Review

by

Biopharmaceutical protein production by Saccharomyces cerevisiae: current state and future prospects

In the past few decades there has been an increasing demand of biopharmaceutical proteins in the market. Several types of cell factories are applied to produce different pharmaceutical proteins. However, manufacturers prefer to use a few favorable biological platforms to undertake the production tasks with low cost, high productivity and proper post-translational modifications. The yeast *Saccharomyces cerevisiae* is one of these preferred cell factories as it meets many of the requirements. There are several reports on improvement of recombinant protein production by *S. cerevisiae* through rational engineering of different stages of the protein secretion pathway. However, recent developments of new technologies like systems biology and synthetic biology open new doors to design *S. cerevisiae* as an ideal production platform.

Market demand, policy support

In the past half century great progress in our ability to produce advanced medicines has contributed to improving the health and longevity of people. Pharmaceutical proteins represent one of the fast growing groups of medicines and currently play an important role in treatment of many diseases [1]. Molecular biology techniques, developed in the 1970s, open a new gate to production of pharmaceutical proteins (often referred to as biopharmaceuticals) as these proteins could be produced by cell factories in a correct form, in high purity and in a scalable fashion [2,3]. Following the first biopharmaceutical, human insulin, which was approved in the market in 1982, hundreds of biopharmaceutical proteins have been launched over the past three decades and many more are under clinical trials [2]. Biopharmaceuticals represent 25% of commercial pharmaceuticals and account for approximately 40% of the total pharmaceutical sales [45]. Total biopharmaceutical sales reached approximately US\$125 billion in 2012, which doubled from approximately \$64 billion in 2006 [6]. Compared with sales in 2011 most of the biopharmaceuticals have achieved significant growth in sales. The top three categories, anti-TNF antibodies, cancer

antibodies and insulin and insulin analogs, represented more than half of the sales with a 10–20% annual growth rate (Figure 1) [7]. With the knowledge and technology development, less conservative attitudes towards biopharmaceuticals from the US FDA and European Medicines Agency have also stimulated the biopharmaceutical industry.

Platforms for production of pharmaceutical proteins

In theory, from prokaryotes to eukaryotes, from single cell to multicell organisms, including bacteria, yeasts, insect cells, mammalian cells, plants and animals, there are many expression systems that can be employed for production of recombinant proteins for different applications [8,9]. However, from an industrial point of view, in consideration of product quality, production timescale, scale-up capacity and downstream processing, mammalian cells, Escherichia coli and Saccharomyces cerevisiae are the most commonly used biopharmaceutical expression systems, by which 43%, 31% and 15% of biopharmaceuticals are produced, respectively [2,10]. In the last few years two biopharmacueticals produced by Pichia pastoris have been approved by US FDA, and several oth-

Mingtao Huang¹, Jichen Bao¹ & Jens Nielsen^{*,1,2}

¹Novo Nordisk Foundation Center for Biosustainability, Department of Chemical & Biological Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden ²Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Fremtidsvej 3, DK-2970 Hørsholm, Denmark *Author for Correspondence: Tel.: +46 31 772 3804 Fax: +46 31 772 3801 nielsenj@chalmers.se



Key Terms

N-glycosylation: Describes an important and conserved post-translational modification manner for protein; the glycan is linked to the carboxamido nitrogen on asparagine residue within the conserved motif (asparagine – X – serine/threonine) of the target protein, where X is any amino acid except proline.

Protein misfolding and aggregation: Many neurodegenerative diseases are caused by the accumulation of misfolded proteins, such as Alzheimer's disease, Parkinson's disease and Huntington's disease.

ers are in the pipeline to be approved, which makes *P. pastoris* an important alternative expression system for biopharmaceutical production [11].

Mammalian cells, which allow human-like N-glycosylation, can produce properly glycoproteins, which have correct function and show good pharmacokinetics *in vivo* [3,12]. However, these cells are not very resistant to bioprocessing, and the medium required for mammalian cell cultures is complex and more expensive than that for microorganisms [13]. *E. coli* was the expres-

sion system for the first FDA-approved rDNA pharmaceutical - human insulin in 1982. Simple medium, easy culture condition and high cell density cultivation combined with rapid growth and protein production rates makes E. coli a good expression system for production of biopharmaceuticals [14,15]. However, poor protein folding capacity and limitation in its secretion capacity limits its application in some cases [14,16]. S. cerevisiae, a single-cell eukaryotic organism, has bacteria and eukaryotes characteristics - it is easy to culture, grows fast, can give high productivity, can give high density fermentations [17], can ensure proper protein folding and post-translational modifications (PTMs) [18], and it can secrete the product to the extracellular medium which simplifies purification [4,19,20]. Moreover, as a generally recognized as safe (GRAS) organism, free of pyrogens also makes S. cerevisiae a favorable expression system for biopharmaceutical production [16]. Here, we will mainly discuss biopharmacecutical protein production by S. cerevisiae and methods for improving protein production properties of S. cerevisiae.



Figure 1. Sales of biopharmaceuticals in 2012. The size of the circles represents sales; the color of the circles represents the percentage change from 2011, red: <-5%; purple: -5% to 5%; blue: 5% to 20%; green: >20%. Data taken from [7].

Current state & recent advances in production of biopharmaceutical proteins by *S. cerevisiae*

Advantages by using *S. cerevisiae* for pharmaceutical proteins production

S. cerevisiae, the most widely used yeast for recombinant protein production, has a long history in industrial food application due to its GRAS status. It has the dual characteristic of being a unicellular and eukaryotic organism [21]. As a unicellular microbial organism, it is easy to culture, grows fast, has resistance and tolerance to chemical and secondary metabolites, and adjusts well to industrial processing [22]. It has complete sub-cellular organelles and membrane-bound compartments, including nucleus, endoplasmic reticulum, Golgi apparatus, vesicles, vacuoles, mitochondria and microbodies [10], and it is therefore able to correctly produce and fold many eukaryotic proteins (including human proteins) as well as perform proper PTMs of such proteins, e.g. proteolytic processing of signal peptides, disulfide bond formation, subunit assembly, acylation and glycosylation [2]. The PTMs are a key point for a biological active protein and it is as important as the correct amino acid sequences [23]. Proper PTMs will maintain correct secondary structures of proteins and/or their active catalytic sites. Another advantage of S. cerevisae for production of biopharmaceuticals is that the proteins can be secreted out by cells, which significantly reduces the costs for downstream purification [2]. By-products not only reduce the yields of desire protein by consuming substrate and energy, but accumulation of by-products during fermentation also may cause denaturation of biopharmaceuticals and increase purification costs. It is possible and necessary to maximize production of protein by optimization of fermentation processes in industrial manufacture [24]. Recently, a multivariate Bayesian predictive approach was presented for identification of critical process parameters and optimization of pharmaceutical protein production by S. cerevisiae [25].

S. cerevisiae has also developed as an eukaryotic model organism for studying diseases associated with **protein misfolding and aggregation**, as the protein-homeostasis networks and vesicular trafficking are conserved between yeast and humans [26,27]. *S. cerevisiae* was the first eukaryotic organism whose whole DNA was completely sequenced [28]. A huge amount of information about this organism is available, and a well-established molecular toolbox enables easy genetic modifications. Different promoters with a wide range of transcriptional activity are available and well characterized, including the strong glycolytic promoters like pTDH3, pPGK1, pTPI1 and pADH1, which will maintain a constitutive and higher transcriptional

level [2]. Unlike these constitutive expression promoters, conditional induced promoters like pGAL1, pGAL7, pGAL10, pPHO5 and pMET25 are also very useful for production of biopharmaceuticals, which hereby can be expressed in a regulated manner. For example, high-level expression of human serum albumin (HSA) and HSA-fusion proteins will be triggered by methionine depletion in the late log phase, if they are under control of pMET25 promoter [29]. This can enable separation of the growth phase and the protein production phase which may be beneficial for preventing unintentional selection of more rapidly growing, non-protein-expressing cells or for the production of toxic proteins [29].

