

# Modeling for Understanding and Engineering Metabolism

Jens Nielsen<sup>1,2,3,4\*</sup> and Dina Petranovic<sup>2,3</sup>

<sup>1</sup> BiolInnovation Institute, Ole Maaløes Vej 3, DK2200 Copenhagen N, Denmark; <sup>2</sup> Department of Life Sciences, Chalmers University of Technology, SE412 96 Gothenburg, Sweden; <sup>3</sup> Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kgs. Lyngby, Denmark; <sup>4</sup> Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, DK2200 Copenhagen N, Denmark

\*For correspondence: [jni@bii.dk](mailto:jni@bii.dk)

## Abstract

Metabolism is at the core of all functions of living cells as it provides Gibbs free energy and building blocks for synthesis of macromolecules, which are necessary for structures, growth and proliferation. Metabolism is a complex network composed of thousands of reactions catalyzed by enzymes involving many co-factors and metabolites. Traditionally it has been difficult to study metabolism as a whole network and most traditional efforts were therefore focused on specific metabolic pathways, enzymes and metabolites. By using engineering principles of mathematical modeling to analyze and study metabolism, as well as engineer it, i.e. design and build, new metabolic features, it is possible to gain many new fundamental insights as well as applications in biotechnology. Here we present the history and basic principles of engineering metabolism, as well as the newest developments in the field. We are using examples of applications in: 1) production of protein pharmaceuticals and chemicals; 2) basic studies of metabolism; and 3) impacting health care. We will end by discussing how engineering metabolism can benefit from advances in artificial intelligence (AI)- based models.

## Introduction

Metabolism is at the core of all cellular functions providing not only the energy equivalents for driving all chemical processes but also the building blocks for all cellular structures, growth and proliferation. The metabolism is a network of many interconnected chemical reactions that in a coordinated fashion ensure degradation and modification of nutrients, i.e. from carbon and nitrogen sources, into different chemicals that the cells can use for synthesis of building blocks like amino acids, fatty acids and nucleotides. These building blocks are subsequently assembled into macromolecules like proteins, lipids, DNA and RNA that make up the cell. Major breakthroughs in our understanding of metabolism were made in the beginning of the 20<sup>th</sup> century when Otto Meyerhof, Gustav Embden and Jakub Karol Parnas identified the individual chemical reactions involved in the conversion of glucose to pyruvate. This pathway is today known as the Embden-Meyerhof-Parnas (EMP) pathway, but also often referred to as glycolysis. In 1922 Otto Meyerhof received the Nobel Prize in Physiology or Medicine for his work on mapping the conversion of glucose to lactic acid in muscle cells deprived of oxygen, so-called fermentative metabolism. In 1933 Otto Warburg received the Nobel Prize in Physiology or Medicine for his discovery of respiratory metabolism, but his name is today more known for the “Warburg effect” and his discovery that cancer cells tend to bypass respiration, in presence of oxygen, as they convert glucose to lactate. This observation was also made by Herbert Crabtree in studies of cancer cells, but his name is today associated with the same phenomenon in yeast that produce ethanol at high glucose concentrations even in the presence of oxygen, the so-called “Crabtree effect” In 1953 Hans Krebs received the Nobel Prize for identifying the steps involved in the tricarboxylic acid (TCA) cycle, also often referred as Krebs cycle, where acetyl-CoA is degraded in a metabolic cycle to carbon dioxide resulting in production of “energy equivalents”. Fritz Lipmann shared the

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

10.1017/qrd.2025.1

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Nobel Prize in 1953 with Hans Krebs for identifying acetyl-CoA as a crucial molecule in linking the EMP pathway with the TCA cycle, via the enzyme Pyruvate Dehydrogenase. In 1978 Peter Mitchell received the Nobel Prize in chemistry for proposing the chemiosmotic theory, that describes how electrons are transferred from the co-factor NADH, generated in the TCA cycle, to oxygen reducing it to water. According to this theory protons are being pumped against a concentration gradient out of the mitochondrial matrix when electrons transfer through the so-called electron transport chain. Paul D. Boyer and John E. Walker, who received the Nobel Prize in Chemistry in 1997 for the identification of the mechanisms of ATP Synthase, that forms ATP when protons re-enter the mitochondria, together with Jens C. Skou, who discovered the Na<sup>+</sup>-K<sup>+</sup>-ATP pump, and hereby all the mechanisms involved in both oxidative and fermentative metabolism of glucose had been mapped. Other Nobel Prizes in Chemistry were given for discovery of specific metabolic pathways, i.e. to Lord Todd in 1957 for his discovery of the biosynthesis of nucleotides, to Melvin Calvin in 1960 for his discovery of the mechanisms behind carbon dioxide fixation, and to Luis Leloir in 1970 for his discovery of the role of sugar nucleotides in sugar metabolism, e.g. for metabolism of galactose, now known as the Leloir pathway. Many other Nobel Prizes have been given for discoveries of vitamins and other bioactive chemicals such as molecules, enzymes and pathways of the so-called secondary metabolism e.g. terpenes, alkaloids, carotenoids, antibiotics etc.

The more than 100 years of research on metabolism has resulted in extensive mapping of most core reactions engaged in energy generation and cell synthesis, but we don't always know the identity and characteristics of the enzymes carrying out the reactions. There are also many dark spots in certain parts of metabolism, especially in plants as their cells have evolved the capability to perform advanced chemistry, but modern molecular biology techniques have enabled transfer and assembly of whole plant pathways into a host organism, e.g. Baker's yeast, and this has enabled mapping of pathways leading to complex natural products. This type of heterologous expression assembles all the enzymes of many different biosynthetic pathways, even from different organisms, leading to creation of various complex natural products in so-called cell factories (Gelanie *et al.*, 2015; Zhao *et al.*, 2023). Information about metabolic reactions and their associated enzymes is available in several different databases, e.g. Kyoto Encyclopedia of Genes and Genomes (KEGG), which is a valuable resource that significantly advanced our ability to understand and thus engineer metabolism.

Several studies, including the seminal work related to developing Metabolic Control Analysis (Kacser and Burns, 1973; Kell *et al.*, 1989), have shown a big difference between enzymes operating in isolation (test tube, *in vitro*) and how enzymes operate within a pathway (in cell, *in vivo*). Gaining insight into how sets of enzymes interact in a cell can therefore only be obtained through the introduction of mathematical models and in engineering disciplines, mathematical models play an important role in design and development of new or improved products and processes and for analyzing data. Biological systems are inherently hard to engineer and describe mathematically using different models, especially as we do not know most of the components of the system, their roles and functions, as well as inherent redundancies, inefficiencies and complexity. Even for the simplest cell we still have a challenge to know all the "parts", their individual functions and how they interact, as well as the unique features of that system. However, as we progress towards a complete overview of metabolism it is possible to build comprehensive models with remarkably wide application for studying and designing cellular metabolism, with the objective of developing new cell factories, for biomarker discovery, for drug discovery, and for design of healthy diets for humans.

