Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation

José L. Martínez\textsuperscript{a,1}, Lifang Liu\textsuperscript{a,1}, Dina Petranovic\textsuperscript{1} and Jens Nielsen\textsuperscript{1,2}

Addresses:

1\textsuperscript{1}Novo Nordisk Center for Biosustainability, Department of Chemical & Biological Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden

2\textsuperscript{2}Novo Nordisk Center for Biosustainability, Technical University of Denmark, DK2970 Hørsholm, Denmark

Corresponding author: Nielsen, Jens (nielsenj@chalmers.se)

These authors contributed equally to this article.
Abstract

Since the approval of recombinant insulin from *Escherichia coli* for its clinical use in the early 80’s, the amount of recombinant pharmaceutical proteins obtained by microbial fermentations has significantly increased. The recent advances in genomics together with high through-put analysis techniques (the so-called –omics approaches) and integrative approaches (systems biology) allow the development of novel microbial cell factories as valuable platforms for large scale production of therapeutic proteins. This review summarizes the main achievements and the current situation in the field of recombinant therapeutics using yeast *Saccharomyces cerevisiae* as a model platform, and discusses the future potential of this platform for production of blood proteins and substitutes.

Introduction

Microorganisms have been extensively used since ancient times for the production of fermented food and beverages, thousands of years before the actual nature of the fermentative processes was known. In the early 20th century the production of citric acid based on microbial fermentation was initiated as the first large scale fermentation product and this was followed by industrial production of penicillin as the first antibiotic. Introduction of the genetic engineering in the 70’s paved the way for the establishment and development of the current biotech industry, allowing the commercial production of industrial enzymes and biopharmaceutical proteins. In 1980, the FDA approved for clinical use the recombinant insulin obtained from *E. coli*, becoming the first recombinant pharmaceutical protein to enter the market [1]. Since then, the biotechnology industry has grown substantially, and currently about 25% of commercial pharmaceuticals are biopharmaceuticals [2] with 2010 sales exceeding USD100 billions [3]. About half of the world-wide sales are in the USA with monoclonal antibodies representing the majority (>USD18 billions) followed by hormones (USD11 billions) and growth factors (>USD10 billions)[4]. Together with the production of industrial enzymes, the recombinant protein production market is expected to rise to 169 billion dollars in 2014 [3] (Figure 1).

