategy that has achieved results as high as a 1500 fold improvement is known as the *HypGlyco* technology. The 0-glycosylation of Hyp residues is unique to higher plants and green algae. The *HypGlyco* technology exploits this by generating a protein tag rich in Hyp sites, which are targeted for extensive o-glycosylation and seem to function as a molecular carrier, improving the secretion of proteins into the media, improving the solubility of the protein, and protecting the protein against proteolytic degradation (Xu & Zhang 2014).

In corroboration of the expectations about O-glycosylation discussed above, the *HypGlyco* tags also improved the serum-half life of small therapeutic proteins by as much as 13 fold without significantly affecting their biological activity. Furthermore the glycosylated tags were shown to be not immunogenic in mice and only mildly so when injected as a fusion molecule. For applications where the tag would interfere with biological activity, a cleavage marker can be designed. This will increase post-purification processing costs but these are expected to be offset by the increase in yield obtained (Xu & Zhang 2014).

Process development strategies to improve yield include media optimization, cell immobilization and in situ protein removal, and optimization of both bioreactor and processing mode, i.e. batch versus fed-batch versus continuous culture.

An example of a media improvement strategy in the moss-based platform involves the addition of agents such as human serum albumin (HSA) and/or polyvinylpyrrolidone (PVP) to the media to improve the stability of the expressed protein. However, when added to the media at high concentrations, they can cause foam, which interferes with both cultivation and downstream processing. Furthermore, the addition of commercial HSA would introduce a contamination risk, negating the inherent safety profile of the plant-based system. This problem was elegantly overcome by the co-expression of the HSA gene with the protein of interest. The resulting low levels of HSA would not have been enough if added separately to the media but were enough to provide some stability benefits when co-expressed, an effect thought to be mediated by early

interactions during the joint passage of the proteins through the endoplasmic reticulum (Decker & Reski 2007).

Both Protalix and Greenovation are well poised to take advantage of the trend towards disposable technology in bioprocessing having both developed their own disposable bioreactor systems. Protalix's system, ProCellEx, is based on a design originally designed by fellow-Israeli company Osmotek, and involves a simple bag supported by an external wire cage. Greenovation's system is based on the Wave bioreactor system, where a disposable bag is rocked back and forth on a moving platform (Xu, Ge & Dolan 2011) (Protalix 2018) (Greenovation 2018).

One final issue that could impact the adoption of certain plant cells as an expression platform is the frequently seen issue of genetic instability in cultures of undifferentiated higher level plants, of which most plant cell lines are, resulting in loss of transgene expression. This problem is usually solved by the creation of master and working cell banks, which is a prerequisite for regulatory approval for expression in mammalian systems. Successful cryopreservation protocols have been developed for a few elite cell lines but there is no universal protocol developed yet; protocols would have to be adapted to each cell line (Xu & Zhang 2014).

However, genetic drift does not seem to be a problem in moss-based platforms. Instead of undifferentiated cells, with their inherent tendency towards genetic drift, fully intact moss cells are used and genetic instabilities have not been observed. Furthermore, cryopreservation protocols for long-term storage of strains have been developed so the creation of master and working cell banks is possible (Decker & Reski 2007).

Cell-free expression systems

Given the extensive time, cost and safety concerns associated with using living cells to manufacture recombinant proteins, it should not be surprising that some expression system strategies aim to do away with the cell altogether: cell-free expression systems. The first reported instance of cell free protein production took place in 1948 with the *in vitro* artificial incorporation of carboxy-radio-labeled amino acids into proteins from rat liver cells extracts. Since then it has been shown that disrupted cells from both eukaryotes and prokaryotes are still capable of synthesizing proteins (Legastelois et al. 2017).

The system can be particularly useful in the production of proteins that are difficult to produce in live cells, either because they are inherently cyotoxic, or that tend to aggregate, like membrane proteins. Conditions can be optimized by changing the reaction conditions such as redox potential, pH, hydrophobicity, or temperature. Cell-free expression systems can also allow the incorporation of unnatural, or even toxic amino acid variants (Hunt et al. 2017).

For example, the drug Onconase (ranpirnase) is a ribonuclease used to treat cancer and certain viral infections. Its method of action is to degrade tRNA, which slows and stops protein production, which presents an obvious problem when attempting expression at high-yields in living cells. Production in *E. coli* at high yields is only possible by intentionally encouraging the formation of inclusion bodies of incorrectly folded proteins, which are subsequently refolded during downstream processing. Using a cell-free protein system with supplementation of the reaction with tRNA resulted in 95% of the protein being soluble and the product was found to inihibit cancer cell growth 60 times more effectively than refolded product (Hunt et al. 2017).

The production of potentially therapeutic recombinant protein (in this instance, granulocyte macrophage colony stimulating factor) at a concentration of 700 mg/L in a standard 100L bioreactor with 10 hours was a big breakthrough for the cell-free system and shows that although there are currently no approved biopharmaceuticals produced in cell-free expression systems, the technology could soon have potential in commercial production (Huang, Lin & Yang 2012).

CRISPR Cas9

Just as the discovery of restriction enzymes from the study of the evolutionary battle between bacteria and their infecting phages led to the development of the recombinant DNA technology that has allowed for the rise of the current biopharmaceutical industry, exciting new possibilities for genetic engineering may come from another microbial antiphage defense: the CRISPR-Cas9 system (Hsu, Lander & Zhang 2014).

The story begins in 1987, when researchers studying alkaline phosphatase in *E. coli* noticed an unusual set of repetitive elements downstream of the gene they had been studying. The repeats were 29 nucleotides long and were interspersed with intervening non-repetitive sequences of about 32 nucleotides. The sequencing of more and more microbial genomes kept turning up these same elements, present in about 40% of bacteria and over 90% of archaea. Christened CRISPR in 2001 (for Clustered Regulary Interspaced Short Palidromic Repeats), a cluster of signature CRISPR-associated (*Cas*) genes were soon identified as well-conserved and typically adjacent to the repeating elements, forming the basis of a typical CRISPR loci: clustered set of Cas genes and the signature CRISPR array consisting of a series of repeat sequences (direct repeats) interspersed with variable sequences (spacers). Three different types, depending on the Cas proteins associated with the array were classified. (Hsu, Lander & Zhang 2014).