## History of Engineering Metabolism

In 1973 Herbert W. Boyer and Stanley N. Cohen invented genetic engineering, that laid the foundation for a multibillion dollars industry, underling everything we do today in modern molecular biology, both in academia and industry. Through expression of a heterologous gene in a host cell like the bacterium *Escherichia coli* or Baker's yeast *Saccharomyces cerevisiae* their invention enabled scalable production of human proteins like insulin and growth hormone (Nielsen, 2013). Following success with expressing a single gene encoding a protein of commercial interest, genetic engineering was also exploited for expressing heterologous enzymes with the objective of reconstructing heterologous biosynthetic pathways in various hosts. In 1991 this led to the coining of the term Metabolic Engineering by James E. Bailey and Gregory N. Stephanopoulos (Bailey, 1991; Stephanopoulos and Vallino, 1991), and over the last 30 years Metabolic Engineering has been established as an active research field with a dedicated textbook (Stephanopoulos *et al.*, 1998), dedicated journals and several conference series.

Despite many successes in using Metabolic Engineering to develop new cell factories (see below), it has often been difficult to meet the techno-economic requirements for establishing commercial processes (Konzock and Nielsen, 2024). A major reason for this has been that it is often difficult to engineer the host metabolism in a way that ensures conversion of most of the carbon atoms from substrate towards production of the chemical/

compound of interest. The main reason for this is the trade-off between growth and product formation where most microorganisms have evolved to maximize growth. Re-directing flux towards the product of interest is therefore difficult for two main reasons: 1) metabolism is not organized into linear pathways but is a web (hairball) of interactions between metabolites and enzymes (Fig. 1A), and the conversion of substrate to the product therefore often engages a very large number of reactions not directly involved in this conversion; and 2) there is extensive regulation of flux through the different enzymes (Fig. 1B), at several levels, from genome, transcriptome, proteome and fluxome, and this often results in diversion of flux from the path from substrate to the product. An approach of engineering specific enzymes at a time is therefore often failing, and even though automation has in recent years enabled rapid evaluation of many different engineering targets, it has become clear that a more holistic design approach needs to be applied, using mathematical models.

In 1979 Aiba and Matsuoka presented a simple model of citric acid production by the yeast *Candida lipolytica* (today renamed as *Yarrowia lipolytica*) and they were the first to use simple mass balancing around intracellular metabolites to calculate fluxes through metabolic pathways (Aiba and Matsuoka, 1979) (Fig. 1C). In the 1980s and 90s, more mass balance models of the central metabolism of various bacteria and fungi were made. Pioneers of developing bacterial models were Bernhard Palsson (Varma and Palsson, 1994) and Gregory N. Stephanopoulos (Vallino and Stephanopoulos, 1993), whereas Jens Nielsen pioneered the development of models for eukaryal organisms, i.e. *S. cerevisiae* (Nissen *et al.*, 1997) and *Penicillium chrysogenum* (Jørgensen *et al.*, 1995). With the availability of genome sequences, it became possible to identify most of the enzymatic capabilities of a cell and hereby develop comprehensive mathematical models for metabolism. This led to development of so-called genome scale metabolic models (GEMs) for different bacteria by the Palsson group (Edwards and Palsson, 1999; Edwards and Palsson, 2000; Schilling *et al.*, 2002) and the first eukaryal cell (the yeast *S. cerevisiae*) by the Nielsen group (Förster *et al.*, 2003). *S. cerevisiae* is a widely used model organism in molecular and cell biology, as well as in industry where it is the most widely used cell factory for production of food, beverages, chemicals, fuels and pharmaceuticals (Nielsen, 2019). Its wide use as a model organism is well illustrated by the fact that several Nobel Prizes in Physiology or Medicine have been given to researchers that used yeast in their fundamental studies (Hohmann, 2016), e.g. to Leland Hartwell, Paul Nurse and Tim Hunt in 2001 for their discoveries of key regulators of the cell cycle, to James Rothman, Randy Schekman and Thomas Südhof in 2013 for their discovery of the protein secretory pathway in eukaryal cells, and to Yoshinori Ohsumi in 2016 for elucidating the mechanisms for autophagy.

Following these initial models the Nielsen group reconstructed GEMs for many other important microorganisms, such as *Streptomyces coelicolor* used for antibiotics production (Borodina *et al.*, 2005), *Aspergillus niger* used for production of citric acid and many industrial enzymes (Andersen *et al.*, 2008), *A. oryzae* used for production of many fermented food products and industrial enzymes (Vongsangnak *et al.*, 2008), and *P. chrysogenum* used for production of penicillin and other antibiotics (Aagren *et al.*, 2013). Since this early reconstruction of metabolic networks there have been developed GEMs for many other organisms (Gu *et al.*, 2019). Many models have been updated regularly as new biochemical information become available. For example, our yeast GEMs have had many different updates of the original model, with the most comprehensive model Yeast8 comprising almost 4,000 metabolic reactions linked to more than 1,100 enzymes and genes (Lu *et al.*, 2019). Yeast8 also represented a breakthrough in terms of engaging the scientific community as the model was made available open source via GitHub. Hereby researchers from around the world could contribute with annotation, curation and propose model updates and this has resulted in a recent update of the model (Yeast9), that contains information of 30 additional genes, 203 new reactions and 140 new metabolites (Zhang *et al.*, 2024). These comprehensive models do not only represent an extensive database of cellular metabolism, but also find application in metabolic engineering, basic science, and in biomedical research as we will discuss below. Despite the scale of these models they still only cover the core of metabolism as demonstrated in our discussion on the use of artificial intelligence for gaining new insight into metabolism.

## Engineering metabolism for industrial production

### Building a new (biotech) industry

The invention of genetic engineering resulted in the establishment of an industry for production of recombinant proteins used as therapeutics exceeding USD100B (Nielsen, 2013). Human insulin was launched as the first product by Ely Lilly in 1982, with the second being human growth hormone launched by Genentech in 1985. Today, there are more than 300 different recombinant proteins approved as therapeutics with many more in clinical trials, and among the top 10 selling pharmaceuticals 9 are biopharmaceuticals, which is a significant change from 20 years ago where most top-selling pharmaceuticals were small molecules.

Many different cell factories are being used for production of recombinant proteins. The bacterium *E. coli* is used for small proteins such as interleukins and insulin, but even though this organism enables high expression levels a disadvantage is that the recombinant protein accumulates intracellularly in so-called inclusion bodies, which makes the purification process more cumbersome and costly. *S. cerevisiae* is characterized by having a fully functional protein secretion pathway, and it can therefore secrete recombinant proteins to the extracellular medium, which facilitates purification. This yeast is therefore widely used for recombinant protein production, e.g. for insulin, human serum albumin, and glucagon-like peptide 1 receptor agonists (GLP1) (Huang *et al.*, 2014). The latter group of molecules mimic the action of the endogenous incretin hormone GLP1 that is released after eating, stimulating satiety. These molecules have therefore formed the basis for the active ingredient in the recently launched type-2 diabetes and obesity drugs Ozempic and Wegovy by Novo Nordisk.