A comparison of *S. cerevisiae* with other expression systems for biopharmaceuticals production based on their industrial and biological characteristics is made in Figure 2 [8,11,30]; and it is clear that by combining the advantages of bacteria and eukaryotic organisms, *S. cerevisiae* is an attractive expression platform for production of biopharmaceuticals.

Current state & recent advances

Dozens of pharmaceutical proteins, such as insulin, vaccines and blood factors, produced by S. cerevisae have entered the market [4]. As of June 2012 there are 23 European Medicines Agency-approved rDNA biopharmaceuticals produced by S. cerevisiae (Figure 3), and several of these are blockbusters with sales exceeding \$1 billion [7,31]. Being the first biopharmaceutical human insulin became a typically representative for the successful new biopharmaceutical industry. In 2012, the sales of insulin and its analogs reached \$18.92 billion [7], and the sales are dominated by Novo Nordisk, Eli Lilly and Sanofi-Aventis [32-34]. The leading producer Novo Nordisk (with a 44% market share) is using S. cerevisiae for production of insulin and insulin analogs; all of its four brands are blockbusters, Novolog®/NovoRapid®, Actrapid[®], Levemir[®] and Novomix[®] with sales of \$2.94, \$1.99, \$1.72 and \$1.65 billions of sales in 2012, respectively (Figure 4). For Eli Lilly and Sanofi-Aventis, E. coli is employed as a platform for production of their insulin or insulin analogs.

Even though *S. cerevisiae* is a good platform for biopharmaceuticals production, some of its properties could be improved to meet the commercial production requirements. The most direct and simplest thought is to increase the capacity of protein secretion of *S. cerevisiae*. From a biological point of view, protein secretion by *S. cerevisiae*, covering multiple steps in conversion of DNA coding sequences to mature proteins, is complex and involves many different levels of processing: transcription, translation, translocation, post translational modifications and folding, peptide cleavage



Figure 2. The criteria of the four expression systems. The cluster rules of assessment criteria are based on the industrial extent and host intrinsic property, such as upfront and production cost, scale and risk, strain producing capacities and safety. Due to the fast growth rate of *Escherichia coli*, it is widely used for production of small and simple pharmaceutical proteins. However, there are some limitations for using this expression system. For instance, the secretion of recombinant proteins is inefficient, which requires more complex downstream purification process. And also it is difficult for *E. coli* to perform complicated PTM. On the contrary, CHO cell has a high similarity to human cells, and this impart CHO cell lines a good ability to perform PTM of proteins for human use. Nevertheless, CHO cell shows other disadvantages, such as low growth rate and high contamination risk, which means that the manufacturers need to spend more time and money to maintain the production process, and pay more attention and cost on the clearance and disinfection. *Saccharomyces cerevisiae* seems to be a good compromise because of the fast growth rate, low cost of medium and downstream processing, low contamination risk, and good secretory capacity. *S. cerevisiae* can carry out N-linked glycosylation modification with a high mannose type, which is a different glycosylation manner than found in humans, and it therefore has limitation for producing complex glycoproteins.

PTM: Post-translational modification.

Data taken from [8,11,30].

and additional glycosylation, sorting, and secretion (Figure 5) [35]. Each step represents an engineered target for improving expression. Even activation of cell stress responses could be considered as a possible way toward a cost-effective production process, due to the increasing cell stress may couple with overexpression of protein [36].

Increasing the gene copy number is an easy way to increase transcription. Two kinds of plasmids are commonly used in *S. cerevisiae*. Centromeric plasmids,



Figure 3. Platforms for production of European Medicines Agency-approved rDNA biopharmaceuticals. (A) Cumulative numbers of platforms used for the production of European Medicines Agency-approved rDNA biopharmaceuticals; (B) European Medicines Agency-approved biopharmaceuticals produced by *Saccharomyces cerevisiae*, authorized year, company.

BPG: BioPartners GmbH; CPL: Canyon Pharmaceuticals Ltd; GSK: GlaxoSmithKline Biologicals SA; JCI: Janssen-Cilag International NV; MSD: Merck Sharp & Dohme Ltd; NVO: Novo Nordisk A/S; SAG: Sanofi-aventis groupe; SPM: Sanofi Pasteur MSD.

⁺This medicine is now withdrawn from use in the EU.



Figure 4. Sales of Insulin and Insulin analog in 2012. Data taken from [32–34].

containing a yeast centromere sequence and an autonomously replicating sequence, maintain about one copy number per cell [37]. 2µ-based plasmids, containing either the whole 2μ sequence or partial 2μ sequence including both the replication origin (ori) and the *REP3* stability locus, maintain 10–40 copies per cell [38],



Figure 5. Processing steps in protein secretion of eukaryotic cells. Protein secretion of eukaryotic cell is like products production in a real factory. The DNA sequences coding of interested proteins is the product blueprint; once the mRNA synthesis is ready, it could be used as template to guide polypeptide synthesis. The polypeptide will be either co- or post-translationally translocated into the ER where there follows a series of modifications (disulfide bond formation and ER-glycosylation). The correct folded polypeptides will be trafficked from the ER to the Golgi apparatus via vesicles coated with COPII and here there will be carried out further post-translational modifications and sorting. Finally, the mature proteins will be secreted out of the cell. ER: Endoplasmic reticulum.

Key Terms

SM protein: Sec1/Munc18-like (SM) protein, shaped like clapes, regulates SNARE proteins and SNARE-mediated membrane fusion.

SNARE: Composition by SNARE proteins for mediation of vesicle fusion.

which are beneficial for high expression of proteins. It should be noted that the whole 2μ sequence plasmid should be used in the cir⁰ strain without native plasmid for avoiding recombination between heterologous and native plasmids; the partial 2μ plasmid should be used in the cir⁺ strain with native plasmid, which provides transacting factors (*REP1* and *REP2*) for keeping plasmid stability [38]. Down-regulation of the selection marker gene combined with destabilization of the marker protein results in higher copy number of 2μ plasmids that can be used for further increasing protein expression [39]. However, plasmid-based expression may encounter problems such as segregational instability, structural instability and allele segregation [40], which are disadvantages in industrial application. Chromosome integration could avoid these problems, but single copy gene is often not enough for high expression and one will therefore typically integrate several copies of the expression gene into the chromosome [41]. By utilization of a multiple chromosome-integrated plasmid, 5–7 copies of human alpha-fetoprotein genes were integrated into the *S. cerevisiae* chromosome resulting in successful secretion of human alpha-fetoprotein to the culture medium [42]. In case of cloning of large DNA fragments, yeast artificial chromosomes may be useful [43].