The biological significance of the loci remained a puzzle until a systematic analysis of the sequence of the spacer elements in the arrays in 2005 suggested that they came from extra-chromosomal or phage-associated DNA. Combined with previous studies that had shown that the CRISPR loci are transcribed and that phages are unable to infect archaeal cells that had CRISPR spacers matching the phage genome, this discovery led to the speculation that the CRISPR loci functioned as an immune memory and defense mechanism against infection (Hsu, Lander & Zhang 2014).

The exact mechanism was elucidated in 2007, showing that it was a nucleic acid based system where the CRISPR spacers gave target DNA specificity to the Cas proteins, usually nuclease enzymes, which would target and degrade invading phage DNA. The Cas proteins also directed

spacer acquisition (Hsu, Lander & Zhang 2014). Recognition of the phage DNA (known as the protospacer) by the Cas proteins is mediated via a variable nucleotide sequence specific to the Cas ortholog known as the protospacer adjacent motif (PAM). The Cas proteins will capture the spacer from the phage DNA and then integrate it as a new spacer between two adjacent direct repeat sequences in the CRISPR locus (Singh, Braddick & Kumar 2017).

In all three types of CRISPR loci, the direct repeat-spacer array is transcribed into a pre-CRISPR RNA (pre-crRNA). In types I and III, the pre-crRNA is processed into mature crRNA that then bind to a complex of multiple Cas proteins that help with target DNA recognition and degradation. In type II loci, a second non-coding trans-activating RNA (tracrRNA) hybrizes with the pre-crRNA to form an RNA duplex before it is cleaved by endogenous RNA nucleases. This tracrRNA-crRNA complex becomes a scaffold around which the Cas9 protein assembles, targeting it to the complementary DNA sequence in the target DNA. Cas9 is a nuclease with two active domains that performs blunt double stranded breaks. The Cas9-RNA complex scans the double stranded DNA until it encounters and binds to the PAM motif. This binding initiates strand separation, permitting the RNA portion of the Cas9-crRNA complex to bind to the DNA. If the target DNA matches the crRNA sequence, then this binding triggers the Cas9 nuclease activity (Hsu, Lander & Zhang 2014).

Because genome editing seeks to make permanent modifications to the genome, the target specificity of the site-directed system is important. For protein based systems like zinc finger nucleases (ZFN) or transcription activator-like effectors (TALEs), to achieve cleavage at a wide variety of different sites in the genome requires the costly and time consuming creation of a library of large proteins. The targeting in the type II Cas9 system, (the system most frequently used at the current time) is, by contrast, based on Watson-Crick base pairing. The system is very simple, requiring only the Cas9 protein and the mature tracrRNA-crRNA complex to function. Targeting is effected by the short RNA guide sequence, a library of which is easily created (Hsu, Lander & Zhang 2014).

The only requirement for the selection of a target is the presence of the PAM downstream of the target site. The complexity of the PAM can thus determine the overall targeting potential: longer

more stringent PAMs mean that there will likely be fewer target sequences in a genome. However, with Cas9 from *Streptococcus pyogenes* (SpCas9), the most thoroughly characterized ortholog, the PAM of NGG ensures that, on average, every 8 bp of the human genome could be targeted. The protein was also able to bind to NAG sequences, albeit with less efficiency, expanding its targeting range even further (Hsu, Lander & Zhang 2014).

However, a drawback of a greater targeting range however can be an increase in off-target binding and cleavage. SpCas9 specificity has been assayed and these studies show that the enzyme can tolerate mismatches throughout the guide sequence, depending on the number (up to five mismatches), position (the more distal to the PAM the less sensitive to mismatch) and concentration (mismatches tolerated more at higher enzyme concentrations). Again, because the enzyme can bind to both NGG and NAG sequences in its PAM regions, it is important to consider both possible PAMs when determining potential off-target cleavage sites (Hsu, Lander & Zhang 2014).

There are other interesting features of the Cas9 protein that expand the versatility of the CRISPR/Cas9 system as a genome engineering tool. First, the Cas9 protein cleaves DNA through the activity of its RuvC and HNH nuclease domains, each of which nicks a strand of DNA to create double stranded breaks. By inactivating either the RuvC or the HNH domain, Cas9 can be converted into a 'nickase' capable of performing only single-stranded breaks, which are repaired using the more high-fidelity base excision repair (BER) pathway. To improve the on target specificity of double stranded breaks, a pair of guide RNAs can be used with such nickase Cas9 mutants to increase the number of bases that are effectively recognized in the target DNA. The two cuts mimic a double stranded break, which is usually repaired via the lower fidelity non-homologous end joining (NHEJ), leading to insertion/deletion (indels) which usually lead to frame mutations causing effective knock outs of the target region. Off target single strand nicks will be more precisely repaired. This strategy has improved Cas9 specificity by up to 1500x compared to wild-type Cas9 (Hsu, Lander & Zhang 2014).

Double stranded breaks can also be repaired via the homologous repair (HR) pathway, which relies on a repair template, which can be either endogenous (sister chromosome) or exogenous

(donor plasmid or vector) but compared to NHEJ, efficiency rates are low. Single stranded nicks using Cas9 nickases could also be used with donor recombination by adding 'homology arm' sequences on either side of the donor gene of interest in the expression cassette. These 'arms' also help prevent off-target integration (Hsu, Lander & Zhang 2014).