Platforms for production of pharmaceutical proteins

Industrial biotechnology has traditionally used numerous bacterial and eukaryal cells as production platforms, with the main criterion for host selection being the ability to produce the desired compound. However, with the advent of genetic engineering it became possible to introduce heterologous genes and create new traits in non-natural producers, allowing the development of cell factories for the production of chemicals through metabolic engineering. *E. coli* was the earliest
platform to be exploited, and is still nowadays the most used production platform for recombinant proteins [5], covering approx. 30% of the total production of recombinant proteins [1] (Figure 1). In general terms, bacteria have been considered to be the most efficient producers of heterologous proteins due to several reasons: i) well developed molecular tools for genetic manipulation, ii) annotated genomes and metabolic pathways, iii) high cell density cultivation capacity and growth rate and iv) high yield of recombinant proteins, up to 80% of its dry weight [6-8]. However, standard prokaryotic systems have some limitations for production of human proteins. For example, bacteria are unable to perform some of the complex post-translational modifications [1], which itself represents a limitation, since many proteins require further processing to become fully active. In particular glycosylations that are needed to ensure proper function and activity, by influencing proper charge, solubility, folding, serum half live of the protein, in vivo activity, correct cellular targeting and immunogenicity, among others, cannot be often be fully accomplished in bacterial systems [9] [10]. These limitations have paved the way towards eukaryotic expression systems and there exists several eukaryotic systems that are currently in use for large scale production of different therapeutic proteins (Table 1), with the most studied being hybridoma cells, Chinese Hamster Ovary (CHO) cells [11], insect cells [12,13] and yeast cells [14]. Mammalian systems like hybridoma and CHO cells clearly have the highest similarity to human cells, and proteins produced by these systems are often properly folded and glycosylated. However the costs for their cultivation are high (e.g. expensive media and growth factors, contaminations with microorganisms and viruses), they have a limited secretion capacity and protein yields are usually low [15]. On the other hand fungal expression systems, and in particular yeast, can grow in relatively cheap and defined media, decreasing the production costs. Besides, they are not so susceptible to contaminations and in addition, the yeast cells are less sensitive since the wall makes them more resistant to shear stress during the production process [10]. Yeast expression systems also provide higher protein titers (>1g/l) in fermentation processes that even last shorter time (only few days) [16]. Based on this, we propose yeast as an attractive choice, and recent advances in genetic and metabolic engineering, and tools in genomics and systems biology could make S. cerevisiae a preferred production platform for a range of pharmaceutical proteins[17]. However, even though yeasts are eukaryotic systems, the glycosylations of proteins may differ substantially from that performed by mammalian cells, a difference that can be in some cases detrimental for its subsequent therapeutic use. N-glycosylation in yeast, for example, is of the high mannose type whereas human N-glycans are mainly of the complex or hybrid type. In addition to N-glycosylation, yeast O-glycosylation characterized by shorter glycan structures, also differs from the human type, which is mucin-type in contrast to the oligomannosyl-glycans in yeasts [10]. Very promising attempts have been recently achieved to introduce human glycosylation patterns in yeast (humanized yeast platforms). To date, only Pichia
species have been successfully engineered to produce specific human-like glycoforms of proteins, however recent advances reported in this field in *S. cerevisiae* seem to indicate that, with further development, it may increase the array of strains available that are able to produce human-type glycosylated proteins, and these strains will become a valuable platform for the production of glycoproteins for therapeutic use [10,18,19].

How to make *Saccharomyces cerevisiae* a better producer of pharmaceutical proteins?

The technology for industrial production of recombinant pharmaceutical proteins in *S. cerevisiae* is well established and currently applied for production of human insulin, hepatitis virus vaccines and human papilloma virus vaccines, and its potential to be used for large scale production of many other proteins in the forthcoming years is therefore high. Furthermore, the advent of systems biology allowing global metabolism analysis and the application of so-called “omics” approaches such as transcriptome, proteome and metabolome data, facilitates the identification of the bottlenecks and factors limiting the full potential of this yeast to become a better producer [20,21], and consequently the application of metabolic engineering to overcome constraints in productivity could definitely allow the establishment of *S. cerevisiae* as a suitable platform for large scale production of heterologous (including human) proteins [22].

There are several reports describing how either genetic or metabolic engineering can be successfully performed in *S. cerevisiae* [22,23] resulting in the generation of strains showing an enhanced production capacity of heterologous proteins [17,24,25]. Often just introducing an entire new pathway for the production of the desired compound does not result in high levels of production, because protein folding and secretion can represent the major limitation in terms of protein yields in yeast [20]. Folding and secretion are complex processes and the molecular machineries are composed of large number of components, so further modifications and development of these pathways requires integrative analysis of the whole secretory pathway. Such approach has been successfully carried out by engineering different elements of the secretory pathway, and by combining different expression systems in order to optimize the production of several kinds of different proteins showing different biochemical properties (i.e. size, type of modification (glycosylation and/or disulfide bond formation)), such as human insulin precursor or α-amylase [17,26,27]. Through a combination of these approaches it is possible to select the best protein producers for further optimization, and this may lead to generic protein producing strains that can be used as general platforms for the production of bio-based pharmaceutical proteins (Figure 2).

Production of Recombinant Human Blood Proteins
Among the 58 biopharmaceuticals approved in the United States and/or Europe from 2006 to 2010, four are blood related proteins, including a rh coagulation factor VIII produced in CHO cells, a rh antithrombin from milk of transgenic goats, a plasma kallikrein inhibitor produced in *Pichia pastoris*, and a rh thrombin produced in CHO cells [28]. All have therapeutic use for treatment of hemophilia.