Nascent peptides are involved in many processing steps before secretion as mature proteins. Table 1 lists some examples for improving biopharmaceuticals production by engineering of the protein processing and secretory pathway of *S. cerevisiae*. Protein folding in correct conformation in the ER is very important as it determines whether the protein is targeted for the secretory pathway or whether it is assigned for ER-associated degradation [44]. If the nascent peptide is synthesized too fast, there is not enough time for the peptide to fold correctly resulting in misfolding [45]. This will cause luminal burden resulting in ER stress, which lowers

Table 1. Examples of improving biopharmaceutical production by engineering of *Saccharomyces*

Protein	Category	Size (AA)	Disulfide bonds/ N-glycosylation	Production level	Host modification targets#	Production pathway affected	Ref.	
Hirudin	Anticoagulants	65	3/0	1.1 g/l 36 mg/l	PDI1+, ERO1+ BiP+	Disulfide bond formation Regulate unfolded protein response	[50] [49]	
Human serum albumin	Blood factors	585	17/0	6 g/l	<i>SIL1+, LHS1+, JEM1</i> + and <i>SCJ1</i> +	Regulate the ATPase cycle of Kar2p	[19]	
Human transferrins	Blood factors	679	19/2	2.25 g/l	PDI1+, YPS1-, HSP150-	Disulfide bond formation, Reduce protein degradation	[69,70]	
Insulin precursor	Hormone	51	3/0	85 mg/l 19 mg/l	SLY1+, SEC1+ Mutant HSF1+	ER to Golgi transport, exocytosis Activate heat shock response	[61] [53]	
Human parathyroid hormone	Hormone	84	0/0	350 mg/l	YPS1-, YPS2-, YPS3-, YPS6-, YPS7-	Reduce protein degradation	[66]	
Glucagon	Hormone	29	0/0		YPS1-	Reduce protein degradation	[63]	
sHBsAg	Antigen	226	-	74.4 mg/l	PDI1+	Disulfide bond formation	[51]	
PDGF	Human Growth Factor	241	5/1	1.5 mg/l	PDI1+	Disulfide bond formation	[52]	
-: Deletion or down regulation; +: Overexpression or up regulation; AA: Amino acid; sHBsAg: S domain of hepatitis B virus surface antigen.								

the efficiency of protein synthesis, and ER processing becomes a limited step. The ER luminal binding protein (BiP, Kar2p in S. cerevisiae) is a chaperone belonging to the member of the HSP70 family, and mediates protein folding in the ER and export of soluble proteins [46,47]. It also participates in the ER-associated degradation and regulates the unfolded protein response by interaction with Ire1p through controlling HAC1 activation [19,48]. The production of human serum albumin could be increased by overexpression of SIL1, LHS1, and JEM1, which regulate Kar2p in S. cerevisiae [19]. Antithrombotic hirudin secretion by S. cerevisiae could also be enhanced by overexpression of KAR2 [49]. Moreover, overexpression of PDI1 together with ERO1, which are responsible for disulfide bond formation, also increases antithrombotic hirudin production by maintaining proper redox balance in the ER [50]. It seems that overexpression of PDI1 is generally a good strategy for increasing protein production; for example it is also benefit for production of sHBsAg and PDGF in S. cerevisiae [51,52].

The heat shock response (HSR) is a well-ordered reaction to environmental and physiological stress. Hundreds of genes coding for molecular chaperones that help protein folding will be elevated through activation of a primary HSR transcription factor Heat shock factor (Hsf1p) [53]. By overexpressing a mutant HSF1 for constitutively activating HSR, insulin precursor production could be improved [53]. Similarly, glycosylation, one kind of post-translational modification, is as important as the folding process as it affects protein folding, quality control, protein stability and solubility [54,55]. Proper glycosylation is required by many pharmaceutical proteins for their full active biological function when applied to human therapy. Although CHO cells are widely used to produce glycoproteins with human-like N-glycosylation patterns [56], glycan structures are still not identical to those of native human protein and may sometimes cause immunogenic reactions [57]. S. cerevisiae can perform glycosylation of proteins, but mainly in a high-mannose manner, which is immunogenic [58]. The ALG3 and ALG11 double mutant S. cerevisiae is disable to perform its native high-mannose modification on the secretory proteins; coexpression of mammalian GnTI and GnTII in this strain allows production of complex-type human-like glycans [59,60]. This enabled production of the monoclonal antibody HyHEL-10 with humanized N-glycans [59].

Heterologous protein production by *S. cerevisiae* may not only be limited at the transcriptional, translational and PTM level. Thus, improving the trafficking between different organelles could also enhance protein secretion. Overexpression of Sec1p, which is a SM protein regulating the soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) com-

plex, improves insulin secretion by helping the vesicle trafficking from the Golgi to the cellular membrane [61]. Overexpression of Kex2p, a Golgi associated endopeptidase, may facilitate the cleavage of pro-peptide from pro-protein hence increase production of mature protein [51]. The yapsins are a family of aspartic proteases, which are located at the cell surface through glycosyl-phosphatidylinositol anchoring, exhibit a common specificity cleavage activity at basic amino acid residues of pro-proteins [62]. Some proteins are sensitive to proteases, which will degrade proteins subsequently and reduce the final production of proteins. The Yps1p is a major protease for the aberrant proteolytic processing, deletion of YPS1 results in improving of glucagon production in S. cerevisiae [63]. Similar results were observed for production of human parathyroid hormone (hPTH) [64]. However, even though a YPS1 disrupted strain was used for production of hPTH, significant proteolysis still appeared in the later stages of fed-batch fermentation [65]. Multiple-yapsins-deficient mutant strains were therefore constructed by deletion of YPS1 as well as YPS2, YPS3, YPS6 and YPS7, to further reduce potential cleavages of pro-proteins. This quintuple disruptant strain showed improved efficiency in preventing proteolysis of hPTH in fed-batch fermentation compared with the single deletion strain [66]. Elimination of other kinds of proteinases may also increase protein production. Deletion of the PRB1 gene, coding for vacuolar proteinase B [67], resulted in a five-fold increase of the human mechano-growth factor titer [68].

Modification of several targets, overexpression of PDI1 together with deletion of YPS1 and HSP150, has been applied on engineering of S. cerevisiae for improving production of human transferrins and their quality [69,70]. Pdilp helps protein folding, hence improving protein production in its correct conformation. YPS1 deficiency reduces the degradation of target protein. Hsp150p is a secretory protein and will be co-purified with recombinant proteins. Deletion of HSP150 could remove a potential contaminant and is therefore beneficial for downstream purification. Method for enhancing recombinant human albumin production by overexpression of SIL1, LHS1 and JEM1 was also useful for increasing human transferrin production [19]. Such multitargets strategy for improved proteins production is favorable, it could release the limitation at different levels and the combinatorial effects will be accessed. A significant enhancement of intracellular human hemoglobin production by yeast was achieved by an optimal combination of heme group supply and globin production through metabolic engineering of heme biosynthetic pathway and altering two globin subunits ratios at the same time [71].

Key Term

RNA-seq: A technology utilizes next-generation sequencing for full-length cDNA sequencing, which provides comprehensive, quantitative result of complete RNA portion of transcriptome.

Enhanced protein production by *S. cerevisiae* using new technologies

Systems biology tools

An ideal expression platform is to convert substrate to final product without the formation of any by-product; however, no such expression platforms exist in nature. Engineering of S. cerevisiae is a useful way to improve its properties for production of proteins. Key questions are which genes should be manipulated and to which extent changes should be made. Literature reading or preliminary experiments could provide some useful information. However, due to the complexity of the protein secretory machinery, a successful strategy used for production of a certain protein sometimes is invalid when applied for production of another protein [72]. Systems biology provides a global view of the physiology and may involve the use of different -omics for analysis (i.e., genomics, transcriptomics, proteomics, metabolomics and fluxomics) [73-76].

Genetic information stored in the genome is the most basic level for understanding higher-order biological systems. Even small changes in the genome may influence cellular function or its products. With well-developed DNA sequencing methods and bioinformatics, which are two major parts of genomics, it is possible to relatively easy access the coding, function and structure of the genome [77]. Metabolic engineering is a rational design-based method and widely used for improving desired phenotypic properties [78]. However, the application may be limited if not enough information is available. Unlike metabolic engineering, adaptive evolution can be used to obtain favorable phenotypic strains even if there is not enough information at first. Here one relies on the appearance of random mutations in the chromosome and selection of mutations contributing to a desired phenotype may accumulate by imposing the cells under a specific selective pressure. High throughput deep DNA sequencing can be applied on the mutants, for identification of beneficial mutations, and other omics (transcriptomics and metabolomics) may assist in identifying causal relationships between identified mutations and the identified phenotype [79]. This approach may lead to new insight into genotype-phenotype relationships and hereby provide novel targets for metabolic engineering. Similar to evolutionary engineering, mutants with desired phenotype can be identified by random mutagenesis generated by chemical or physical agents followed by selective screening. Also here genomics analysis combined with reverse metabolic engineering can be applied for improving performance of the strains.