Finally, by inactivating both cleavage domains to produce 'dead Cas9' (dCas9), the DNA-binding properties are retained, essentially turning it into an RNA-guided homing device. The fusion of various effector proteins can then be used to carry out a variety of genetic engineering tasks. For example, if targeted to a promoter, dCas9 can block transcription of a gene or dCas9 can be fused to transcriptional repressor domains. Conversely, dCas9 can be converted into a synthetic transcriptional activator when fused to transcriptional activator proteins. By attaching epigenetic modulating enzymes, epigenetic control can be investigated. Attaching fluorescent in situ hybridization tags can be used for live cell imaging or to study complex chromosomal architecture and nuclear organization. (Hsu, Lander & Zhang 2014).

To date, the Cas9 system has been used in a wide variety of species, including human, bacteria, yeast, zebrafish, mice, pigs, monkey and common crops (Hsu, Lander & Zhang 2014). Given the exciting implications for genetic engineering the system offers, it is not surprising that the ubiquitous CHO cells would get their turn as well; in 2015, using the CRISPR/Cas9 and a donor plasmid containing short homology arms on either side of the gene of interest to encourage homologous recombination, a targeting efficiency of between 7.4% to 24.8% (depending on the target locus) was obtained and a comparison of clones transfected by random instead of targeted integration showed no affect on expression level or cell viability. While such efficiencies are quite low compared to those achieved using CRISPR/Cas9 in bacteria, yeast and even human expression systems, it corroborates the finding of other studies showing that CHO cells show lower levels of homologous repair (Lee et al. 2015).

Yeast exhibit a large variability in their preference for HR or NHEJ DNA repair pathways when introducing double stranded breaks using traditional genetic engineering methods. *S. cerevisiae* tends to exclusively tend towards HR, with a close to 100% correct integration of donor vectors with short (approximately 50 base pair) homology arms. In contrast, *P. pastoris*, donor vectors

require long homology arms (greater than 1000 base pair) and HR occurs at a frequency of 1% to 30% depending on the targeted site of integration. (Rashmanova et al. 2018).

In addition, conventional expression/knockout cassettes usually require selection markers such as antibiotic resistance or an auxotrophic marker. The inclusion of drug resistance markers in particular can have impacts on the eventual approval of the strain for use in industrial applications due to public health and safety concerns; the need for marker-less modifications and higher rates of HR have been sought after (Rashmanova et al. 2018).

The use of the CRISPR-Cas9 system in *S. cerevisiae,* first established in 2013, achieved a 99% recombination frequency. Such high integration frequencies mean that selection markers can often be omitted. Such precise and efficient targeting has meant that manipulations of the *S. cerevisiae* genome have been greatly facilitated and a number of different deletions, insertions, and even introduction of point mutations has been achieved (Rashmanova et al. 2018).

When introducing specific single nucleotide mutations, it is necessary to prevent Cas9 from repeated cutting of the first HR-directed repair because Cas9 can tolerate single nucleotide mismatches and possibly cleave the region repeatedly. This is avoided by designing the introduction of an additional change surrounding the point mutation such that the PAM or crRNA target site is disrupted, preventing Cas9 from cleaving the site again (Rashmanova et al. 2018).

In addition to the CRISPR/Cas9 system's wide-ranging potential as a research tool for the development of biopharmaceuticals, there is also the exciting potential of its use as a therapeutic itself. For recessive disorders caused by a single gene, such as cystic fibrosis, sickle-cell anemia, or Duchenne muscular dystrophy, Cas9 could be used to repair the mutation. This is an improvement over other proposed methods, such as delivering functional copies via viral vectors, as the gene is expressed in its natural context. For dominant disorders where only one working copy is needed, the affected allele could be inactivated. Protective mutations may be introduced to help prevent against infection by certain infections like HIV or to overcome therapeutic resistant mutations like statin-resistant hypercholesterolemia. Finally, similar to the currently

marketed prostate cancer drug Provenge, certain specialized cells can be extracted from patients are edited ex vivo before being reinfused into the patient (Hsu, Lander & Zhang 2014).

There are still hurdles to overcome however, before Cas9-therapies become a reality. First, the Cas9 cassettes must be able to be efficiently delivered to the target cell. Unfortunately, the Cas9 gene is quite large and unable to fit into most commonly used expression vectors. Potential strategies to overcome this include removing extraneous domains from the protein, or by searching for smaller Cas9 orthologs. Expression vectors will also have to be designed for tightly regulated expression in target tissues to restrict Cas9 function only to areas where it is required. Off target activity in the context of a therapeutic could result in unwanted disruptions to other genes, which could potentially be oncogenic. Avoiding an immune response to the Cas9 protein is also an important consideration. Strategies include humanizing Cas9 peptides in a similar manner to that performed for antibodies or by using the system itself to disrupt major histocompatibility complex trafficking, in a manner similar to viruses (Mali, Esvelt & Church 2013).

Excitement regarding the potential of the CRISPR/Cas9 system has also been dampened somewhat by an ongoing fight over intellectual property rights. The Broad Institute of Cambridge, Massachusetts filed arguments in October 2017, defending a key patent regarding the use of CRISPR/Cas9 in eukaryotic cells, of which the implications will be obviously very broad and potentially expensive for research and development labs hoping to use the technology. The University of California, Berkeley have appealed, arguing that their earlier patent for CRISPR/Cas9 in prokaryotic systems rendered the Broad Institutes' patent for use in eukaryotic cells 'obvious' and thus invalid. The case is expected to be heard by the US Federal Circuit court of appeals sometime this year (Ledford 2017).

Chapter 3: Overall Conclusions and Recommendations

The first biopharmaceutical produced in a prokaryotic expression system was in *Eschericia coli* in 1982. The first biopharmaceutical produced in a single celled eukaryotic expression system was in *Saccharomyces cerevisiae* in 1984. The first biopharmaceutical produced in mammalian expression systems was in Chinese Hamster Ovary cells in 1987. As proof of the inherent inertia and conservatism of the pharmaceutical industry, in the thirty plus years since then, the vast majority of biopharmaceuticals are produced in exactly the same three expression systems: *E. coli*, CHO and *S. cerevisiae.*

As commonly used and well studied model organisms in earlier molecular biology work, these three organisms would have had a substantial head start on other expression systems, and the intervening years of experience, investment and impressive safety record have helped to solidify and entrench their dominance.

