Production of biopharmaceutical proteins by yeast
Advances through metabolic engineering

Jens Nielsen
Novo Nordisk Foundation Center for Biosustainability; Chalmers University of Technology; Gothenburg, Sweden; Novo Nordisk Foundation Center for Biosustainability; Technical University of Denmark; Hørsholm, Denmark; Science for Life Laboratory; Stockholm, Sweden

Keywords: Saccharomyces cerevisiae, systems biology, secretory pathway, insulin, industrial biotechnology

Production of recombinant proteins for use as pharmaceuticals, so-called biopharmaceuticals, is a multi-billion dollar industry. Many different cell factories are used for the production of biopharmaceuticals, but the yeast Saccharomyces cerevisiae is an important cell factory as it is used for production of several large volume products. Insulin and insulin analogs are by far the dominating biopharmaceuticals produced by yeast, and this will increase as the global insulin market is expected to grow from USD12B in 2011 to more than USD32B by 2018. Other important biopharmaceuticals produced by yeast are human serum albumin, hepatitis vaccines and virus like particles used for vaccination against human papillomavirus. Here is given a brief overview of biopharmaceutical production by yeast and it is discussed how the secretory pathway can be engineered to ensure more efficient protein production. The involvement of directed metabolic engineering through the integration of tools from genetic engineering, systems biology and mathematical modeling, is also discussed.

The introduction of generic engineering by Cohen and Boyer in 1973 laid the foundation for the current biotech industry, which is based on using microorganisms or cell cultures for production of proteins that can serve as pharmaceuticals, often referred to as biopharmaceuticals. A few years later, researchers at Genentech cloned the genes for human insulin and growth hormone, and expressed them in E. coli, hereby demonstrating the utility and applicability of genetic engineering in creating genetically engineered bacteria that produce these two human proteins. In 1982 this led to marketing of the first biopharmaceutical, human insulin, by Eli Lilly, who licensed the technology from Genentech. In 1985 Genentech received FDA approval to market their own first product, Protropin, the human growth hormone, and to express them in E. coli. This enabled the production of a human-like N-glycosylation pattern and enhanced the biological activity of the hormone. Further advances were seen with the production of human growth hormone, and its analogs, for clinical applications. In 1986, Genentech launched human insulin produced by the yeast Saccharomyces cerevisiae as a replacement for their human insulin enzymatically derived from porcine insulin. Shortly following this early development, many other products were launched and today there are more than 300 biopharmaceutical proteins and antibodies on the market with sales exceeding USD100B, with monoclonal antibodies representing the majority (USD188B) followed by hormones (USD11B) and growth factors (USD10B). Furthermore, biopharmaceuticals have the fastest growth in the market with an annual growth of about 19%, and there are currently more than 240 monoclonal antibodies and 120 recombinant proteins in clinical trials.

About 40% of the biopharmaceuticals are currently being produced by mammalian cell cultures, mainly using Chinese Hamster Ovarian cell lines (CHO cells), as these allow for production of proteins with very similar glycosylation patterns as human proteins. E. coli is used as cell factory for production of another 30% of the biopharmaceuticals whereas 20% are being produced by S. cerevisiae. The dominant biopharmaceuticals produced by S. cerevisiae are insulin (and insulin analogs), human serum albumin, hepatitis vaccines and virus like particles, e.g., for vaccination against human papillomavirus. The advantages of using yeast S. cerevisiae as a cell factory for the production of biopharmaceuticals are that this eukaryal model system enables production and proper folding of many human proteins. Furthermore, the proteins can be secreted to the extracellular medium and this facilitates subsequent purification. A further advantage is that in many cases yeast can perform proper post-translational modifications of the protein, including proteolytic processing of signal peptides, disulphide bond formation, subunit assembly, acylation and glycosylation. S. cerevisiae is also widely used as an eukaryal model organism and there is therefore much information available about this organism through high-throughput studies, databases, sequenced genomes and extensive tools for molecular modification, which provides an extensive knowledge base for further engineering of this organism. One of the limitations with the use of yeast is, however, that it performs high-mannose type N-glycosylation. This confers a short half-life of the modified protein in vivo, which then can have a reduced efficacy for therapeutic use. Much work has been performed on engineering yeasts, both S. cerevisiae and Pichia pastoris, so that they can carry out human-like N-glycosylation patterns that even includes terminal addition of sialic acid to the
glycoprotein. This has opened for an even wider use of S. cerevisiae as a cell factory for production of biopharmaceuticals and there is therefore much interest in further engineering of yeast for ensuring efficient production of recombinant proteins.