The transcriptome is a dynamic state of gene expression linking genotype to phenotype, which is affected by or responding to the extra- and intra-cellular environment [80]. The expression level of different genes at the same condition or the same gene at different conditions could vary several orders of magnitude; and this can be used for quantifying gene activities at different environmental conditions. Using a network component analysis of transcriptome data from microarrays, Contador et al. identified transcription factors changed at different fermentation stages when expressing a heterologous protein in S. cerevisiae [81]. This assisted in gaining insight into the cellular response to high level expression of proteins. RNA-seq is a revolutionary quantitative approach for studying transcriptomics, providing accurate measurement of the transcriptome [82]. Compared with microarrays, RNA-seq is a cost saving high-throughput sequencing technology with precise resolution at the single base pair; it barely relies on genomic sequence information and it is able to distinguish between different isoforms and allelic expression and has much larger dynamic range for quantification of gene expression levels due to low background noise [82]. Many novel transcribed regions in yeast were identified by RNA-seq and hence a more comprehensively map of the transcriptome was obtained [83].

The entire protein properties, including protein quantity, post-translational modification and interaction, are described by proteomics [84-86]. Relationships between the proteome and the transcriptome could help to identify cellular regulation mechanisms following transcription [87]. A weak positive correlation of the quantity between protein levels and corresponding mRNA levels implied a variety in post-transcription processing of mRNA. There can be many explanations for this, but different codon composition of genes may lead to differences in codon usage hence contributes to variation in translational efficiency, and it has been found that genes with similar functions tend to have similar codon frequencies resulting in similar correlations between protein and mRNA levels [87]. Proteomics could not only be used as a validation of transcriptomics (by protein quantity), but also provide information about PTM. In many cases, PTMs are essential for proteins to carry out biological function and hence to regulate cellular responses and module metabolism [88]. A recent study highlighted the importance of PTMs in metabolic control and variant of life-history traits by using alcoholic fermentation in S. cerevisiae as a model [89]. Most cellular processes in the cell are carried out or regulated by multiprotein complexes (protein—protein interaction) rather than only by a single protein [90]. Building protein—protein interaction networks provides knowledge to understand cellular function; an affinity purification coupled with MS/MS method is useful to for analysis of protein—protein interactions under near physiological conditions [91].

The metabolome refers to the complete set of metabolites that are generated by a biological system. Metabolomics, systematic study of the metabolome, could be used for construction of biochemical networks in the organism [92]. Integrated with other omics data, metabolomics can provide a more detailed landscape of cellular processes and map the metabolic pathways to reveal cellular states under different conditions. Important metabolic nodes may be identified for releasing the limitation in protein production. Thus, a recombinant host Bacillus megaterium exhibited lower GFP yield under large-scale fermentation condition [93]. Quantitative analysis of metabolites (intracellular amino acids) revealed potential limitations in the available of tryptophan, aspartate, histidine, glutamine and lysine as precursors during large-scale fermentation; additional supplementation of these amino acids to the medium resulted in increased GFP production [93]. Metabolic fluxes are the rates of metabolites passing through metabolic pathways or reactions in a biological system. Fluxomics, genome-scale measurement of metabolic fluxes, is a dynamic reflection of how the cells are processing carbon and energy sources and combined with transcriptomics, proteomics and metabolomics it is a powerful technique for gaining insight into regulation of metabolism [94]. Thus, by integrating transcriptome, fluxome and metabolome data, detailed kinetic models could be established and this could be used for understanding and predicting the behavior of S. cerevisiae under different environments or stresses [95].

A recent study described a genome-scale model for the protein secretory machinery in *S. cerevisiae* [96]. This model, consisting of 163 core components in the secretory machinery and 137 reactions, was constructed by using a bottom-up approach. To reduce the complexity, the machinery was divided into 16 subsystems, which covered all the secretory machinery processes. Seven secretory-related features about all proteins from the complete yeast proteome (5882 proteins) were extracted and condensed into the so-called yeast Protein Specific Information Matrix. Proteins were also assorted into 186 secretory classes, based on the possible combinations of the seven secretory-related features. With the information of reactions and secretory classes, protein specific reaction in the secretory pathway could be generated. 1197 potential ER-Golgi secretory proteins were calculated and a genome-scale protein specific reaction list for 552 of these proteins was obtained (a total of 11,684 reactions). This model helps in understanding the protein secretory machinery not only in yeast but also in other eukaryotic organisms. More importantly, it could help to estimate the energy and metabolic demand in the secretory machinery and hereby evaluate metabolic engineering targets for improving protein secretion. Integration of multi-omics data into this model may lead to a better understanding of the secretory machinery.

Synthetic biology tools

Through systems biology it is possible to provide a global overview of biological systems and narrow the gap between genotype and phenotype, hence figure out the possible limitation in protein production. Besides traditional molecular methods and tools for manipulation of *S. cerevisiae*, a recently fast developing new discipline – synthetic biology – provides more efficient and useful methods and tools for engineering *S. cerevisiae*. These tools could target different levels and allow multilevel modifications of *S. cerevisiae* (Table 2).

While microbial cells are utilized to offer the basic biological processing functions as a producing platform (biological chassis), construction of expression plasmids or cassettes is needed in most cases for production of chemicals and proteins or for enhancement of their production [97-99]. Restriction enzyme and ligase-based cloning is the conventional method for this purpose; however, it is limited in some cases, when there are no suitable restriction sites or several genes need to be cloned at the same time. Easy and highthroughput construction methods not constrained by DNA sequences are therefore desired.

Gateway cloning is a method that is based on phage lambda site-specific recombination in vitro [100]; it provides a fast way to transfer hundreds of DNA sequences to the same plasmid [101] or transfer the same DNA sequence to different plasmids without any ligation steps [102]. To facilitate expression of protein and gene functional analysis in S. cerevisiae, a set of threesegment multisite gateway vectors was developed for rapid assembly of any combination of promoter, gene and tag in a one-step reaction [103]. Gibson assembly is another *in vitro* method that can be used for rapid assembly of multilinear DNA fragments in a single reaction [104]. Here the DNA fragment should be designed with overlapping sequences at the end of its neighbor fragments. The reaction relies on a mixture of three different enzymes, a mesophilic 5' to 3' exonuclease (T5 exonuclease) for double-stranded DNA (dsDNA) 5'-recession, a high-fidelity polymerase

Table 2. Some recently developed methods and tools for host modifications								
Modification level	Method/tool	Application	Ref.					
Pre-transcriptional	Gateway recombination; Gibson cloning; DNA assembler	Rapid pathway construction	[102,104,106,131]					
Pre-transcriptional	Yeast oligo-mediated genome engineering	Achieve high gene modification frequencies	[109]					
Pre-transcriptional; transcriptional	CRISPR-Cas systems	Genome engineering, transcriptional regulation	[113,115]					
Transcriptional	Synthetic promoter libraries	Tunable components for synthetic genetic networks	[119]					
Post-transcriptional; translational	Design synthetic ribosome binding sites	Control protein production	[121]					
Post-transcriptional	RNA-based control modules	Predictable tuning of expression levels	[122]					
Transcriptional; translational	Gene codon optimization	Reduce mRNA secondary structure, improve translation rates	[124,125]					
Post-translational	Genetically-encoded biosensors	Optimize and regulate metabolic pathways	[128]					
Post-translational	Synthetic protein scaffolds	Provide modular control over metabolic flux	[130]					

(Phusion DNA polymerase) for annealed complementary single-stranded DNA (ssDNA) trimming and extension, and a thermophilic ligase (Taq DNA ligase) for sealing the gap. Gibson assembly is very useful for large DNA assembly [104] and dsDNA construction from numerous chemical synthesis oligonucleotides [105]. The DNA assembler enables rapid construction of an entire biosynthetic pathway in one step by means of in vivo homologous recombination in S. cerevisiae [106]. Significantly, not only plasmids from DNA fragments can be built up via the DNA assembler but integration of DNA fragments into chromosome is also available, which maintains the heterologous pathways (genes) stable along with host replication which is beneficial for industrial application [106]. The DNA assembler facilitates the comprehensive evaluation of pathway variants involving multiple genes and hence the best one can be chosen [107]. To take full advantage of this method, improvements have been made for increasing the efficiency and accuracy of the DNA assembler [108].