The methylotrophic yeast *Komagataella phaffii* (formerly known as *Pichia pastoris*), is also a relevant cell factory as it ensures very high productivity due to a very efficient methanol induced expression system and the ability for the yeast to grow to very high cell densities. We developed a GEM for *K. phaffii* (Caspeta *et al.*, 2012) and expanded this model to describe how metabolism could be engineered to enable humanized glycosylation of recombinant proteins (Irani *et al.*, 2015). This enabled identification of metabolic engineering targets for improved production of a range of different recombinant proteins having industrial relevance. For production of more complex proteins such as antibodies, erythropoietin, and blood coagulation factors, e.g. Factor VII and Factor VIII, that require proper glycosylation it is necessary to apply a more advanced cell factory that can ensure that the protein is produced with the proper human glycosylation pattern. The Chinese Hamster Ovary (CHO) cells are often used but the originally productivity of the CHO cells was relatively low and there were issues with strain stability. However, there has been significant advancement in developing detailed mathematical models for CHO cells and our focus was to model the secretory pathway as this is crucial for ensuring efficient protein secretion by this cell factory (Gutierrez *et al.*, 2020). At the same time, through advancement in abilities to engineer CHO cells it has been possible to improve productivity significantly, and CHO cells has therefore become the cell factory of choice for proteins that require proper human glycosylation.

As the secretory pathway is of such critical importance for many processes in engineering, we also focused on developing detailed mathematical models for the protein secretory pathway in *S. cerevisiae*. In one study the protein secretory pathway was completely mapped, with 162 proteins in this pathway engaged in processing 1,190 proteins (Feizi *et al.*, 2013). Very few of the endogenous proteins being processed by the pathway are eventually secreted, whereas the remainder are processed through this pathway to be directed to the cell membrane where they function as transporters or receptors. Production of a heterologous protein will therefore compete with capacity of this pathway, and if a heterologous protein is expressed at a very high level, it will therefore drain capacity for processing membrane proteins, which will impact overall cellular functions. To quantify the demand for resources in the complex protein secretory pathway a detailed enzyme constrained mathematical model was established (Li *et al.*, 2022a), and using this model it was possible to quantify the impact of producing different recombinant proteins. It was further possible to use the model for design how the protein secretory pathway could be improved for production of a specific protein, and many of these designs could be experimentally validated (Li *et al.*, 2022). Furthermore, the design strategies proposed by the model matched many earlier targets that had been identified through sequencing of different mutant strains with natural varying secretion capacities (Huang *et al.*, 2015; Huang *et al.*, 2017; Huang *et al.*, 2018).

### **Transforming the old (chemical) industry**

With the current wide use of microbial fermentation for production of recombinant proteins used as advanced medicines, it is interesting to note that the use of microbial fermentation for production of chemicals dates to the mid-19<sup>th</sup> century, where ethanol produced through yeast fermentation was used as a lighting fuel. In 1908 the demand for ethanol increased as it was used to fuel Henry Ford's Model T. With the oil boom gasoline took the lead, but in the 1920s and 1930s ethanol was used as an octane booster, and it was in high demand during World War II (WWII) due to fuel shortages. In the 1970s with the rise in oil prices, ethanol was gaining renewed interest as a blend-in fuel, and this resulted in the establishment of the current very large ethanol industry; e.g. in 2023 more than 110 billion of liters of ethanol was produced through yeast fermentation. In 1908 there was another landmark in microbial production of chemicals, namely the development of the acetone-butanol production by Chaim Weismann, a lecturer at Manchester University and later the first president of Israel, who developed a process based on fermentation with the bacterium *Clostridium acetobutylicum*. During World War I there was a large demand for acetone to be used in gunpowder. Acetone had earlier been produced from calcium acetate imported from Germany, and the Weismann process therefore became an important new route to obtain acetone for the United Kingdom. Production of ethanol and the Weismann process are both anaerobic

processes, i.e. there is no need for provision of oxygen to the cells. A key landmark was therefore the production of citric acid through fermentation with the filamentous fungus *Aspergillus niger*, that was introduced in 1919. This fungus is extremely tolerant to low pH, and the fermentation process can therefore be operated at pH 2-3, which reduces the demand for maintenance of aseptic conditions as very few other microorganisms can survive at these conditions. Following the discovery of penicillin by Alexander Flemming in 1928, there was extensive attempts to chemically synthesize this new bioactive to be used as an antibiotic, but during WWII it was decided to start fermentation-based production using the filamentous fungus *Penicillium chrysogenum*. This resulted in development of the first aerobic fermentation process that required supply of large amounts of aseptic air to the fermentation process, and establishment of this process therefore laid the basis for production of many different chemicals using microbial aerobic fermentation. Retrospectively it turned out to be a wise decision to choose the fermentation route for production of penicillin as it was first demonstrated possible to chemically synthesize penicillin in 1956, and the chemical synthesis route cannot compete with the fermentation route. Today most antibiotics in the world are being produced by microbial fermentation.

Today microbial fermentation represents an industry exceeding USD100 billion, and many different chemicals are being produced through this route. Many new processes have been developed and the application of Metabolic Engineering basically falls into four different applications: 1) Engineering central metabolism to improve titers, rates and yields (TRYs); 2) Extension of substrate range of the cell factory; 3) Improving tolerance of the cell factory; 4) Expression of heterologous pathways for production of valuable chemicals.

**Improvement of titers, rates and yields** is essential for the establishment of an industrially viable process (Konzock and Nielsen, 2024). Even though it may be possible to extend endogenous pathways with a few enzymatic steps leading to the product of interest, the yield of product from the feedstock, typically glucose, and the rate of production, may often be low and it is therefore necessary to engineer the central carbon metabolism of the cell factory to ensure that carbon is directed from glucose to the product at a high rate. This is well illustrated in work on engineering *Escherichia coli* for production of 1,3 propanediol (Nakamura and Whited 2003) and 1,4 butanediol (Yim *et al.*, 2011). In both cases the bacterium was heavily engineered with more than 10 different genetic modifications that ensured a high rate and yield. Today these two chemicals are being produced commercially and used in the production of various plastics. Lactic acid has been commercially produced through fermentation with lactic acid bacteria since the early 20<sup>th</sup> century for use in the food industry, but with development of technology for polymerization of lactic acid to polylactate (PLA), that is a polymer with valuable properties such as biodegradability, good layer adhesion, and high strength, there was a need for large volumes of pure lactic acid, and not the lactate salt. Lactic acid bacteria require supplementation of complex feedstocks, and this makes it expensive and difficult to obtain pure lactic acid from these fermentation. The company Cargill therefore developed a new process based on an engineered yeast that can tolerate low pH for production of lactic acid, and this process resulted in a significant expansion of lactic acid production. More recently the company Corbion developed a process for purification of lactic acid from a traditional lactic acid bacterial fermentation process, and hereby enabled supply to production of PLA. Engineering of the central metabolism has also been shown to enable improvement of ethanol production and reduction of production of the by-product glycerol. Production of ethanol from glucose is completely balanced in terms of redox potential as there is also formed one molecule of carbon dioxide per molecule of ethanol formed, but as part of the glucose is used for production of more yeast cells that are more reduced than glucose, there is a need for an electron sink, and converting part of the glucose to glycerol represents such an electron sink (glycerol is more reduced than glucose). Cells are using different co-factors to balance electron flows within metabolism, and based on metabolic modeling it was identified that by engineering pathways for ammonia assimilation it should be possible to reduce the requirement for glycerol production and hereby increase ethanol production by 5-8%, which was experimentally validated (Nissen *et al.*, 2000). Also, mathematical modeling of the central carbon metabolism guided how changing the co-factor usage in glycolysis could give a similar effect (Bro *et al.*, 2006). GEMs have also played an important role in identification of targets for engineering yeast for over-production and secretion of free fatty acids by engineering both the central carbon metabolism and the fatty acid metabolism (Zhou *et al.*, 2016). Through pursuing additional model guided targets it was further possible to engineer the central carbon metabolism such that yeast could be transformed from alcoholic fermentation leading to ethanol production to efficient conversion of glucose to free fatty acids (Yu *et al.*, 2018). As the conversion of glucose to free fatty acids is redox im-balanced the yield was, however, found to be relatively low, but through reconstruction of a completely synthetic glycolysis identified through modeling it was possible to overcome this challenge and hereby an even better producing strain was obtained (Qin *et al.*, 2023).