To date, most of the recombinant blood related biopharmaceuticals approved for clinical treatment are coagulation factors, including factor VIII, factor VIIa, and factor XI [28] with recombinant human serum albumin (rHSA) as an exception. Due to the fact that it is not glycosylated, a variety of expression hosts have been screened to express rHSA, including bacterium *Bacillus subtilis*, yeast *S. cerevisiae*, *Kluuyveromyces lactis* and *P. pastoris* etc [29]. So far, Novozymes has commercialized two rHSA produced by *S. cerevisiae*, namely Recombumin and Albacult. Since the clinical dosage of HSA is usually quite high, normally over 10g/L, many studies have tried to express rHSA also in *P. pastoris* for its high capacity in heterologous protein production [30]. The rHSA produced from *P. pastoris* has gone through the clinical trials and confirmed the safety and efficacy to treat different diseases [31].

Aside from rHSA, many other human blood proteins are also under active studies [32-34]. Human fibrinogen (Hf) is a large plasma glycoprotein that plays a critical role in the last stage of coagulation. It is dimeric and comprised of two sets of three different polypeptides, namely Aα, Bβ and γ. The protein was expressed in *P. pastoris* protease deficient strain by constructing an expression vector containing the cDNA of three individual peptide chains. Even though the peptides expressed were of different N-glycosylation patterns as that of native Hf, they were correctly assembled to a functional rHf that is capable of forming a clot in the presence of factor XIIIa [32]. Human α1-antitrypsin (hAAT) was produced in tomato: the codon modified cDNA sequence was expressed and the mRNA 5’ and 3’ flanking regions were modified to achieve a high-level expression by eliminating mRNA destabilizing sequences, which are ATTTA and its variants, splice sites and A/T strings. In contrast to the unglycosylated rAAT expressed in *E. coli*, the glycosylated rATT from transgenic tomato was biologically active [33]. One more example is the production of human transferrins (Tf) which are a family of monomeric proteins that are of different sizes depending on the extent of glycosylation. Besides its central role to facilitate iron transport and metabolism, a lot more other functions have been evidenced, e.g. acting as a growth factor for mammalian tissue cells, as a neurotropic factor during neural stem cell development and as an angiogenic factor to promote endothelial cell migration etc, enabling a development of many novel practical applications in medicine [34]. To date, several heterologous systems including *E. coli*, yeast, transgenic plants, mammalian and insect cells have been developed for rHf production [34], among which *E. coli* was reported to be inefficient due to the production of inactive hTf. Successful expressions of hTf were reported in *S. cerevisiae* and *P.
pastoris with the resulting proteins being functional independent of use of only a partial sequence or a full sequence, or with mutated glycosylation sites [34].

S. cerevisiae as a cell factory for human hemoglobin production

All the proteins described above are blood plasma components which contribute to different roles of blood such as coagulation, clotting, transport of iron, maintain blood osmotic pressure and blood volume. The additional crucial role of blood is the transport of oxygen and the only component in blood that possesses oxygen carrier function is hemoglobin (found in erythorocytes), and this is therefore a key component for development of human blood substitutes for treatment of patients with injuries, anemia or in post-operational recovery.