Gene deletion or replacement in the chromosome is often undertaken by heterologous or endogenous recombination. High frequency recombination is usually obtained via dsDNA fragments with long homologous arms, ssDNA is of limited use due to its low frequency recombination unless there is a specific selection for this type of recombinants [109]. Dicarlo *et al.* identified key parameters in improvement of the recombination efficiency in *S. cerevisiae* by using oligonucleotides, which is easy to obtain commercially without the requirement for further PCR amplification and purification. This method, named Yeast Oligo-Mediated Genome Engineering, can be applied for achieving high gene modification frequencies on chromosomes with oligonucleotides directly [109]. These modifications can be accumulated and enriched through several rounds of manipulation and it is a very fast way towards rational strain engineering.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) systems provide resistance for Bacteria and Archaea against phages by RNA-guided nuclease activity [110]. It has been adapted as an efficient RNA-guided genome editing technology in eukaryotic systems [111,112]. A type II bacterial CRISPR-Cas system has been applied to engineer S. cerevisiae, and it shows increased recombination efficiency and feasibility for in site-specific mutation and allelic replacement [113]. The CRISPR-Cas system has a great potential to become a valuable genome engineering tool, because multiple targets could be modified simultaneously with corresponding guided RNA (gRNA) cassettes, which are easy to be constructed and co-transformed [113]. The mutant Cas9 deficient in its endonuclease activity can still form a DNA recognition complex together with the gRNA and target it to a specific DNA region (complementary to gRNA); if the DNA region is for transcriptional initiation or gene coding, the transcriptional initiation or elongation will be blocked by the complex [114]. This CRISPR interference (CRISPRi) is not limited in use for efficient gene silencing in *E*.coli [114], but has also been shown to work in eukaryotic organisms including *S. cerevisiae* for gene expression regulation [114,115]. A tunable orthogonal transcriptional regulation system could be established in *S. cerevisiae* via CRISPRi [115].

As one of the core transcriptional elements, promoters control gene expression and are essential for understanding regulatory mechanisms. Utilization of different strength promoters for control of gene expression at the transcriptional level is the most common method to alter cellular metabolism. However, native promoters may not offer a continual gradient transcriptional strength and hereby limit the ability to fine-tune gene expression. Synthetic promoter libraries, which have been created via random mutagenesis of a natural promoter [116], can be purposely designed and generated rapidly using DNA synthesis technology, and hereby used for unraveling transcription regulatory mechanisms [117,118]. As substitution for native promoters, these synthetic promoters could be applied as predictive transcriptional components in optimization of metabolic networks due to their large dynamic range but still the ability to have small changes in gene expression [119].

Gene expression will be affected by its 5' UTR (untranslated region) and 3' UTR [120], in which both natural element and artificial element could be utilized for control of expression. Instead of laborious work in screening from a large random library of ribosome binding sites, Salis *et al.* presented a predictive method for rational design of synthetic ribosome binding sites for controlling protein expression [121]. Synthetic RNA-based control modules located at the 3' UTR exhibited a range of gene regulatory activities at the post-transcriptional level, through variations of RNase III (Rn1p) processing efficiency that regulates mRNA stability [122]. Those modules can be used in combination with promoter libraries for fine-tuning and precise control of gene expression.

Codon utilization bias differs from organism to organism [123]. Amplification of a gene via PCR method will retain its codon usage pattern that may not be compatible in a heterologous expression platform. Gene codon optimization according to the expression platform could avoid unusual codon usage and alteration of the mRNA secondary structure and hence increase protein expression [124-126]. The signal sequence of a protein not only determines the final location of the protein but also affects protein secretion; for example, a synthetic leader was found to be more efficient for insulin precursor production by S. cerevisiae compared with the endogenous alpha factor leader [72]. Variants from the same signal sequence resulted in different secretory efficiency, and were utilized for improving protein secretion [127].

Genetically-encoded biosensors are genetic devices for sensing and responding to the change of environment and metabolic status of the cell [128], and these offer great prospects for being employed as regulatory elements for circumventing imbalanced metabolism and hence increase product yield [129]. The introduction of a heterologous pathway is potentially to cause metabolic imbalance in hosts due to the lack of regulatory mechanisms present in the original organism [130]. Synthetic protein scaffolds allow combination of relevant enzymes in an optimized stoichiometry and leads to a way for fine-tuning metabolic flux through modular control; it avoids the loss of intermediates during diffusion or to competing pathways, protects unstable intermediates and bypasses the unfavorable kinetics [130]. It is obvious but still interesting to evaluate the use of biosensors for improving recombinant protein production in the future, for example, for evaluation of ER stress associated with improper protein folding.

Future perspective

To date, the global biopharmaceutical market exceeds \$125 billion per year and is expected to increase continuously with a high annual growth rate. There is therefore much interest in continuously improve protein expression platform, both to enable production of novel proteins and to reduce the costs of producing recombinant proteins. Besides the most common used expression platforms, non-conventional microbial systems are also under development or evaluation to be alternative cell factories for biopharmaceutical production.

S. cerevisiae, which possesses the advantages of bacteria and eukaryotes, is an attractive choice for industrial biopharmaceutical production. Many biopharmaceuticals produced by S. cerevisiae have been commercialized including one of the most successful biopharmaceuticals - human insulin. Because of the upcoming patent cliff, improved expression systems and optimized processes are required for the originators to cope with competition from the potential biosimilars. Expression systems with a general efficient productivity for most proteins have significant advantages, optimization of biopharmaceutical production could begin from a high level with these outstanding platforms instead of from zero each time. Metabolic engineering of S. cerevisiae for enhancing protein production has been quite successful, but sometimes effective strategies for one protein are invalid for improving production of another protein. With the improvement of -omics measurements, systems biology offer the opportunity to provide a comprehensive understanding of biological functions of S. cerevisiae, and hereby potential limitations in protein production can be identified and thus become targets for metabolic engineering of S. cerevisiae. Fundamental research on the protein secretion system of *S. cerevisiae*, due to its role as a model organism, also provides useful information for unraveling mechanisms for protein production in this organism. Furthermore, novel methods and tools in synthetic biology allow rapid and efficient modifications of *S. cerevisiae*. Thus, we predict that by combining systems biology and synthetic biology, *S. cerevisiae* may be designed and developed to become a robust biological chassis for biopharmaceutical protein production.