Many cell factories have a **limited range of carbon sources** they can use efficiently. Thus, yeast, the most widely used cell factory, is not very efficient in using galactose and it cannot use xylose and arabinose as carbon and energy sources. This is important in the context of using lignocellulosic materials as feedstock for production of fuels and chemicals, as galactose, xylose and arabinose are abundant sugars in these feedstocks. Galactose is metabolized via the Leloir pathway, and even though this pathway only involves a few additional steps compared with the metabolism of glucose, the pathway is quite inefficient, and the growth rate of yeast on galactose is less than 50% compared with on glucose (Ostergaard *et al.*, 2000). The Leloir pathway is tightly regulated, but through engineering the regulatory machinery rather than simply over-expressing the enzymes it was possible to significantly increase the growth rate of yeast on galactose (Ostergaard *et al.*, 2000). Using mathematical modeling we later found that this was due to a requirement for balanced expression of the individual enzymes in the pathway as the pathway intermediates can inhibit enzymes in the pathway (de Jongh *et al.*, 2008). This led to the identification of a downstream enzyme, phosphoglucomutase (Pgm2) that converts glucose-1-phosphate to glucose-6-phosphate and has traditionally not been considered part of the Leloir pathway, as a flux controlling enzyme, and by over-expressing this single enzyme it was also found possible to significantly increase the specific growth rate (Bro *et al.*, 2005). In a later study where we used adaptive laboratory evolution, we found that combined over-expression of Pgm2 and engineering of a pathway engaged in glucose regulation it was possible to further enhance the growth rate on galactose as this ensured balancing of flux through all the individual reactions and therefore no accumulation of intermediates (Hong *et al.*, 2011). Engineering yeast to efficiently use xylose and arabinose has been attempted since the 1980s, but the discovery of a fungal xylose isomerase that converts xylose to xylulose, that yeast can metabolize, in a single step represented a breakthrough (Kuyper *et al.*, 2005). This also opened for engineering yeast to use arabinose (Wisselink *et al.*, 2007). However, yeast has not evolved to use xylose efficiently and the central carbon metabolism is therefore not well balanced to handle this carbon source. Using mathematical modeling to guide extensive engineering of the central carbon metabolism, it was possible to pinpoint key targets among more than 100 different genetic modifications, and hereby engineer yeast metabolism to efficiently grow on xylose as the sole carbon source (Li *et al.*, 2021).

Tolerance towards environmental stress is an important feature of cell factories, and it is therefore interesting to find strategies to **improve tolerance** of cell factories towards chemicals present in the medium, chemicals produced by the cell factory, low/high pH, and low/high temperature. Often a cell factory is chosen for its tolerance towards specific environmental conditions, e.g. *A. niger* is a well-suited cell factory for production of citric acid as it can tolerate low pH and *S. cerevisiae* is a well-suited cell factory for production of ethanol as it is very tolerant towards this chemical. However, in the current ethanol industry it is desirable to operate at a higher temperature than the optimum temperature for *S. cerevisiae*, but the biology associated with ensuring temperature tolerance is not known in detail. Here the concept of adaptive laboratory evolution was shown to be efficient as it enabled, through sequential selection of yeast clones, to identify clones that could grow faster at elevated temperatures, i.e. 40°C versus the optimum for the yeast of 35°C (Caspeta *et al.*, 2014). Through genome sequencing combined with metabolic modeling of isolated clones, it was then possible to identify causal mutations, and in particular one mutation was identified to cause a loss of function of an enzyme involved in biosynthesis of the membrane component ergosterol. The loss of function of this enzyme caused production of another sterol, namely fecosterol, and this resulted in a slightly stiffer membrane property that could enable the cells to better function at higher temperatures (Caspeta *et al.*, 2014). A similar approach has been applied to make yeast more tolerant to lactic acid at low pH (Fletcher *et al.*, 2016) and more tolerant to dicarboxylic acids (Pereira *et al.*, 2019), but it has also been used to improve tolerance of *E. coli* to various toxic chemicals (Lenne *et al.*, 2023).

Finally, engineering metabolism has been used to recruit different cell factories such as yeast for expression of **heterologous pathways**, and hereby enabling production of complex natural products. Thus, yeast has been recruited to produce complex plant chemicals such as opioids (Gelanie *et al.*, 2015), monoindole alkaloids that can be used as anti-cancer drugs (Zhang *et al.*, 2022), celastrol that has anti-obesity properties (Zhao *et al.*, 2023), berberine that has anti-diabetic properties (Jiao *et al.*, 2024), polyamines and polyamine conjugates (Qin *et al.*, 2021), and iso-flavonoids such as puerarin and daidzin that have cardioprotective properties (Liu *et al.*, 2021). Yeast is a well-suited cell factory for production of these complex plant natural products as biosynthesis of many of these molecules involves complex oxidation reactions catalyzed by so-called P450 enzymes. These enzymes contain a heme group in the catalytic core and both heme and iron can often become limited for function of these enzymes. However, through engineering of the heme biosynthetic pathway it was shown possible to enable elevated activity of P450 enzymes (Michener *et al.*, 2012). By expanding an enzyme constrained yeast GEM to include incorporation of metals it was possible to predict the effect of iron supplementation on biosynthesis of coumaric acid, that involves a single P450 enzyme (Chen *et al.*, 2021). Hereby it was shown how

GEMs can enable advancement of all the efforts on engineering yeast for production of natural products towards commercial production.

The above examples show that even though there have been many successful examples of engineering metabolism without guidance of mathematical models, these models have today come to a stage where they can significantly impact cell factory design (Domenzain *et al.*, 2024). The field is therefore aligning more with traditional engineering disciplines where mathematical models are actively used in design, and the design-build-test-learn cycle is therefore becoming well integrated into biology research (Nielsen and Keasling, 2016). Here it is interesting to note that even though cell factory development is generally carried out with a clear objective, i.e. to develop a cell factory that has improved properties, analysis of newly engineered cell factories often results in new understanding of cellular metabolism, and the border between engineering and basic biological research is therefore disappearing.