In general, if a therapeutic protein does not require post translational modifications for its clinical function, the overwhelming preference is for manufacture in microbial or lower eukaryotic expression systems due to their ease of genetic manipulation, fast growth rates, high productivity, robust physiology, and less expensive media. However, if a therapeutic protein requires post translation modifications for its clinical function, CHO cells are the expression system of choice. The outsize importance in the industry in terms of profit from therapeutic protein classes such as monoclonal antibodies, which are large, multimeric proteins requiring both glycosylation and disulphide bonds, explains the current dominance of the platform.

Given the long timelines and high costs involved in drug development timelines, with the high risk of failure, it is unsurprising that the biopharmaceutical industry tends to be conservative and risk averse in its choices. Especially with biopharmaceuticals, where even small changes can have a large effect on the final product, resulting in a large validation and regulatory reporting burden, the tendency is to rely on what is already known to pass muster with regulatory agencies. A recent industry survey had 46% of respondents cite a lack of regulatory guidance and 37% cite

an increase in regulatory oversight as some of the greatest challenges for the biomanufacturing industry (BioPharma-Reporter.com 2017).

For post translation modifications, especially those like glycosylation, where even small changes to the process can result in changes to the glycosylation pattern of the final product, the ability to determine which changes are clinically relevant is very important when considering process changes or even more importantly, when comparing a biosimilar to an innovator product.

For the most part, the impact of such changes can be predicted or risk assessed based on previous literature or product class history. In fact, such knowledge has permitted the specific engineering of post translational modifications to selectively enhance biopharmaceuticals. However, ultimately the only way to assess the clinical impact of such changes is through clinical trials or clinical use and post-market surveillance (Walsh & Jefferis 2006).

Exactly which changes to critical quality attributes can be altered without affecting the efficacy and safety of the product have long been debated (Schiestl et al. 2011). While both European and the United States regulatory agencies allow that 'identical' is not appropriate when considering biopharmaceuticals, defining the terms 'comparable', 'similar' and 'very similar' for setting acceptance levels still presents a challenge (Berkowitz et al. 2012). The existence of products on the market that display glycosylation differences pre-and post change, without any alteration to the product label indicates that regulatory bodies are willing to be pragmatic in their assessments if there is enough evidence to prove that the differences not likely to result in any significant clinical impacts (Schiestl et al. 2011).

So far, existing expression platforms have served the industry well and business forecasts for the biopharmaceutical market expect growth and investment to continue as more and more therapeutic proteins are discovered. However, there is a growing awareness by both governments and patients of rising health care costs and biopharmaceuticals represent some of the most costly therapies available. Consider that for every story expressing excitement about the appearance of the world's first gene therapy, there was one expressing horror at its million-dollar price tag.

In addition, as legacy products lose patent protection and clearer understanding of regulatory rules regarding the approval of biosimilars increases the number of 'me too' drugs competing for market share, there is growing pressure on biopharmaceutical companies to reduce their cost of goods. These pressures have led to a reevaluation of all aspects of the manufacturing process. As the choice of expression system represents such a large investment of time and money, the potential savings opportunities offered by alternative expression systems have begun to slowly make inroads into resisting the inertia of the industry.

The same industry survey mentioned above showed that 45% of respondents felt that the industry was too reliant on the CHO cell platform, and 33% cited its being approved as safe by regulators as the reason for this. Interestingly, a lack of trust and innovation in alternative cell lines were ranked as the least likely reasons for this over-reliance, indicating that fear of regulatory issues, and not fear of innovation, is what is likely holding back more moves into alternative systems. A few brave (and/or well funded) companies have started to make investments in a few niche products, as evidenced by the slow but growing trickle of products approved in alternative systems. Clearly however, there is a need for regulatory agencies to provide more guidance, encouragement and incentive for industry to move beyond the status quo.

When considering advancements in expression systems, there are two main directions to consider: first is the re-engineering of existing systems, and second is the identification of new alternative host organisms. In terms of product quality, efforts have mainly focused on re-engineering existing systems, usually through manipulation of the host genome or culture conditions to improve the glycosylation profile. However, in terms of productivity, there are certain physiological limits that can't be extended indefinitely. Solutions in this area are more likely to come from new organisms with a natural capacity for high protein production and secretion, fast doubling time, and ability to grow in simple and inexpensive media (Legastelois et al. 2017).

For all the decades of experience and cutting edge technology employed in the biopharmaceutical industry, there is still the sense that the cell is a little 'black box' where expression vectors and put in and, it is hoped that a high producing clone comes out. Huge improvements in yield have

been effected mainly through empirical approaches like screening or process optimization. In addition, these improvements are often specific to a single product, necessitating the repeat of often laborious and expensive work for each new therapeutic (Lee et al. 2015). Considering how important the species is to the industry, it is surprising that it took until 2011for the first CHO genome to be produced, given that complete genome sequencing had been possible since the turn of the millennium. The growing availability of genomic and proteomic data has helped to accelerate efforts to improve existing platforms and given current drug development and approval timelines, it is expected that they will continue to dominate the industry into the next decade.

However, despite improvements, the costs of manufacturing biopharmaceuticals in mammalian cells remain high: culture durations are necessarily longer due to the inherently slower growth rate, requirements for complex media and expensive chromatographic resins (i.e. protein A used in the initial capture of monoclonal antibodies from the harvest pool) for their purification. These high costs are preventing CHO-based manufacturing from being able to supply markets in the developing world. The Gates Foundation, the charitable foundation founded by Bill and Melinda Gates from their Microsoft fortune, works with pharmaceutical companies to supply vaccines, drugs and diagnostics to developing countries at reduced costs. The foundation advises that for routine treatment of HIV or infectious diseases in the developing world, the cost of monoclonal antibodies will need to drop to $10/g or less, a formidable challenge for systems currently ranging in price from $50-$300g/L (Matthews et al. 2017).