Financial & competing interests disclosure

This work is funded by the Novo Nordisk Foundation and the European Research Council (contract no. 247013). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Introduction

- The development of biopharmaceuticals industry is driven by huge market demand and policy support.
- Escherichia coli, Saccharomyces cerevisiae and Mammalian cells are the most common used biopharmaceutical production platforms.
- Current state & recent advances in production of biopharmaceutical proteins by S. cerevisiae
- *S. cerevisiae*, a unicellular microbial organism with advantages of bacteria and eukaryotes, is an attractive choice for biopharmaceutical production.
- Dozens of pharmaceutical proteins, such as insulin, vaccines and blood factors, produced by *S. cerevisae* have entered the market; several of them are blockbusters.
- Metabolic engineering of *S. cerevisiae* can enhance its protein production capacity, improve product quality and ease downstream processing.
- Enhanced protein production by S. cerevisiae with new technologies
- Systems biology could provide global landscapes and deepen our insight of *S. cerevisiae* for improving our understanding of protein production mechanisms. Potential limitations in protein production can be figured out and become engineering targets.
- Novel methods and tools in synthetic biology allow rapid and efficient modification of S. cerevisiae, based on rational design.

Perspective

• It is possible to design and generate *S. cerevisiae* to be a robust biological chassis for biopharmaceutical protein production through the recruitment of systems and synthetic biology.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

- Rowland M, Noe CR, Smith DA *et al.* Impact of the pharmaceutical sciences on health care: a reflection over the past 50 years. *J. Pharm. Sci.*101(11), 4075–4099 (2012).
- 2 Nielsen J. Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered*4(4), 207–211 (2013).
- 3 Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol.*32(1), 54–60 (2014).
- 4 Martinez JL, Liu LF, Petranovic D, Nielsen J. Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.*23(6), 965–971 (2012).
- 5 Goodman M. MARKET WATCH Sales of biologics to show robust growth through to 2013. *Nat. Rev. Drug Discov.*8(11), 837–837 (2009).
- 6 Blockbuster Biologics 2012: sales are double that of 2006 driven by strong growth of recombinant antibodies, La Merie Business Intelligence website:

www.pipelinereview.com/index.php/2013050950921/Press-Room/Blockbuster-Biologics-2012-sales-are-double-that-of-2006-driven-by-strong-growth-of-recombinant-antibodies.html 7 Blockbuster Biologics 2012, La Merie Business Intelligence website:

www.pipelinereview.com/index.php/2013050850905/FREE-Reports/Blockbuster-Biologics-2012.html

- 8 Ma JK, Drake PM, Christou P. The production of recombinant pharmaceutical proteins in plants. *Nat. Rev. Genet.* 4(10), 794–805 (2003).
- 9 Ferrer-Miralles N, Domingo-Espin J, Corchero JL, Vazquez E, Villaverde A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* 8, 17 (2009).
- 10 Berlec A, Strukelj B. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J. Ind. Microbiol. Biotechnol.* 40(3-4), 257–274 (2013).
- 11 Vogl T, Hartner FS, Glieder A. New opportunities by synthetic biology for biopharmaceutical production in *Pichia* pastoris. Curr. Opin. Biotechnol. 24(6), 1094–1101 (2013).
- 12 Zhang Y, Meibohm B. Pharmacokinetics and pharmacodynamics of therapeutic peptides and proteins. In: *Pharmaceutical Biotechnology*. Wiley-VCH Verlag GmbH & Co. KGaA, Germany, 337–367 (2012).
- 13 Shibui T, Bando K, Misawa S. High-level secretory expression, purification, and characterization of an antihuman Her II monoclonal antibody, trastuzumab, in the methylotrophic yeast *Pichia pastoris*. Adv. Biosci. Biotech. 4(5), 640–646 (2013).

Biopharmaceutical protein production by Saccharomyces cerevisiae Review

- 14 Swartz JR. Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol.* 12(2), 195–201 (2001).
- 15 Choi JH, Keum KC, Lee SY. Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chem. Eng. Sci.* 61(3), 876–885 (2006).
- 16 Shusta EV, Raines RT, Pluckthun A, Wittrup KD. Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat. Biotechnol.* 16(8), 773–777 (1998).
- 17 Hong KK, Nielsen J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. Cell. Mol. Life Sci. 69(16), 2671–2690 (2012).
- 18 Jonikas MC, Collins SR, Denic V *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* 323(5922), 1693–1697 (2009).
- 19 Payne T, Finnis C, Evans LR *et al.* Modulation of chaperone gene expression in mutagenized *Saccharomyces cerevisiae* strains developed for recombinant human albumin production results in increased production of multiple heterologous proteins. *Appl. Environ. Microbiol.* 74(24), 7759–7766 (2008).
- 20 Ilmen M, den Haan R, Brevnova E *et al.* High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 4, 30 (2011).
- 21 Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, Porro D. Recombinant protein production in yeasts. In: *Recombinant Gene Expression* Humana Press, NY, USA, 329–358 (2004).
- 22 Benjaphokee S, Hasegawa D, Yokota D *et al.* Highly efficient bioethanol production by a *Saccharomyces cerevisiae* strain with multiple stress tolerance to high temperature, acid and ethanol. *New Biotechnol.* 29(3), 379–386 (2012).
- 23 Beltrao P, Albanese V, Kenner LR *et al.* Systematic Functional Prioritization of Protein Posttranslational Modifications. *Cell* 150(2), 413–425 (2012).
- 24 Gnoth S, Jenzsch M, Simutis R, Lubbert A. Control of cultivation processes for recombinant protein production: a review. *Bioproc. Biosyst. Eng.* 31(1), 21–39 (2008).
- 25 Fu ZB, Leighton J, Cheng AL, Appelbaum E, Aon JC. Optimization of a *Saccharomyces cerevisiae* fermentation process for production of a therapeutic recombinant protein using a multivariate bayesian approach. *Biotechnol. Prog.* 28(4), 1095–1105 (2012).
- 26 Treusch S, Hamamichi S, Goodman JL et al. Functional Links Between A beta Toxicity, Endocytic Trafficking, and Alzheimer's Disease Risk Factors in Yeast. Science 334(6060), 1241–1245 (2011).
- A case study of human diseases by utilizing *Saccharomyces cerevisiae* as a model organism.
- 27 Gitler AD, Chesi A, Geddie ML *et al.* Alpha-synuclein is part of a diverse and highly conserved interaction network that includes PARK9 and manganese toxicity. *Nat. Genet.* 41(3), 308–315 (2009).
- 28 Goffeau A, Barrell BG, Bussey H *et al.* Life with 6000 genes. *Science* 274(5287), 546, 563–547 (1996).

- 29 Solow SP, Sengbusch J, Laird MW. Heterologous protein production from the inducible MET25 promoter in *Saccharomyces cerevisiae*. *Biotechnol. Prog.* 21(2), 617–620 (2005).
- 30 Dingermann T, Zündorf I. Overview and classification of approved recombinant drugs. In: *Pharmaceutical Biotechnology*. Wiley-VCH Verlag GmbH & Co. KGaA, Germany, 153–178 (2012).
- 31 Sarantos K, Cleo K. Analysis of the landscape of biologicallyderived pharmaceuticals in Europe: dominant production systems, molecule types on the rise and approval trends. *Eur. J. Pharm. Sci.* 48(3), 428–441 (2013).
- 32 Novo Nordisk 2012 Annual Report. www.novonordisk.com/investors/annual-report-2012/ default.asp
- 33 Eli Lilly 2012 Annual Report. http://investor.lilly.com/annuals.cfm
- 34 Sanofi-aventis 2012 Annual Report. http://en.sanofi.com/investors/events/ corporate/2013/2013-02-07_Results_2012.aspx
- 35 Hou J, Tyo KEJ, Liu ZH, Petranovic D, Nielsen J. Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12(5), 491–510 (2012).
- 36 Tyo KEJ, Liu Z, Petranovic D, Nielsen J. Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. *BMC Biology* 10(1), 16 (2012).
- 7 Mumberg D, Muller R, Funk M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156(1), 119–122 (1995).
- 38 Da Silva NA, Srikrishnan S. Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12(2), 197–214 (2012).
- 39 Chen Y, Partow S, Scalcinati G, Siewers V, Nielsen J. Enhancing the copy number of episomal plasmids in *Saccharomyces cerevisiae* for improved protein production. *FEMS Yeast Res.* 12(5), 598–607 (2012).
- 40 Friehs K. Plasmid copy number and plasmid stability. In: New Trends and Developments in Biochemical Engineering. Springer, Heidelberg, Germany. 47–82 (2004).
- 41 Yamada R, Taniguchi N, Tanaka T, Ogino C, Fukuda H, Kondo A. Cocktail delta-integration: a novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. *Microb. Cell Fact.* 9, 32 (2010).
- 42 Dudich E, Dudich I, Semenkova L *et al.* Engineering of the *Saccharomyces cerevisiae* yeast strain with multiple chromosome-integrated genes of human alpha-fetoprotein and its high-yield secretory production, purification, structural and functional characterization. *Protein Expr. Purif.* 84(1), 94–107 (2012).
- 43 Burke DT, Carle GF, Olson MV. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236(4803), 806–812 (1987).
- 44 Schroder M. Engineering eukaryotic protein factories. Biotechnol. Lett. 30(2), 187–196 (2008).
- 45 Zhang G, Ignatova Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.* 21(1), 25–31 (2011).