## Systems Biology of Metabolism

Systems biology is mathematical analysis and modeling of complex biological systems. It has two historical roots (Westerhoff and Palsson, 2004): 1) from theoretical biology, and 2) from genome-sequencing. The root from theoretical biology is often referred to as a bottom-up approach as it is based on detailed mathematical modeling of specific molecular processes whereas the root from genome-sequencing is often referred to as a top-down approach (Nielsen, 2017). Theoretical biology developed together with new understanding of molecular mechanisms in biology, as illustrated by the classical discovery of the mechanism on how expression of genes encoding for enzymes involved in metabolism of lactose in *E. coli*, present in the so-called Lac-operon. This genetic system was discovered by Francois Jacob and Jacques Monod, who received the Nobel prize in Physiology or Medicine in 1965. The molecular understanding of the Lac-operon has formed the basis for the development of detailed mathematical models (Lee and Bailey, 1984), and these pioneering studies led to an expansion of the field of mathematical models of biological systems. However, most of these models only describe a specific process within the cell and do not capture overall cellular metabolism. There have been attempts to develop whole cell models, e.g. a comprehensive model for *E. coli* (Karr *et al.*, 2012), but these models are often relying on description of several processes through empirical expressions rather than detailed molecular, mechanistic models. GEMs are in principle a merger of bottom-up and top-down approaches as these models are capturing individual enzymatic reactions, but by doing it genome-scale the model is relying on genome information. The models do, however, deviate from many mechanistic bottom-up models as they do not rely on mechanistic models describing the kinetics of each enzymatic reaction, and the models therefore only have very few parameters that needs to be estimated based on fitting to experimental data. This feature makes them attractive for wide use as it is hereby relatively easy to build the models based only on the genome sequence and stoichiometry of reactions.

Whereas the first GEMs were reconstructed in a bottom-up fashion, the availability of GEMs from many different organisms and extremely well curated models like Yeast8 (Lu *et al.*, 2019), it has become possible to almost automatically generate GEMs for different organisms, i.e. more than 350 different yeast species were reconstructed by using Yeast8 as a template (Lu *et al.*, 2021). These models were used to gain insight into many different aspects of fundamental knowledge of metabolism and evolution, i.e. how metabolism has evolved in different yeast species to adapt to various ecological niches. A similar approach was taken in using the well curated GEM for human metabolism, Human1, as a template model for generating GEMs for different model organisms like mouse, rat, zebrafish, fruit fly and nematode (Wang *et al.*, 2021). The hereby generated GEMs were used for identification of differences and commonalities in terms of metabolism between these very diverse species that are all used as model organisms for studying biology of human cells, tissues and organs. Specifically our mouse GEM was used to analyze data from mouse to gain new insight into how the development of Alzheimer's disease is associated with dramatic metabolic alterations in neuronal cells (Wang *et al.*, 2021).

Even though GEMs have a remarkable predictive strength, it is required to impose key constraints to the model for predicting a relevant phenotype. For example to predict the growth rate of a cell it is necessary to constrain the nutrient uptake rate or vice versa (Fig. 2A). Furthermore, GEMs can predict very large fluxes through pathways that in practice may have very little capacity, either due to low concentrations of the enzymes or due to low catalytic efficiency of the enzymes, i.e. low enzyme turnover numbers. To overcome this problem we developed the concept of enzyme constrained GEMs (ecGEMs) using the modeling framework we call GECKO (Sanchez *et al.*, 2017). In ecGEMs the flux through each individual enzyme is constrained by the turnover number and the enzyme concentration (Fig. 2B). This enables significant improvement in the predictive strength of these models, both in terms of flux distribution, growth on different carbon sources and prediction of overflow metabolism, i.e. the Crabtree effect (Sanchez *et al.*, 2017). A trade-off with these models is, however, that they

require information about the turnover numbers for all the enzymes, i.e.  $k_{cat}$ 's, and information about the enzyme levels. For well-studied organisms like *S. cerevisiae* and *E. coli* there are extensive databases, e.g. BRENDA, of  $k_{cat}$  values, and if there are no value available for some enzymes  $k_{cat}$  values determined for similar enzymes in other organisms can be used as a default. For many well studied enzymes there are even several different  $k_{cat}$  values reported, and in some cases these values even vary by an order of magnitude. In these cases, GECKO selects the largest values in order not to over-constrain flux through the reaction (Sanchez *et al.*, 2017). The GECKO framework has been further developed to automatically sample  $k_{cat}$  values from databases, and this has enabled faster reconstruction of ecGEMs for different microorganisms (Domenzain *et al.*, 2022). There is less information about the enzyme concentrations, and even though quantitative proteomics has advanced significantly, in particular for *S. cerevisiae* (Lahtvee *et al.*, 2017; Yu *et al.*, 2020; Di Bartolomeo *et al.*, 2020), it is still laborious to obtain high-quality proteome data, but in these cases there can be used a constraint about a total proteome allocation to metabolic enzymes (Fig. 2B), and it turns out that this allocation is remarkable constant across strains and different environmental conditions (Sanchez *et al.*, 2017).

ecGEMs have been shown to have strong predictive strength, most likely as they are rooted in a biological constraint that is deeply rooted in evolution, namely a constraint on protein synthesis rate by the ribosomes. For many organisms there is a linear correlation between ribosomal RNA content and specific growth rate, and quantitative proteomics has confirmed this for ribosomal protein content (Xia *et al.*, 2022). If the cell can reduce proteome required for certain part of metabolism, e.g. for biosynthesis of amino acids if these are supplied to the medium, this proteome mass can be allocated to ribosomes, and hereby the cell can grow faster (Björkeröth *et al.*, 2020). This clearly shows that proteome allocation within the cell is important and therefore also imposes overall constraints on metabolism. ecGEMs can therefore describe a phenomenon that has been a conundrum for many years, why do fast growing cells use metabolic pathways that provide less energy per unit glucose, i.e. overflow to ethanol in yeast cells (the Crabtree effect), to acetate in *E. coli* cells and to lactic acid in human cells, instead of using complete respiration that results in extraction of far more energy. In fact, ecGEMs can very well describe this overflow metabolism (Sanchez *et al.*, 2017) and using quantitative proteomics data it has been shown that the underlying reason is that overflow metabolism is more efficient in terms of energy not per glucose but **per proteome mass of the energy generating pathway** (Chen *et al.*, 2019). Thus, in nutrient excess, at fast growth, the cells have evolved to prioritize proteome allocation towards ribosomes rather than efficient energy generation, where fully respiratory metabolism is “proteome-expensive”.

GEMs are an excellent platform for integrative analysis of so-called -omics data, i.e. transcriptome, proteome and metabolome data. As these models are comprehensive in terms of covering all enzymes in the metabolic network of a cell and as there is a direct link to the encoding genes it is possible to directly overlay different -omics data onto the metabolic networks. More importantly it is even possible to combine this with statistical methods for identification of what we defined as so-called reporter metabolites (Fig. 3A) (Patil and Nielsen, 2005; Oliveira *et al.*, 2008). These are metabolites in the metabolic network for which there are significant alterations in expression of the enzymes that engage in chemical reactions with the metabolite, either using the metabolite as a substrate or producing the metabolite. Altered expression can be quantified from measuring gene expression, e.g. through mRNA sequencing, or through quantitative proteomics, and reporter metabolites therefore identifies spots within the metabolic network where there are altered enzyme levels either to maintain homeostasis of metabolism or to drive a metabolic change. With presence of high-quality metabolomics data it is also possible to identify reporter enzymes, that points to key enzymes involved in handling altered metabolite levels within the metabolic network (Cakir *et al.*, 2006). The concept of reporter metabolites has gained much traction and led to the development of PIANO, our platform that enables easy analysis of different types of data (Väremo *et al.*, 2013).