Production of recombinant hemoglobin (rHb) has been attempted since the late 80’s. A variety of strategies have been applied using several different expression systems, ranging from bacteria to higher organisms such as transgenic plants and animals [35]. The work was first done in E. coli, where a single β globin was expressed with a cleavable linker and refolded in vitro with native α globin and exogenous heme [35]. The work was quite laborious and researchers therefore tried to express α and β globins simultaneously in vivo with endogenous heme incorporated [36]. It was observed that the essential parameters for normal human hemoglobin, namely Bohr effect and 2,3-BPG effects of the rHb were reduced which very likely resulted from the methionine termini at the end of the globin chains [37]. The amount of methionine modified α and β globins were significantly reduced by co-expression of the methionine amino-peptidase (Met-AP) gene with the globin genes resulted in an increased yield of rHb [36]. Further optimizations of the E. coli expression system included codon optimization for globin expression in a T7 promoter system [38]; Site-directed mutations in β globin chain in order to reduce the extreme oxygen affinity to rHb (no release of oxygen) due to the lack of 2, 3-BPG allosteric regulation; and a tandem fused α globin to prevent the dissociating of the tetramer into a αβ dimer [39]. In some resent studies, researchers have tried to co-express α globin with its molecular chaperon, α-hemoglobin stabilizing protein (AHSP) and revealed its mechanisms on preventing α globin precipitation. [40-42]. Correct expression and folding of human Hb have also been accomplished in animals, e.g. pig [43] and mice [44,45], and in plants, e.g. tobacco[46].

The previous examples and the state-of-the-art methodologies and approaches show that S. cerevisiae can be engineered to become an even better producer for a wider range of pharmaceutical and blood proteins. Comparing to E. coli, heterologous proteins produced in S. cerevisiae do not have methionine modification which affecting the biological function of the rHb.
Compared to plant and animal expression systems, the yeast system is cheaper and faster to manipulate. As an outlook, we propose several potential strategies for increasing human hemoglobin production in *S. cerevisiae* e.g. globin folding, heme uptake, and subunit assembling.

Additionally to the existing strategies such as site directed mutagenesis of the amino acids that are important for stabilization [47,48], co-overexpression of α and β globin genes, cross-linking two α globins as di-α to avoid αβ dimer formation, it is worth trying to co-express the AHSP gene together with the α and β globin genes to increase the α globin stability [42,49,50] thus further enhancing the production of rHb. It has also been reported that heme can accelerate hemoglobin accumulation in immature cultured erythroid cells [51] and heme is not only the indispensable prosthetic group but also essentially involved in assembling and ensuring a stable tetramer structure [52]. As heme supplementation is expensive and the mechanism behind is poorly understood [53], engineering the heme synthesis pathway is proposed to be a better choice to increase heme levels in the cell.

Examples for engineering the heme biosynthesis pathway include over-expression of the rate-limiting enzymes in the synthesis pathway [54], engineering ALA (aminolevulinic acid) synthesis since it is the first intermediate involved in heme synthesis, as well as engineering the iron uptake pathway.

In conclusion, the recent advances in the field of metabolic engineering allowed that *S. cerevisiae* become an efficient cell factory for the production of heterologous proteins. By a systems biology approach, further improvements might be implemented through integrative analysis and the development of mathematical predictive models, being this yeast expected to become the suitable platform for sustainable large scale production of protein therapeutics in the forthcoming future.

**Acknowledgments**

We would like to thank Zihe Liu and Dr. Jin Hou for suggestions and comments on the manuscript. This work has been funded by the Chalmers Foundation, European Research Council project INSYSBIO (Grant no. 247013) and the Novo Nordisk Foundation.
References and recommended reading

Papers of particular interest have been highlighted as:

* of special interest

** of outstanding interest


Figure 1. (A) Global market for recombinant protein drugs and (B) percentage of protein-based recombinant pharmaceuticals, produced by different systems [1].
Figure 2. Workflow overview for the development of yeast cell factories by metabolic engineering from a systems biology approach. Engineering for protein production improvement can be implemented either at a host level (e.g. metabolic engineering of gene pathways related to different steps concerning protein processing and secretion) and/or be applied to the improvement of expression systems by addition/testing of different features (e.g. the suitable selection marker for each system, promoter sequences, etc). High throughput analysis methodologies allow then the generation or large data sets, which can be processed and integrated in mathematical models for the identification of new potential targets, allowing further improvement via retrofitting the system, and therefore resulting in an enhanced protein production capability of the cell platform.