- 46 Normington K, Kohno K, Kozutsumi Y, Gething MJ, Sambrook J. S. cerevisiae encodes an essential protein homologous in sequence and function to mammalian BiP. Cell 57(7), 1223–1236 (1989).
- 47 Flynn GC, Pohl J, Flocco MT, Rothman JE. Peptidebinding specificity of the molecular chaperone BiP. *Nature* 353(6346), 726–730 (1991).
- 48 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.* 13(3), 349–356 (2001).
- 49 Kim MD, Han KC, Kang HA, Rhee SK, Seo JH. Coexpression of BiP increased antithrombotic hirudin production in recombinant *Sacchdromyces cerevisiae*. J. *Biotechnol.* 101(1), 81–87 (2003).
- 50 Kim MD, Park EH, Cho JW *et al.* Enhanced production of antithrombotic hirudin by coexpression of Pdi1 and Ero1 in recombinant *Saccharomyces cerevisiae*. J. Biotechnol. 131(2), 147 (2007).
- 51 Kim EJ, Park YK, Lim HK, Park YC, Seo JH. Expression of hepatitis B surface antigen S domain in recombinant *Saccharomyces cerevisiae* using *GAL1* promoter. *J. Biotechnol.* 141(3-4), 155–159 (2009).
- 52 Robinson AS, Hines V, Wittrup KD. Protein disulfideisomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 12(4), 381–384 (1994).
- 53 Hou J, Osterlund T, Liu ZH, Petranovic D, Nielsen J. Heat shock response improves heterologous protein secretion in *Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol.* 97(8), 3559–3568 (2013).
- 54 Rich JR, Withers SG. Emerging methods for the production of homogeneous human glycoproteins. *Nat. Chem. Biol.* 5(4), 206–215 (2009).
- 55 Ciplys E, Sasnauskas K, Slibinskas R. Overexpression of human calnexin in yeast improves measles surface glycoprotein solubility. *FEMS Yeast Res.* 11(6), 514–523 (2011).
- 56 Gerngross TU. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat. Biotechnol.* 22(11), 1409–1414 (2004).
- 57 Casademunt E, Martinelle K, Jernberg M *et al.* The first recombinant human coagulation factor VIII of human origin: human cell line and manufacturing characteristics. *Eur. J. Haematol.* 89(2), 165–176 (2012).
- 58 Wildt S, Gerngross TU. The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* 3(2), 119–128 (2005).
- 59 Nasab FP, Aebi M, Bernhard G, Frey AD. A combined system for engineering glycosylation efficiency and glycan structure in *Saccharomyces cerevisiae. Appl. Environ. Microbiol.* 79(3), 997–1007 (2013).
- Describes a strategy to produce humanized N-glycosylated protein by *S. cerevisiae*.
- 60 Helenius J, Ng DTW, Marolda CL, Walter P, Valvano MA, Aebi M. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* 415(6870), 447–450 (2002).

- 61 Hou J, Tyo K, Liu Z, Petranovic D, Nielsen J. Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Metab. Eng.* 14(2), 120–127 (2012).
- 62 Krysan DJ, Ting EL, Abeijon C, Kroos L, Fuller RS. Yapsins are a family of aspartyl proteases required for cell wall integrity in *Saccharomyces cerevisiae*. *Eukaryotic Cell* 4(8), 1364–1374 (2005).
- 63 Egel-Mitani M, Andersen AS, Diers II et al. Yield improvement of heterologous peptides expressed in yps1disrupted Saccharomyces cerevisiae strains. Enzyme Microb. Technol. 26(9-10), 671–677 (2000).
- 64 Kang HA, Kim SJ, Choi ES, Rhee SK, Chung BH. Efficient production of intact human parathyroid hormone in a *Saccharomyces cerevisiae* mutant deficient in yeast aspartic protease 3 (YAP3). *Appl. Microbiol. Biotechnol.* 50(2), 187–192 (1998).
- 65 Song GY, Chung BH. Overproduction of human parathyroid hormone by fed-batch culture of a *Saccharomyces cerevisiae* mutant lacking yeast aspartic protease 3. *Process Biochem*. 35(5), 503–508 (1999).
- 66 Cho EY, Cheon SA, Kim H et al. Multiple-yapsindeficient mutant strains for high-level production of intact recombinant proteins in *Saccharomyces cerevisiae*. J. Biotechnol. 149(1-2), 1–7 (2010).
- 67 Moehle CM, Aynardi MW, Kolodny MR, Park FJ, Jones EW. Protease B of *Saccharomyces cerevisiae*: isolation and Regulation of the *PRB1* Structural Gene. *Genetics* 115(2), 255–263 (1987).
- 68 Morozkina EV, Marchenko AN, Keruchenko JS et al. Proteinase B disruption is required for high level production of human mechano-growth factor in *Saccharomyces cerevisiae*. J. Ind. Microbiol. Biotechnol. 18(3), 188–194 (2010).
- 69 Finnis CJ, Payne T, Hay J et al. High-level production of animal-free recombinant transferrin from *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 9, 87 (2010).
- 70 Finnis C. Optimizing Saccharomyces cerevisiae for the secretion of difficult proteins, Novozymes website posters and presentations. www.biopharma.novozymes.com/en/information-centre/ posters-and-presentations/Documents/PEGS_Boston_ CJAF_170510.pdf
- 71 Liu L, Martinez JL, Liu Z, Petranovic D, Nielsen J. Balanced globin protein expression and heme biosynthesis improve production of human hemoglobin in *Saccharomyces cerevisiae*. *Metab. Eng.* 21, 9–16 (2013).
- 72 Liu Z, Tyo KEJ, Martínez JL, Petranovic D, Nielsen J. Different expression systems for production of recombinant proteins in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 109(5), 1259–1268 (2012).
- 73 Snyder M, Gallagher JEG. Systems biology from a yeast omics perspective. FEBS Lett.583(24), 3895–3899 (2009).
- 74 Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11(7), 476–486 (2010).
- 75 Reaves ML, Rabinowitz JD. Metabolomics in systems microbiology. *Curr. Opin. Biotechnol.* 22(1), 17–25 (2011).