## Understanding metabolism for human health

GEMs have also been developed for human cells (Duarte *et al.*, 2007; Ma *et al.*, 2007; Mardinoglu *et al.*, 2013) and the consensus model Human1 was established using the same concept as we used for building the consensus Yeast8 (Robinson *et al.*, 2020). These models have been used extensively for mapping metabolic changes associated with disease development, e.g. for analyzing the metabolic changes in adipocytes in response to obesity (Mardinoglu *et al.*, 2013) and how liver metabolism becomes serine deficient in response to development of non-alcoholic fatty liver disease (NAFLD) (Mardinoglu *et al.*, 2014). GEMs have also been used to analyze metabolism during high-intensity exercise (Nilsson *et al.*, 2019). Besides enabling new understanding how metabolism responds to disease development, human GEMs find wide application for biomarkers identification and drug discovery.



Identification of reporter metabolites has shown to be very useful for **biomarker identification**. In a study of how metabolism is changing in 10 different cancers, we found that metabolism in the different cancers is more similar to the tissue of origin than among the different cancers (Gatto *et al.*, 2014). The metabolism in one cancer type, namely clear cell renal cellular carcinoma (ccRCC), was found to be quite distinct, and through further analysis a larger number of reporter metabolites were identified to be associated with glycosaminoglycan metabolism (Gatto *et al.*, 2014). Through measurements of 19 glycosaminoglycans in blood and urine in patients with metastatic ccRCC and healthy controls, a systems biomarker derived using machine learning from concentrations of the 19 metabolites could be identified (Gatto *et al.*, 2016). Further analysis showed that the systems biomarker could be used to detect early-stage solid tumors of ccRCC (Gatto *et al.*, 2018). These findings are now taken forward in clinical trials to validate a biomarker approved for detection of recurrence in ccRCC patients from both urine and blood (clinical study AURORAX-0087A; NCT04006405). We also found that the systems biomarker has been shown to have much wider applicability as it has been found possible to enable early detection of more than 14 different cancer types from a blood sample (Bratulic *et al.*, 2022). Cancer detection using liquid biomarkers has become attractive with a potential future market exceeding USD10B, much thanks to demonstration of early cancer detection through deep sequencing of DNA in blood samples (Jamshidi *et al.*, 2022). However, the systems biomarker based on measurements of glycosaminoglycans is attractive due to its lower cost and its combination of high selectivity and sensitivity (Bratulic *et al.*, 2022).

GEMs can also be used for drug discovery as was illustrated for both cancer and for disrupted metabolism associated with ageing (Folger *et al.*, 2011; Yizhak *et al.*, 2013). In one study we found that fatty acid oxidation in mitochondria plays a central role in hepatocellular carcinoma (HCC), the most dominant form of liver cancer, and by blocking transport of fatty acids to the mitochondria cancer cell growth could be blocked (Aagren *et al.*, 2014). Another example is the identification of serine deficiency in patients with NAFLD leading to use of our GEMs for drug discovery: serine deficiency, together with other metabolic alterations, led to proposing a cocktail of metabolites, i.e. serine, L-carnitine, nicotinamide riboside and N-acetyl-L-cysteine, that was found in clinical trials to aid the treatment of NAFLD (Zeybel *et al.*, 2021) and of mild-to-moderate COVID-19 (Altay *et al.*, 2021). The cocktail has also passed phase 2 in clinical trials for improving cognitive function in patients with Alzheimer Disease (Yulug *et al.*, 2023). In the cocktail serine and N-acetyl-L-cysteine serves as precursors for biosynthesis of the important anti-oxidant glutathione, L-carnitine ensures efficient transport of fatty acids to the mitochondria for  $\beta$ -oxidation and nicotinamide riboside serves as a precursor for the co-factor nicotinamide adenine dinucleotide (NAD<sup>\*</sup>).

Besides our own human cell, tissues and organs, we also have a very important metabolic organ composed of the human gut microbiome, which is a complex biological system comprising more than 1,000 different microorganisms, who communicate and exchange metabolites among each other and with our human cells. We have about as many microbial cells as human cells in our bodies, so it is no surprise that the quality and composition of this microbiome has been shown to impact human health in different ways (Karlsson *et al.*, 2013; Ji and Nielsen, 2015; Schmidt *et al.*, 2018). Most studies in this field are associative, but an increasing number of studies have now identified causality between the human gut microbiome composition and disease development, e.g. how the gut microbiome composition alters intestinal inflammation in colitis (Zhu *et al.*, 2018) and in the field of cancer treatment with immune therapies where we have shown that the presence of specific bacteria increases the response to treatment with check point inhibitors (Limeta *et al.*, 2020). The gut microbiome composition evolves based on dietary intake, but due to extensive metabolic interactions between the many different microorganisms it is difficult to predict how the composition exactly changes in response to diet. GEMs represent an excellent platform for analysis of metabolic interactions between the many different bacteria and the host (Karlsson *et al.*, 2011), and through reconstructing GEMs for three dominant gut bacteria, we showed that it is possible to simulate bacterial interactions, and how growth of the three bacterial species depends on the nutrients provided (Shoai *et al.*, 2013). This concept further enabled simulation of how bacteria in the human gut microbiome contribute to amino acid biosynthesis in the human body, and hereby how the levels of amino acids change in the blood when human subjects undergo dietary interventions (Shoai *et al.*, 2015). In this study a group of overweight individuals were provided with a low-calorie diet for 6-weeks, and it was found that the response to the dietary change was dependent on the gut microbiome composition. Modeling showed that subjects with a less diverse microbiome responded better to dietary intervention, i.e. their health status measured by several blood markers, including amino acid levels, than subjects with a very diverse microbiome (Shoai *et al.*, 2015). The findings were later confirmed in a controlled mice study (Mardinoglu *et al.*, 2015). With the objective of enabling prediction of how diet impacts the human gut microbiome composition, modeling of the microbiome was combined with a detailed mathematical model of the entire human gastrointestinal system, and this enabled for the first time simulation of how the gut microbiome evolves in infants

when they change their diet from breast milk to solid food (Geng *et al.*, 2021). These modeling efforts of the human gut microbiome will enable better design of dietary interventions with the objective of modulating the human gut microbiome, but it will also enable identification of new probiotics that can be used to ensure maintenance of healthy gut microbiome composition, and hereby contribute to improved human health. Currently the world market for probiotics exceeds USD50B and this is likely going to increase significantly in the future when it will be possible to design better products that have clinically validated health claims.