A further challenge relates to potential treatments for prevalent chronic diseases such as Alzheimer's and heart disease. However the likely required amounts of product are likely to present a manufacturing challenge. For example, assuming a monthly dose of 1g of therapeutic antibody is prescribed to a patient population of one million people. This will require 12 tons of product, a figure that is already double the expected amount of product produced in CHO in 2016. Furthermore, these one million patients would comprise less than 20% of the potential patient population in the United States alone. To meet these projected needs, a massive investment in additional production sites and/or bioreactor capacity would be required(Matthews et al. 2017).

Alternative mammalian cell lines, including human, transgenic animals, insects and plants, all have the same limiting issue of slower growth rates to permit the massive jump in productivity required. Bacterial systems, while offering the required increase in growth rate, are still, despite ongoing cell line engineering efforts, are still unable to efficiently fold and glycosylate large glycoproteins like monoclonal antibodies. Eukaryotic microorganisms, such as yeast, fungi, microalgae and protozoa offer a potential middle ground, growing quickly like bacterial systems but possessing the required organelles, cellular environments and enzymes required to properly produce, fold, secrete and glycosylate large complex biopharmaceuticals (Matthews et al. 2017).

Genome size can affect costs of development: cost of sequencing and mutations can be more consistently identified: makes strain engineering both quicker and cheaper. For example, currently to sequence an entire P. pastoris (yeast) genome is $100 per sample compared to tens of thousand for CHO. This means that for the cost of developing one CHO cell line, multiple strains can be developed. Furthermore, genome organization is important for regulatory approval as the number of chromosomes affects the cell line stability: CHO cells are notoriously heterogeneous in the number of chromosomes between cell families and clones (Brown et al. 2017). Smaller genomes also mean there is less repetitive or redundant sequences, meaning cells are less likely to survive the loss of part or all of a chromosome. This means that clones are likely to be more stable in a smaller genome. Yeast cells are also highly amenable to genetic modification, showing much higher homologous recombination rates than CHO cells (Matthews et al. 2017).

Although cell growth rates are similar to mammalian expression systems, there are tantalizing cost savings to industry and safety benefits to patients associated with both the insect and plant based platforms. From the estimated cost of goods savings alone (at least a factor of ten according to estimates (Huang & McDonald 2009)), there should be a rush by industry to adopt these platforms. Genzyme's public relations nightmare with Cerezyme should be a lesson to all companies using mammalian cells that sometimes the best risk management strategy is risk elimination. And yet…the small but significant differences in glycosylation patterns between insects, plants and mammals remains the biggest obstacle to plants and insects challenging mammalian systems.

For insect-based therapeutics, the presence of already approved products with more already in the pipeline, marks the emergence of IC-BEVS as an accepted commercial platform with exciting potential for growth. IC-BEVS is proving to be a safe, cost effect, versatile and robust platform, especially for the production of vaccines and personalized immunotherapies. For this class of biopharmaceutical, the immunogenic properties of insect glycosylation patterns are not a risk but a clear therapeutic benefit.

The low yields that had typically held back interest in commercializing plant systems are being overcome with the same molecular and process improvements that have increased mammalian cell yields and the success of Protalix has shown that the differing glycosylation patterns between plants and mammals need not be a deal-breaker and can sometimes constitute a competitive advantage in cost, safety and therapeutic function. However, like their insect counterparts, such advantages are only available for therapeutics where plant glycosylation patterns are actively desired. A ray of hope is the amenability of the moss-based platform to glycoengineering.

Despite the high costs and safety risks of mammalian cell cultures, the fact remains that for certain classes of therapeutics, there is a requirement for glycosylation patterns that are as close to human patterns as possible and even with the ongoing glycoengineering efforts to humanize the glycosylation patterns of both plant and insect models, such attempts have not yet reached the stage where it is more efficient or effective than using mammalian cells. A major breakthrough overcoming this issue is needed before insect or plant cells will displace mammalian cells in this arena and one must not forget that research in mammalian cell culture is not stagnating either. However, with the pressures of the increasing number of biosimilars, costs, and safety pushing glycoengineering efforts, the question should not be if these alternative systems will ever mount a challenge but when.

The breakthrough that finally permits alternative systems to break out of their niches may already be here in the form of the CRISPR/Cas9 system for genetic manipulation and investigation, the most exciting development in molecular biology in the last decade. By bringing together all three major classes of biochemical molecules, (DNA, RNA and protein) in a versatile and programmable way, it is a powerful tool for the study and manipulation of complex biological

systems. The ability to efficiently perform manipulations at specific locations in the genome will help to improve the performance and drive down the cost of producing cell lines in existing expression systems as well as opening up the possibility of combining the best traits of disparate expression systems into one, as strategy that has been already put to great use in yeast systems.

One area where the technology will be of great use is in the characterization and manipulation of of post translational modification pathways, especially glycosylation, as there is much work to be done in determining just which post translational modifications are necessary for the clinical function of a given therapeutic. If it can be more easily determined, through cheap and rapid genetic manipulation that a given modification is clinically irrelevant, it may permit the switch from more complex and expensive mammalian systems to cheaper alternative systems. Moreover, it also facilitates engineering of native proteins to include non-natural modifications that may confer a therapeutic improvement, such as afucoslyated mAbs or additional glycosylation sites on erythropoietin.

While one optimal expression system for the production of all possible therapeutic proteins may not be possible or desirable, the recent development of a minimal bacterial genome (Hutchison et al. 2016), combined with the ability to transfer desired genes from one system into another, means that the ability to build the optimal expression system for a particular class of therapeutic may be on the horizon. Interestingly, 149 genes found to be essential in the minimal bacterial genome were of unknown function, suggesting that there are still undiscovered cellular functions that are essential to life (Hutchison et al. 2016). Now that the genes, if not the functions, have been identified, they offer potential targets for optimizing current and existing systems. At the very least they identify genes that should not be interfered with or omitted during genetic engineering, potentially saving wasted time and resources.