- 76 Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 80, 273–299 (2011).
- 77 Kim IK, Roldao A, Siewers V, Nielsen J. A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res.* 12(2), 228–248 (2012).
- 78 Nielsen J. Metabolic engineering. Appl. Microbiol. Biotechnol. 55(3), 263–283 (2001).
- A detailed review on fundamentals and applications of metabolic engineering.
- 79 Hong KK, Vongsangnak W, Vemuri GN, Nielsen J. Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proc. Natl Acad. Sci. USA* 108(29), 12179 (2011).
- 80 Velculescu VE, Zhang L, Zhou W *et al.* Characterization of the yeast transcriptome. *Cell* 88(2), 243–251 (1997).
- 81 Contador CA, Andrews BA, Liao JC, Asenjo JA. Identification of transcription factors perturbed by the synthesis of high levels of a foreign protein in yeast *Saccharomyces cerevisiae*. *Biotechnol. Prog.* 27(4), 925–936 (2011).
- 82 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10(1), 57–63 (2009).
- 83 Nagalakshmi U, Wang Z, Waern K *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320(5881), 1344–1349 (2008).
- 84 Rogers LD, Overall CM. Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol. Cell. Proteomics* 12(12), 3532–3542 (2013).
- 85 Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17(3), 121–127 (1999).
- 86 Gstaiger M, Aebersold R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat. Rev. Genet.* 10(9), 617–627 (2009).
- 87 Olivares-Hernandez R, Usaite R, Nielsen J. Integrative analysis using proteome and transcriptome data from yeast to unravel regulatory patterns at post-transcriptional level. *Biotechnol. Bioeng.* 107(5), 865–875 (2010).
- 88 Oliveira AP, Sauer U. The importance of post-translational modifications in regulating *Saccharomyces cerevisiae* metabolism. *FEMS Yeast Res.* 12(2), 104–117 (2012).
- 89 Albertin W, Marullo P, Bely M *et al.* Linking posttranslational modifications and variation of phenotypic traits. *Mol. Cell. Proteomics* 12(3), 720–735 (2013).
- 90 Gavin AC, Bosche M, Krause R *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415(6868), 141–147 (2002).
- 91 Wepf A, Glatter T, Schmidt A, Aebersold R, Gstaiger M. Quantitative interaction proteomics using mass spectrometry. *Nat. Methods* 6(3), 203–205 (2009).
- 92 Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7(2), 129–143 (2009).
- 93 Korneli C, Bolten CJ, Godard T, Franco-Lara E, Wittmann C. Debottlenecking recombinant protein production in

Bacillus megaterium under large-scale conditions-targeted precursor feeding designed from metabolomics. *Biotechnol. Bioeng.* 109(6), 1538–1550 (2012).

- 94 Cascante M, Marin S. Metabolomics and fluxomics approaches. *Essays Biochem.* 45, 67–81 (2008).
- 95 Vital-Lopez FG, Wallqvist A, Reifman J. Bridging the gap between gene expression and metabolic phenotype via kinetic models. *BMC Syst. Biol.* 7, 63 (2013).
- 96 Feizi A, Osterlund T, Petranovic D, Bordel S, Nielsen J. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS ONE* 8(5), (2013).
- •• A well-performed study of the eukaryal protein secretion pathway by genome-scale modeling approach.
- 97 Xie W, Liu M, Lv X, Lu W, Gu J, Yu H. Construction of a controllable beta-carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng*, 111(1), 125–133 (2014).
- 98 Paddon CJ, Westfall PJ, Pitera DJ *et al.* High-level semisynthetic production of the potent antimalarial artemisinin. *Nature* 496(7446), 528–532 (2013).
- 99 Jiang M, Pfeifer BA. Metabolic and pathway engineering to influence native and altered erythromycin production through *E. coli. Metab. Eng.* 19, 42–49 (2013).
- 100 Walhout AJM, Temple GF, Brasch MA et al. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* 328, 575–592 (2000).
- 101 Rual JF, Hill DE, Vidal M. ORFeome projects: gateway between genomics and omics. *Curr. Opin. Chem. Biol.* 8(1), 20–25 (2004).
- 102 Freuler F, Stettler T, Meyerhofer M, Leder L, Mayr LM. Development of a novel Gateway-based vector system for efficient, multiparallel protein expression in *Escherichia coli*. *Protein Expr. Purif.* 59(2), 232–241 (2008).
- 103 Nagels Durand A, Moses T, De Clercq R, Goossens A, Pauwels L. A multisite gateway vector set for the functional analysis of genes in the model *Saccharomyces cerevisiae*. *BMC Mol. Biol.* 13, 30 (2012).
- 104 Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6(5), 343–345 (2009).
- 105 Gibson DG, Smith HO, Hutchison CA 3rd, Venter JC, Merryman C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* 7(11), 901–903 (2010).
- 106 Shao ZY, Zhao H, Zhao HM. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37(2), e16 (2009).
- 107 Zhou YJJ, Gao W, Rong QX *et al.* Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *J. Am. Chem. Soc.* 134(6), 3234–3241 (2012).
- 108 Kim B, Du J, Eriksen DT, Zhao HM. Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. *Appl. Environ. Microbiol.* 79(3), 931–941 (2013).

- 109 Dicarlo JE, Conley AJ, Penttila M, Jantti J, Wang HH, Church GM. Yeast oligo-mediated genome engineering (YOGE). ACS Synth. Biol. 2(12), 741–749 (2013).
- 110 Barrangou R, Fremaux C, Deveau H *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819), 1709–1712 (2007).
- 111 Mali P, Yang LH, Esvelt KM *et al.* RNA-guided human genome engineering via Cas9. *Science* 339(6121), 823–826 (2013).
- 112 Cong L, Ran FA, Cox D *et al*. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121), 819–823 (2013).
- A great example of application of the novel genome engineering technology CRISPR/Cas system.
- 113 DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41(7), 4336–4343 (2013).
- 114 Qi LS, Larson MH, Gilbert LA *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152(5), 1173–1183 (2013).
- 115 Farzadfard F, Perli SD, Lu TK. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. ACS Synth. Biol. 2(10), 604–613 (2013).
- 116 Alper H, Fischer C, Nevoigt E, Stephanopoulos G. Tuning genetic control through promoter engineering. *Proc. Natl Acad. Sci. USA* 102(36), 12678–12683 (2005).
- 117 Rajkumar AS, Denervaud N, Maerkl SJ. Mapping the fine structure of a eukaryotic promoter input-output function. *Nat. Genet.* 45(10), 1207–1215 (2013).
- 118 Sharon E, Kalma Y, Sharp A *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* 30(6), 521–530 (2012).
- An excellent study for understanding the regulatory mechanism in cell by systems biology and synthetic biology.
- 119 Rajkumar AS, Maerkl SJ. Rapid synthesis of defined eukaryotic promoter libraries. ACS Synth. Biol. 1(10), 483–490 (2012).
- 120 Wilkie GS, Dickson KS, Gray NK. Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends Biochem. Sci.* 28(4), 182–188 (2003).

- 121 Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27(10), 946–950 (2009).
- 122 Babiskin AH, Smolke CD. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol. Syst. Biol.* 7, 471 (2011).
- 123 Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2(1), 13–34 (1985).
- 124 Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends Biotechnol.* 22(7), 346–353 (2004).
- 125 Kim HJ, Kwag HL, Kim HJ. Codon optimization of the human papillomavirus type 58 L1 gene enhances the expression of soluble L1 protein in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 35(3), 413–421 (2013).
- 126 Bentele K, Saffert P, Rauscher R, Ignatova Z, Bluthgen N. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675 (2013).
- 127 Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E, Wittrup KD. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae. Biotechnol. Bioeng.* 103(6), 1192–1201 (2009).
- 128 Michener JK, Thodey K, Liang JC, Smolke CD. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab. Eng.* 14(3), 212–222 (2012).
- 129 Zhang F, Carothers JM, Keasling JD. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* 30(4), 354–359 (2012).
- 130 Dueber JE, Wu GC, Malmirchegini GR *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27(8), 753–759 (2009).
- 131 Kuijpers NG, Solis-Escalante D, Bosman L et al. A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.* 12, 47 (2013).