## AI for metabolism

Use of computational tools for analysis and engineering of biological complex systems, such as metabolism is at a breaking point as standard mathematical modeling and GEMs have not fully solved the three major challenges in current biotechnology: 1) recombinant proteins used as pharmaceuticals (with an increasing market share of the total pharmaceutical market, which in itself is growing), are relatively expensive and it will be necessary to reduce production costs in order to make the drugs more widely available at fair price; 2) engineered microbes for production of chemicals and new products need not only new cell factories but also improved design processes in order to reduce the development costs (Nielsen *et al.*, 2022); and 3) finding biomarkers and drug targets for improving human health needs faster validation, and as new experimental platforms like human organoids are increasingly more useful there is a need for holistic understanding of metabolism in the whole human body. For all these challenges we have been using modeling and GEMs, but the outcomes will significantly improve when we combine them with use of artificial intelligence (AI).

GEMs developed for protein secretion by yeast (Li *et al.*, 2022a), CHO cells (Gutierrez *et al.*, 2020), and human cells (Feizi *et al.*, 2017; Robinson *et al.*, 2019) are already now used for design of cell factories that can more efficiently and cheaper produce a wide range of different recombinant proteins. However, even though we have considerable knowledge of the protein secretory pathway in these organisms, there are significant gaps in our understanding: the combinatorial space for engineering many target proteins involved in this pathway makes it difficult to find optimal design strategies but we predict that AI can improve the effectiveness in target identification. With the development of ecGEMs we have already demonstrated optimal cell factories design that can be used for production of a range of different chemicals. However, ecGEMs rely on kinetic parameters, i.e. turnover number, that in many cases are unknown. Additionally, many enzymes are catalytically promiscuous or reversible, resulting in reactions that lead to undesirable by-products or degradation of the product of interest. To address this challenge, we have built an AI model that helped us obtain  $k_{cat}$  values for all enzymes present in more than 350 yeast species (Li *et al.*, 2022b). These parameters could be then used to populate functional ecGEMs for all these microbial species. We have democratized this approach by establishing an open access database, GotEnzymes, that holds estimates of  $k_{cat}$  values for more than 25 million enzyme-compound pairs across 8,099 organisms (Li *et al.*, 2023). We also created an AI model to map the so-called under-ground metabolism, i.e. promiscuous functions of enzymes. The model trained on known enzyme-compound interactions, and we identified about 15,000 new reactions in yeast, that produce 15,873 new metabolites, of which the majority are engaged in lipid metabolism (Wu *et al.*, 2025). This demonstrated that metabolism is much more diverse than captured by traditional GEMs. Using the model, it was possible to identify many metabolites that can be formed as by-products, and this can now be used to guide the design of less promiscuous enzymes in metabolism that will improve biotechnological based production. In the future AI could enable development of better ecGEMs, but also enable combination of model simulations with large experimental data obtained from large research programs using biofoundries, for even stronger prediction for optimized cell factory designs.

In the area of human health AI will also play an important role in the future. So far most applications have been for image processing and more recently for protein structure predictions, provided by AlphaFold (Jumper *et al.*, 2021), for which Demis Hassabis and John Jumper received half of the 2024 Nobel Prize in Chemistry. Going further using AI for novel biologics design is, however, more challenging as we are dealing with an almost infinite number of combinations of amino acid sequences, with a desired targeted structure and properties. Breakthroughs are initially likely going to come through design of smaller peptides or through evaluation of small molecule-protein interactions, where AI is very well suited (Watson *et al.*, 2023), and for which David Baker received the other half of the 2024 Nobel Prize in Chemistry. For analysis of human metabolism and identification of novel drugs that target metabolism, the combination of GEMs with AI may be useful as GEMs can be used to generate large datasets that can then be used to train new AI models. For this there is a need for whole body metabolic models that are built based on GEMs for different organs combined with a model describing blood circulation. With such models the use of GEMs will be integrated into any drug development strategy as it will

enable reduced experimental costs and reduced use of experimental animals, and such comprehensive models are very well suited to combine with AI for drug discovery.

In conclusion, we foresee that the application of GEMs, both alone and in combination with AI, will enable provision of many new solutions that can improve both human and planetary health.

## Acknowledgements

We would like to acknowledge all our fantastic students, post docs and research collaborators over the last 30 years.

## Author Contributions

JN and DP designed the structure of the manuscript. JN wrote the manuscript. DP edited and revised the manuscript.

## Financial Support

This work was financially supported by the Novo Nordisk Foundation (grant no. NNF20CC0035580).

## Conflict of interest declarations

Jens Nielsen is a shareholder of Elypta AB, Melt & Marble AB and Chrysea Inc.

## Ethical standards statement

Do not apply as no new research was conducted in connection with this paper.

## Data availability statement

Do not apply as no new data were generated in connection with this paper.

## References

- Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I and Nielsen J (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Computational Biology* 9:e1002980
- Agren R, Mardinoglu A, Kampf C, Asplund A, Uhlen M and Nielsen J (2014) Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Molecular Systems Biology* 10:721
- Aiba S and Matsuoka M (1979) Identification of metabolic model: Citrate production from glucose by *Candida lipolytica*. *Biotechnology and Bioengineering* 21, 1373-1386
- Altay O, Arif N, Li X, Yang H, Aydin M, Alkurt G, Kim W, Akyol D, Zhang C, Dinler-Doganay G, Turkez H, Shoaie S, Nielsen J, Boren J, Olmuscelik O, Doganay L, Uhlen M and Mardinoglu A (2021) Combined metabolic activators (CMA) accelerates recovery in mild-to-moderate COVID-19. *Advanced Science* 8:2101222
- Andersen MR, Nielsen ML and Nielsen J (2008) Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Molecular Systems Biology* 4:178
- Bailey JE (1991) Towards a science of metabolic engineering. *Science* 252, 1668-74
- Björkeröth J, Campbell K, Malina C, Yu R, Di Bartolomeo F and Nielsen J (2020) Proteome re-allocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proceedings of National Academy of Science USA* 117:21804-21812
- Borodina I, Krabben P and Nielsen J (2005) Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Research* 15:820-829
- Bratulis S, Limeta A, Dabestani S, Birgisson H, Enblad G, Ståhlberg K, Hesselager G, Häggman M, Höglund M, Simonson OE, Ståhlberg P, Lindman H, Bång-Rudenstam A, Ekstrand M, Kumar G, Cavarretta I, Alfano M, Pellegrino F, Mandel-Clausen T, Salanti A, Maccari F, Galeotti F, Volpi N, Daugaard M, Belting M, Lundstam S, Steiner U, Nyman J, Bergman B, Edqvist PH, Levin M, Salonia A, Kjölhede H, Jonasch E, Nielsen J and Gatto F (2022) Noninvasive detection of any-stage cancer using free glycosaminoglycans. *Proceedings of National Academy of Science USA* 119:e2115328119
- Bro C, Knudsen S, Regenberg B, Olsson L, Nielsen J (2005) Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucomutase: Example of transcript analysis as a tool in inverse metabolic engineering. *Applied and Environmental Microbiology* 71:6465-6472