The advent of such precise, simple, efficient and cost-effective genetic manipulation technology means exciting times are on the horizon for the range of expression systems available to the biopharmaceutical industry.

Bibliography

Ajinomoto Althea 2016, *ContractPharma*, viewed May 2018, <https://www.contractpharma.com/csd/profile/ajinomoto-althea-inc/view_biologics-candidate-enters-clinical-trials-for-first-time-using-corynex-protein-expression-system/>.

Berkowitz, SA, Engen, JR, R, MJ & Bones, GB 2012, 'Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars', *Nature Reviews Drug Discovery*, vol 11, no. 7, pp. 527-540.

Berlec, A & Strukelj, B 2013, 'Current state and recent advances in biopharmaceutical production in Eschericia Coli, yeasts and mammalian cells', *Journal of Industrial Microbiology & Biotechnology*, vol 3, no. 4, pp. 257-274.

Bertolini, LR, Meade, H, Lazzarotto, CR, Martins, LT, Tavares, KC, Bertolini, M & Murray, JD 2016, 'The transgenic animal platform for biopharmaceutical production', *Transgenic Research*, vol 25, no. 3, pp. 329-343.

Betting, DJ, Mu, XY, Kafi, K, McDonnell, D, Rosas, F, Gold, D & Timmerman, JM 2009, 'Enhanced immune stimulation by a therapeutic lymphoma tumor antigen vaccine produced in insect cells involves mannose receptor targeting to antigen presenting cells', *Vaccine*, vol 27, pp. 250-259.

Biology Online Dictionary 2018, viewed April 2018, <https://www.biology-online.org/dictionary/Expression_system>.

BioPharma-Reporter.com 2017, 'State of the Global Manufacturing and Bioprocessing Industry 2017', Survey, William Reed.

Brown, A, Kalsi, D, Fernandez-Martell, A, Cartwright, J, Barber, NOW, Patel, YD, Turner, R, Bryant, CL, Johari, YB & James, DC 2017, 'Expression Systems for Recombinant Biopharmaceutical Production by Mammalian Cells in Culture', in T Vaughan, J Osbourn, B Jallal (eds.), *Protein Therapeutics*, 1st edn, Wiley-VCH Verlag GmbH & Co. KGaA.

Butler, M & Spearman, M 2014, 'The choice of mammalian cell host and possibilities for glycosylation engineering', *Current Opinion in Biotechnology*, vol 30, pp. 107-112.

Chang, G-D, Chen, C-J, Lin, C-Y, Chen, H-C & Chen, H 2003, 'Improvement of glycosylation in insect cells with mammalian glycosyltransferases', *Journal of Biotechnology*, vol 102, pp. 61-71.

Chinyere Anyaogu, D & Hasbro Mortensen, U 2015, 'Manipulating the glycosylation pathway in bacterial and lower eukaryotes for production of therapeutic proteins', *Current Opinion in Biotechnology*, vol 36, pp. 122-128.

Cilian 2018, viewed May 2018, <http://www.cilian.com/technology-and-products/>.

Corchero, JL, Gasser, B, Resina, D, Smith, W, Parrilli, E, Vasquez, F, Abasolo, I, Guiliani, M, Jantti, J, Ferrer, P, Saloheimo, M, Mattanovich, D, Schwartz, SJ, Tutino, ML & Villaverde, A 2013, 'Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics', *Biotechnology Advances*, vol 31, pp. 140-153.

Cox, MMJ 2012, 'Recombinant protein vaccines produced in insect cells', *Vaccine*, vol 30, pp. 1759-1766.

Davidson, E 2012, 'History of Insect Pathology', in F Vega, H Kaya (eds.), *Insect Pathology*, 2nd edn.

Decker, EL & Reski, R 2007, 'Moss bioreactors producing improved biopharmaceuticals', *Current Opinion in Biotechnology*, vol 28, pp. 393-398.

Declerk, PJ 2012, 'Biologicals and Biosimilars: a review of the science and its implications', *Generics and Biosimilars Initiative Journal*, vol 1, no. 1, pp. 13-16.

Dendreon 2018, *Empowering the immune system to detect and attack cancer cells*, viewed April 2018, <https://www.dendreon.com/Our-Science>.

Drugmand, J-C, Schneider, Y-J & Agathos, SN 2012, 'Insect Cells and factories for biomanufacturing', *Biotechnology Advances*, vol 30, pp. 1140-1157.

Dumont, J, Euwart, D, Mei, B, Estes, S & Kshirsagar, R 2016, 'Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives', *Critical Reviews in Biotechnology*, vol 36, no. 6, pp. 1110-1122.

European Medicines Agency 1999, 'Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biotech Products ICH Q6B'.

Farzaneh, M, Hassani, S-N, Mozdziak, P & Baharvand, H 2017, 'Avian embryos and related cell lines: a convenient platform for recombinant proteins and vaccine production.', *Biotechnology Journal*, vol 12.

Felberbaum, RS 2015, 'The baculovirus expression vector system: A commercial manufacturing platform for viral vaccines and gene therapy vectors', *Biotechnology Journal*, vol 10.

Geisler, C & Jarvis, D 2012, 'Innovative use of a bacterial enzyme involved in sialic acid degradation to initiate sialic biosynthesis in glycoengineered insect cells', *Metabolic Engineering*, vol 14, pp. 642-652.

Ghaderi, D, Zhang, M, Hurtado-Ziola & Varki, A 2012, 'Production platforms for biotherapeutic glycoproteins. Occurence, impact, and challenges of non-human sialylation', *Biotechnology and Genetic Engineering Review*, vol 28, pp. 147-176.

GlaxoSmithKline 2018, , viewed April 2018, <https://au.gsk.com/media/324141/cervarix-pi_012_approved.pdf>.

Gomord, V & Faye, L 2004, 'Posttranslation modification of therapeutic proteins in plants', *Current Opinion in Plant Biology*, vol 7, pp. 171-181.