More than 40 different recombinant proteins have been expressed, produced and secreted by S. cerevisiae. This includes several biopharmaceuticals and Table 1 provides an overview of some of these products, i.e., the protein name, their therapeutic application, leader sequence used and the titer reported in the publically available literature. As the table illustrates there are basically used three different types of leader sequences to ensure efficient secretion of the protein through the secretory pathway. S. cerevisiae only secretes few proteins to the extracellular medium, with the α-factor (a yeast hormone involved in mating) being the most studied and therefore most frequently used for efficient secretion of recombinant proteins. To further improve protein secretion in yeast Kjeldsen and coworkers at Novo Nordisk developed a synthetic leader that has been shown to be very efficient in protein secretion. As illustrated in Table 1, yeast can, however, also secrete human proteins that are expressed with their native leader sequences.

The secretory pathway in yeast is quite complex (see Fig. 1 for a schematic overview) as it involves more than 160 proteins that are responsible for different post-translational processes, e.g., folding and glycosylation. The secretory pathway handles more than 550 proteins that have a signal peptide in the yeast proteome, but only very few of these proteins are secreted to the extracellular matrix as all proteins targeted to the endoplasmic reticulum (ER), Golgi, vacuole and cytoplasmic membrane are also processed through the secretory pathway. Protein folding in the ER is of key importance for the secretory pathway as the accumulation of misfolded proteins results in ER stress that is handled then by the unfolded protein response (UPR). Activation of the UPR is a key control mechanism that results in increased production of proteins involved in regulating the ER stress pathway. This is achieved by transcriptionally activating the Hac1p transcription factor. A result of this regulation is upregulation of chaperones and foldases as well as ER associated degradation (ERAD). Based on studies of the UPR many targets for improving protein secretion have been identified and implemented, and it is generally believed that any factor that reduces ER stress and its downstream damage caused by heterologous protein production, results in improved secretion of the produced protein.

Due to the complexity of the protein secretion pathway there has traditionally been focus on transcriptional regulation of protein production. A large number of different promoters have been evaluated for driving expression of the heterologous genes, and among the most widely used promoters are strong glycolytic promoters like pETIH5, pGAL1 or pαHI, pαHI5 or galactose induced promoters like pGALI1, pGAL17 or pGAL10. However, even though changing expression may improve production of one protein, it does not imply there would be improvement in production of another protein, as clearly demonstrated in our recent study on production of insulin precursor and α-amylase. In this study we used an expression system originally developed by Novo for their insulin production. This expression system involves deletion of the TPI1 gene in the chromosome and use of the corresponding gene, pTPI1, from Schizosaccharomyces pombe as a plasmid marker (see Fig. 2). This gives a very stable construct as all cells losing the plasmid will be deficient in triose-phosphate isomerase activity, a key glycolytic enzyme. A particular strength of this method is that it is stable also with use of complex media containing amino acids and nucleotides, which is not the case with standard yeast auxotrophy markers. Using this vector system we evaluated two different promoters (pTPI1 and pPEF1) as well as two different leader sequences (synthetic leader and the α-factor leader) for production of insulin precursor and α-amylase, and found that there was considerable differences in production of the two proteins with the different expression systems evaluated, i.e., with low gene expression α-amylase was produced at higher levels, whereas for high gene expression system insulin precursor was produced at higher levels. This pointed to very different protein processing in the ER, i.e., the larger and more complex α-amylase is more challenging for ER-processing when the flux is high. This hypothesis was confirmed by a measured increased ER-stress (by gene expression analysis) in the α-amylase producing strains. Based on this analysis it was hypothesized that engineering of the downstream secretion pathway may be able to improve the secretion of amylase, and indeed overexpressing regulators of the so-called SNARE complex, Sec71p and Sly1p, resulted in improved protein secretion. Moreover, it was found that overexpression of SEC71, that is involved in regulating vesicle trafficking from Golgi to the cell membrane, resulted in improved production of both insulin precursor and α-amylase, whereas overexpression of SLY1, that is involved in regulating the vesicle fusion from ER to Golgi, increased only the α-amylase production. Through combined overexpression of the Sec71p and Sly1p the overall secretion of α-amylase could be improved by about 70%, whereas insulin precursor production was increased by about 30%, and the study therefore clearly demonstrated that engineering of the secretory pathway can result in significant improvement of recombinant protein production.