<table>
<thead>
<tr>
<th>Protein</th>
<th>System</th>
<th>Production level</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirudin</td>
<td><em>S. cerevisiae</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>60 mg/L</td>
<td>[55]</td>
</tr>
<tr>
<td></td>
<td><em>H. polymorpha</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>-</td>
<td>[56]</td>
</tr>
<tr>
<td>Interferon α-2b</td>
<td><em>H. polymorpha</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>120 mg/L</td>
<td>[57]</td>
</tr>
<tr>
<td>Hepatitis B vaccine</td>
<td><em>H. polymorpha</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>-</td>
<td>[58]</td>
</tr>
<tr>
<td>Angiostatin</td>
<td><em>P. pastoris</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>108 mg/L</td>
<td>[59]</td>
</tr>
<tr>
<td>Anti-HBs Fab</td>
<td><em>P. pastoris</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>50 mg/L</td>
<td>[60]</td>
</tr>
<tr>
<td>Human serum albumin</td>
<td><em>K. lactis</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>3 g/L</td>
<td>[61]</td>
</tr>
<tr>
<td></td>
<td><em>S. cerevisiae</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>3 g/L</td>
<td>[62]</td>
</tr>
<tr>
<td></td>
<td><em>P. pastoris</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>10 g/L</td>
<td>[63]</td>
</tr>
<tr>
<td>Human interleukin 6</td>
<td><em>A. niger</em> &lt;sup&gt;(F)&lt;/sup&gt;</td>
<td>150 mg/L</td>
<td>[64]</td>
</tr>
<tr>
<td>Human apolipoprotein Al</td>
<td>CHO cells &lt;sup&gt;(M)&lt;/sup&gt;</td>
<td>80 mg/ml</td>
<td>[65]</td>
</tr>
<tr>
<td>Insulin precursor</td>
<td><em>P. pastoris</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>3 g/L</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td><em>S. cerevisiae</em> &lt;sup&gt;(Y)&lt;/sup&gt;</td>
<td>98 mg/L</td>
<td>[27]</td>
</tr>
<tr>
<td>Human tPA</td>
<td>CHO cells &lt;sup&gt;(M)&lt;/sup&gt;</td>
<td>34 mg/L</td>
<td>[67]</td>
</tr>
<tr>
<td>Human gonadotropin</td>
<td>CHO cells &lt;sup&gt;(M)&lt;/sup&gt;</td>
<td>3 g/L</td>
<td>[67]</td>
</tr>
<tr>
<td>Erythropoietin (epoetin α)</td>
<td>CHO cells &lt;sup&gt;(M)&lt;/sup&gt;</td>
<td>-</td>
<td>[68]</td>
</tr>
<tr>
<td>Monoclonal Ab</td>
<td>NSO cells (M)</td>
<td>3 g/L</td>
<td>[67]</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>HPV vaccine (Cervarix™)</td>
<td>Insect cells</td>
<td>-</td>
<td>[69]</td>
</tr>
<tr>
<td>Human proapolipoprotein Al</td>
<td>Insect cells</td>
<td>80 mg/L</td>
<td>[70]</td>
</tr>
<tr>
<td>Clotting factor VII a</td>
<td>BHK cells (M)</td>
<td>-</td>
<td>[68]</td>
</tr>
</tbody>
</table>

Table 1. Examples of recombinant therapeutic proteins successfully expressed using different production systems, including highest production levels reported in each organism. CHO = Chinese hamster ovary cells; BHK = Baby hamster kidney cells; NSO = Myeloma cells; (Y) = Yeast; (F) = Filamentous fungi; (M) = Mammalian.

Highlights

Recombinant therapeutic production is a multibillion dollar market.

*E. coli* represents 30% of recombinant protein production but not suitable for human therapeutics.

Eukaryotic systems other than yeast are costly or not so efficient regarding protein yields.

*S. cerevisiae* shows a high potential to be a suitable platform for therapeutic protein.

Human blood proteins are the next candidates to be challenged by *S. cerevisiae* system.