Gomord, V, Fitchette, AC, Menu-Bouaouiche, L, Saint-Jore-Dupas, C, Plasson, C, Michaud, D & Faye, L 2010, 'Plant-specific glycosylation patterns in the context of therapeutic protein production', *Plant Biotechnology Journal*, vol 8, pp. 564-587.

Greenovation 2018, viewed May 2018, <http://www.greenovation.com/services.html#examples>.

Hamilton, SR & Gerngross, TU 2007, 'Glycosylation engineering in yeast: the advent of fully humanized yeast', *Current Opinion in Biotechnology*, vol 18, pp. 387-392.

Hsu, PD, Lander, ES & Zhang, F 2014, 'Development and Applications of CRISPR-Cas9 for Genome Engineering', *Cell*, vol 157, pp. 1262-1278.

Huang, CJ, Lin, H & Yang, X 2012, 'Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements', *Journal of Industrial Microbiology and Biotechnology*, vol 39, pp. 383-399.

Huang, TK & McDonald, K 2009, 'Bioreactor engineering for recombinant protein production in plant cell suspension cultures', *Biochemical Engineering*, vol 45, pp. 168-184.

Hunt, JP, Yang, SO, Wilding, KM & Bundy, BC 2017, 'The growing impact of lyophilized cell-free protein expression systems', *Bioengineered*, vol 8, no. 4, pp. 325-330.

Hutchison, C, Chuang, R-Y, Noskov, V, Assad-Garcia, N, Deerink, TJ, Ellisman, MH, Gill, J, Kannan, K, Karas, BJ, Ma, L, Pelletier, J, Qi, Z-Q, Richter, RA, Strychalski, EA, Sun, L, Suzuki, Y, Tsvetanova, B, Wise, KA, Smith, HO, Glass, JI, et al. 2016, 'Design and synthesis of a minimal bacterial genome', *Science*, vol 351, no. 6280.

Jayapal, KP, Wlaschin, KF, Hu, WS & Yap, MG 2007, 'Recombinant protein therapeutics from CHO Cells - 20 years and counting', *Chemical Engineering Progress*, vol 103, no. 10, pp. 40-47.

Jefferis, R 2016, 'Post translational Modifications and the Immunogenicity of Biotherapeutics', *Journal of Immunology Research*, pp. 1-15.

Kaltashov, IA, Bobst, CE, Abzalimov, RR, Wang, G, Baykal, B & Wang, S 2012, 'Advances and challenges in analytical characterization of biotechnology products: mass spectrometry-based approaches to study properties and behaviour of protein therapeutics', *Biotechnology Advances*, vol 30, no. 1, pp. 210-222.

Kingston, RE, Kaufman, RJ, Bebbington, CR & Rolfe, MR 1993, 'Amplification Using CHO Cell Expression Vectors', *Current Protocol In Molecular Biology*.

Kuriakose, A, Chirmule, N & Nair, P 2016, 'Immunogenicity of Biotherapeutics: Causes and Associations with Post Translational Modifications', *Journal of Immunology Research*, pp. 1-18.
Lai, T, Yen, Y & Ng, SK 2013, 'Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production', *Pharmaceuticals*, vol 6, pp. 579-603.

Lalonde, M-E & Durocher, Y 2017, 'Therapeutic glycoprotein production in mammalian cells', *Journal of Biotechnology*, vol 251, pp. 128-140.

Ledford, H 2017, *Bitter CRISPR patent war intensifies*, viewed May 2018, <https://www.nature.com/news/bitter-crispr-patent-war-intensifies-1.22892>.

Lee, JS, Bercheut Kallehauge, T, Ebdrup Pederson, L & Faustrup Kildegaard, H 2015, 'Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway', *Scientifi Reports*, vol 5, no. 8572.

Lee, JS, Grav, LM, Lewis, NE & Faustup Kildegaard, H 2015, 'CRISP/Cas9-mediated genome engineering of CHO cell factories: application and perspectives', *Biotechnology Journal*, vol 10, no. 7, pp. 979-994.

Lee, YJ & Jeong, KJ 2015, 'Challenges to production of antibodies in bacteria and yeast', *Journal of Bioscience and Bioengineering*, vol 5, pp. 483-490.

Legastelois, I, Buffin, S, Peubez, I, Mignon, C, Sodoyer, R & Werle, B 2017, 'Non-conventional expression systems for the production of vaccine proteins and immunotherapeutic molecules', *Human Vaccines and Immunotherapeutics*, vol 13, no. 4, pp. 947-961.

Liu, X, Yang, Y, Zhang, W, Sun, Y, Feng, P, Jeffrey, L, Harvey, L, McNeill, B & Bai, Z 2015, 'Expression of recombinant protein using Corynebacterium Glutamicum: progess, challenges and application', *Critical Reviews in Biotechnology*, vol 1, no. 13.

Mali, P, Esvelt, KM & Church, GM 2013, 'Cas(as a versatile too for engineering biology', *Nature Methods*, vol 10, no. 10, pp. 957-963.

Matthews, CB, Wright, C, Kuo, A, Colant, N, Westoby, M & Love, JC 2017, 'Reexamining opportunities for therapeutic protein production in eurkaryotic microorganisms', *Biotechnology and Bioengineering*, vol 114, pp. 2432-2444.

Meyer, H-P & Schmidhalter, DR 2012, 'Microbia Expression Systems and Manufacturing from a Market and Economic Perspective', in EC Agbo (ed.), *Innovations in Biotechnology*.
Nielsen, J 2013, 'Production of biopharmaceutical proteins by yeast', *Bioengineered*, vol 4, no. 4, pp. 207-211.

Noh, SM, Sathyamurthy, M & Lee, GM 2013, 'Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production', *Current Opinion in Chemical Engineering*, vol 3, pp. 391-397.

Pollack, A 2010, *Genzyme Drug Shortage Leaves Users Feeling Betrayed*, viewed April 2018, <https://www.nytimes.com/2010/04/16/business/16genzyme.html>.

Protalix 2018, *Protalix*, viewed May 2018, <http://protalix.com/about/elelyso/>.