There are many other studies that clearly demonstrate that engineering of the secretory pathway can result in improved

Table 1. Overview of some biopharmaceuticals produced by S. cerevisiae

| Type          | Protein                  | Therapeutic application       | Leader sequence | Titer     |
|---------------|--------------------------|-------------------------------|-----------------|-----------|
| Blood related | Human Serum Albumin     | Surgery (plasma expander)     | Native          | 3 g/L     |
|               | Hirudin                  | Blood coagulation disorders   | α-Factor        | 460 mg/L  |
| Hormones      | Human transferin        | Anemia                        | Native          | 1.8 g/L   |
|               | Insulin Precursor       | Diabetes                      | Synthetic       | 80 mg/L   |
|               | Glucagon                |                               | α-Factor        | 17.5 mg/L |
| Antigen       | Hepatitis surface antigen | Hepatitis vaccination         | Native          | 19.4 mg/L |
of metabolic engineering also results in improved insight of the cellular metabolism and physiology, but generally to a less extend protein production,12 but a general finding is that most targets are quite specific, depending which particular protein is being overexpressed. Ideally, there would be one (or few) efficient yeast platform strain(s) that can serve as production host for a wide range of different biopharmaceuticals. There is therefore a need for a more rational approach to engineering yeast for improved protein secretion. In such an approach detailed models of the secretory pathway can be used for design, similarly as it has been done with engineering of metabolic pathways, where the use of genome-scale metabolic models has been shown to be of great importance.15,24,25 This approach is referred to as metabolic engineering, and the workflow, often called the metabolic engineering cycle,26 is illustrated in Figure 3.

As illustrated, the workflow involves detailed modeling, often based on detailed analysis of the cellular metabolism and physiology using high-throughput experimental techniques developed in the field of genomics, and the concept of quantitatively describing cellular processes with mathematical models is at the core of systems biology.27 The metabolic engineering cycle is therefore very similar to the workflow of many systems biology studies where perturbation of the cellular system is performed using genetic engineering, e.g., by overexpression or deletion of specific genes, followed by detailed analyses that can be used to define a mathematical model of the biological system. However, there is a major difference in the sense that systems biology is a fundamental science where the primary objective is to gain novel insights, whereas metabolic engineering is an applied science with the primary objective to obtain an improved cell factory. Clearly the process of metabolic engineering also results in improved insight of the cellular metabolism and physiology, but generally to a less extent.

Figure 1. Schematic overview of the secretory pathway in yeast. Proteins targeted for secretion enter the endoplasmic reticulum (ER). If they fold correctly they can enter the secretory pathway, whereas misfolded protein cause ER stress leading to the activation of the unfolded protein response (UPR) that results in activation of a very large number of cellular processes, including activation of chaperones and foldases (like BIP and PDI) that assist with refolding. UPR is also upregulating ER-associated degradation (ERAD) where the unfolded proteins are exported from the ER, ubiquitinated and hereby targeted for degradation by the proteasome (ubiquitin-proteasome system, UPS). Correctly folded proteins can be exported to the Golgi for further processing (including additional glycosylation). The COP1- and COP8-complexes facilitate the ER-Golgi transfer, and from the Golgi the protein may be secreted via the endosome or be targeted to the vacuole for storage and/or degradation. Different colors represent different types of vesicular compartments of the secretory pathway.