Rashmanova, H, Astrid, W, Glieder, A, Kovar, K & Vogl, T 2018, 'Implementing CRISPR-Cas technologies in conventional and non-conventional yeast: current state and future prospects', *Biotechnology Advances*.

Regalado, A 2016, *The World's Most Expensive Medicine Is a Bust*, viewed April 2018, <https://www.technologyreview.com/s/601165/the-worlds-most-expensive-medicine-is-a-bust/amp/>.

Rosales-Mendoza, S, Govea-Alonso, DO, Monreal-Escalante, Fragoso, G & Sciutto, E 2012, 'Developing plant-based vaccines against neglected tropical disease: where are we?', *Vaccine*, vol 31, pp. 40-40.

Rose, T, Winkler, K, Brundke, E, Jordan, I & Sanding, V 2005, 'Alternative strategies and new cell lines for high-level production of biopharmaceuticals', in J Knablein (ed.), *Modern Biopharmaceuticals*, 1st edn, Wliey-VCH.

Roy, D, Power, A, Borgeois-Daigneault, M, Falls, T, Ferreira, L, Stern, A & Tanese de Souza, C 2015, 'Programmable insect cell carriers for systemic delivery of integrated cancer biotherapy', *Journal of Controlled Release*, vol 220, pp. 210-221.

Rudolph, N 1999, 'Biopharmaceutical production in transgenic livestock', *Tibtech*, vol 17, pp. 367-374.

Schiestl, M, Thomas, S, Torella, C, Cepeljnik, T, Toll, H & Grau, R 2011, 'Acceptable changes in quality attributes of glycosylated biopharmaceuticals', *Nature Biotechnology*, vol 29, no. 4, pp. 310-312.

Singh, V, Braddick, D & Kumar, DP 2017, 'Exploring the potential of genome editing CRISPR-Cas9 technology', *Gene*, vol 599, pp. 1-18.

Swiech, K, Picanco-Castro, V & Covas, DT 2012, 'Human cells: New platform for recombinant therapeutic protein production', *Protein Expression and Purification*, vol 84, pp. 147-153.

Toth, AM, Kuo, C-W, Khoo, K-H & Jarvis, D 2014, 'A new insect cell glycoengineering approach provides baculovirus-inducible glycogene expression and increases human-type glycosylation efficiency', *Journal of Biotechnology*, vol 182, pp. 19-29.

Varki, A 2008, 'Sialic acids in human disease and health', *Trends in Molecular Medicine*, vol 14, no. 8, pp. 351-360.

Walsh, G 2003, *Biopharmaceuticals: Biochemistry and Biotechnology*, 2nd edn, John Wiley & Sons, Ltd.

Walsh, G 2010, 'Post-translational modifications of protein biopharmaceuticals', *Drug Discovery Today*, vol 15, no. 17/18, pp. 773-780.

Walsh, G 2014, 'Biopharmaceutical benchmarks 2014', *Nature Biotechnology*, vol 32, no. 10, pp. 992-1000.

Walsh, G & Jefferis, R 2006, 'Post-translation modifications in the context of therapeutic proteins', *Nature Biotechnology*, vol 24, no. 10, pp. 1241-1252.

Wang, Y, Yau, Y-Y, Perkins-Balding, D & Thomson, JG 2011, 'Recombinase technology: applications and possibilities', *Plant Cell Reports*, vol 30, no. 3, pp. 267-285.

Warner, E 2017, *Goodbye Glybera! The World's First Gene Therapy Will Be Withdrawn*, viewed April 2018, <https://labiotech.eu/uniqure-glybera-marketing-withdrawn/>.

Webster, DE & Thomas, MC 2012, 'Post-transational modification of plant-made foreign proteins; glycosylation and beyond', *Biotechnology Advances*, vol 30, pp. 410-418.

Wright, T 2015, *Top 15 Trends in Biopharmaceutical Manufacturing*, viewed April 2018, <https://www.contractpharma.com/contents/view_online-exclusives/2015-07-30/top-15-trends-in-biopharmaceutical-manufacturing/11688>.

Xu, J, Ge, X & Dolan, M 2011, 'Towards high-yield production of pharmaceutical proteins with plant cell suspension cultures', *Biotechnology Advances*, vol 29, pp. 278-299.

Xu, J, Medrano, MC, Cramer, C & Weathers, PJ 2012, 'Green factory: plants as bioproduction platforms for recombinant proteins', *Biotechnology Advances*, vol 30, pp. 1171-1184.

Xun, X, Harish, N, Lewis, NE, Pan, S, Cai, Z, Liu, X, Chen, W, Xie, M, Wang, W, Hammond, S, Andersen, MR, Neff, N, Passarelli, B, Koh, W, Fan, HC, Wang, J, Gui, Y, Lee, KH, Betenbaugh, MJ, Quake, SR, et al. 2011, 'The genomic sequence of the (CHO)-K1 cell line', *Nature Biotechnology*, vol 29, no. 8, pp. 735-742.

Xu, J & Zhang, N 2014, 'On the way to commercializing plant cell culture platforms for biopharmaceuticals: present status and prospect', *Pharmaceutical Bioprocessing*, vol 2, no. 6, pp. 499-518.

Xu, J & Zhang, N 2014, 'On the way to commercializing plant cell culture platforms for biopharmaceuticals: present status and prospects', *Pharmaceutical Bioprocessing*, vol 2, no. 6, pp. 499-518.

Zhang, P, Woen, S, Wang, T, Liau, B, Zhao, S, Chen, C, Yang, Y, Zhiwie, S, Wormald, MR, Yu, C & Rudd, PM 2016, 'Challenges of glycoprotein analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs', *Drug Discovery Today*, vol 21, no. 5, pp. 740-765.

Zitzmann, J, Sprick, G, Weidner, T, Schrieber, C & Czermak, P 2017, 'Process Optimization for Recombinant Protein Expression in Insect Cells', in SJT Gowder (ed.), *New Insights into Cell Culture Technology*.