Figure 2. Illustration of the stable expression system with a glycolytic gene as the selection marker. One of the glycolytic enzymes is used as a marker for plasmid presence: the endogenous gene encoding triosephosphate isomerase (TPI1) is deleted and the corresponding gene (POT1) from Schizosaccharomyces pombe is expressed from a plasmid. The same plasmid carries the gene for the heterologous gene to be expressed (here demonstrated with a gene encoding human insulin). If the plasmid is lost the cells lack a key glycolytic enzyme and the glycolytic flux is therefore reduced dramatically resulting in impaired growth. Cells that are replicating the plasmid in high copy numbers and expressing the genes from the plasmid therefore have an inherent growth advantage.
The metabolic engineering cycle is very similar to the workflow of many systems biology studies where perturbation of the cellular system is performed using genetic engineering, e.g., by overexpression or deletion of specific genes, followed by detailed analysis that can be used to define a mathematical model for the biological system.
21. Patil C, Walter P. Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. Curr Opin Cell Biol 2001; 13:545-55; PMID:1143987; http://dx.doi.org/10.1016/S0952-0009(00)00219-2.

22. Lin Z, Tyo KEJ, Martinez J, Petranovic D, Nielsen J. Different expression systems for production of recombinant protein in Saccharomyces cerevisiae. Biotechnol Bioeng 2012; 109:2179-88; PMID:22040883; http://dx.doi.org/10.1002/bit.24899.

23. Hou J, Tyo KEJ, Lin Z, Petranovic D, Nielsen J. Engineering of vesicle trafficking improves N-linked protein secretion in Saccharomyces cerevisiae. Metab Eng 2012; 14:128-37; PMID:22309423; http://dx.doi.org/10.1016/j.mib.2012.01.002.

24. Bo C, Bregnhøj R, Finner J, Nielsen J. In silico aided metabolic engineering of Saccharomyces cerevisiae for improved ethanol production. Metab Eng 2006; 8:182-91; PMID:16280774; http://dx.doi.org/10.1016/j.mib.2005.09.017.

25. Tyo KEJ, Kucharik K, Nielsen J. Toward design-based engineering of industrial microbes. Curr Opin Microbiol 2010; 13:275-82; PMID:20260723; http://dx.doi.org/10.1016/j.mib.2010.02.001.

26. Nielsen J. Metabolic engineering. Appl Microbiol Biotechnol 2001; 55:263-83; PMID:1143987; http://dx.doi.org/10.1007/s002530000511.

27. Nielsen J. Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. FEMS Yeast Res 2008; 8:123-31; PMID:17727895; http://dx.doi.org/10.1111/j.1567-1364.2007.00302.x.

28. Umaña P, Bailey JE. A mathematical model of N-linked glycoform biosynthesis. Biotechnol Bioeng 1997; 55:890-908; PMID:18636599; http://dx.doi.org/10.1002/(SICI)1097-0290(19970920)55:6<890::AID-BIT7>3.0.CO;2-B.

29. Krambeck FJ, Betenbaugh MJ. A mathematical model of N-linked glycosylation. Biotechnol Bioeng 2005; 92:691-700; PMID:15924151; http://dx.doi.org/10.1002/bit.20645.

30. Mjøen P, Kjensmo K, Nielsen J. Fifteen years of large-scale metabolic modeling of yeast: developments and impacts. Biotechnol Adv 2012; 30:979-88; PMID:21846501; http://dx.doi.org/10.1016/j.bio technol.2011.07.021.

31. Patil C, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci U S A 2005; 102:2685-9; PMID:15671683; http://dx.doi.org/10.1073/pnas.0406811102.