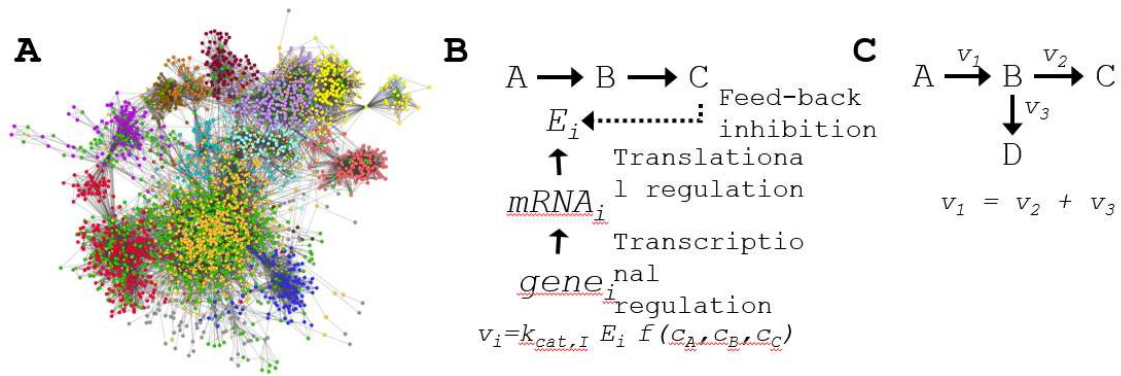
- Bro C, Regenberg B, Förster J and Nielsen J** (2006) *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metabolic Engineering* 8:102-111
- Cakir T, Patil KR, Önsan ZI, Ülgen KÖ, Kirdar B and Nielsen J** (2006) Integration of metabolome data with metabolic networks reveals reporter reactions. *Molecular Systems Biology* 2:50
- Caspeta L, Shoai S, Agren R, Nookaew I and Nielsen J** (2012) Genome-scale metabolic reconstruction of *Pichia stipidis* and *Pichia pastoris* and *in silico* evaluation of their potentials. *BMC Systems Biology* 6, 24
- Caspeta L, Chen Y, Ghiaci P, Feizi A, Buskov S, Hallström BM, Petranovic D and Nielsen J** (2014) Altered sterol composition renders yeast termotolerant. *Science* 346, 75-78
- Chen Y and Nielsen J** (2019) Energy metabolism controls phenotypes by protein efficiency and allocation. *Proceedings of National Academy of Science USA* 116, 17592-17597
- Chen Y, Li F, Mao J, Chen Y and Nielsen J** (2021) Yeast optimizes metal utilization based on metabolic network and enzyme kinetics. *Proceedings of National Academy of Science USA* 118, e2020154118
- Di Bartolomeo F, Malina C, Campbell K, Mormini M, Fuchs J, Vorontsov E, Gustafsson CM and Nielsen J** (2020) Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proceedings of National Academy of Science USA* 117:7524-7535
- Domenzain I, Sanchez BJ, Anton M, Kerkhoven EJ, Millan-Oropeza A, Henry C, Siewers V, Morrissey JP, Sonnenschein N and Nielsen J** (2022) Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO2.0. *Nature Communications* 13, 3766
- Domenzain I, Lu Y, Shi J, Lu H and Nielsen J** (2024) Computational biology predicts metabolic engineering targets for increased production of 103 valuable chemicals in yeast. *Proceedings of National Academy of Science USA*, in revision
- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R and Palsson BO** (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of National Academy of Science USA* 104, 1777-1782
- Edwards JS and Palsson BO** (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry* 274, 17410-17416
- Edwards JS and Palsson BO** (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of National Academy of Science USA* 97, 5528-5533
- Feizi A, Österlund T, Petranovic D, Bordel S and Nielsen J** (2013) Genome-scale modeling of the protein secretory machinery in yeast. *PLOS One* 8:e63284
- Feizi A, Gatto F, Uhlen M and Nielsen J** (2017) Human protein secretory pathway genes are expressed in a tissue-specific pattern to match secretion demand. *Systems Biology Applications* 3:22
- Fletcher E, Feizi A, Hallström B, Bisschops M, Koomrung S, Siewers V and Nielsen J** (2016) Evolutionary engineering renews divergent paths when yeast is adapted to different acidic environments. *Metabolic Engineering* 113:1164-1170
- Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E and Shlomi T** (2011) Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology* 7, 501
- Förster J, Famili I, Fu P, Palsson BO and Nielsen J** (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research* 13, 244-253
- Gatto F, Nookaew I and Nielsen J** (2014) Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proceedings of National Academy of Science* 111, E866-E875
- Gatto F, Nookaew I, Nilsson H, Maruzzo M, Roma A, Johansson ME, Steiner U, Lundstam S, Volpi N, Basso U and Nielsen J** (2016) Measurements of glycosaminoglycans in plasma and urine for diagnosis of clear cell renal cell carcinoma. *Cell Report* 15:1-15
- Gatto F, Blum KA, Hosseini SS, Ghanaat M, Kashan M, Maccari F, Galeotti F, Hsieh J, Volpi N, Hakimi AA and Nielsen J** (2018) Plasma glycosaminoglycans as diagnostic and prognostic biomarkers in surgically treated renal cell carcinoma. *European Urology Oncology* 1:364-377
- Gelanie S, Thodey K, Trenchard IJ, Interrante MF and Smolke CD** (2015) Complete biosynthesis of opioids in yeast. *Science* 349, 1095-1100
- Geng J, Ji B, Li G, Lopez-I sunza F and Nielsen J** (2021) CODY enables quantitatively spatiotemporal predictions on *in vivo* gut microbial variability induced by diet-intervention. *Proceedings of National Academy of Science USA* 118:e2019336118
- Gutierrez J, Feizi A, Li S, Kallehauge T, Lee GM, Palsson B, Nielsen J and Lewis N** (2020) Genome-scale reconstruction of the mammalian secretory pathway predicts metabolic costs and limitations of protein secretion. *Nature Communications* 11:68
- Gu C, Kim GB, Kim WJ, Kim HU and Lee SY** (2019) Current status and applications of genome-scale metabolic models. *Genome Biology* 20, 121
- Hohmann S** (2016) Nobel yeast research. *FEMS Yeast Research* 16, fow094
- Hong KK, Vongsangnak W, Vemuri GN and Nielsen J** (2011) Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proceedings of National Academy of Science USA* 108, 12179-12184
- Huang M, Bao J and Nielsen J** (2014) Biopharmaceutical protein production by *Saccharomyces cerevisiae*: Current state and future prospects. *Pharmaceutical Bioprocessing* 2:167-182
- Huang M, Bai Y, Sjoström SL, Hallström MB, Liu Z, Petranovic D, Uhlen M, Joensson, HN, Andersson-Svahn H and Nielsen J** (2015) Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proceedings of National Academy of Science USA* 112, E4689-E4696

- Huang M, Bao J, Hallström BM, Petranovic D and Nielsen J (2017) Efficient protein production by yeast requires global tuning of metabolism. *Nature Communications* **8**, 1131
- Huang M, Wang G, Qin J, Petranovic D and Nielsen J (2018) Engineering the protein secretory pathway of *Saccharomyces cerevisiae* enables improved protein production. *Proceedings of National Academy of Science USA* **115**:E11025-E11032
- Irani ZA, Kerkhoven EJ, Shojaosadati SA and Nielsen J (2015) Genome-scale metabolic model of *Pichia pastoris* with native and humanized glycosylation of recombinant proteins. *Biotechnology and Bioengineering*. **113**, 961-969
- Jamshidi A, Liu MC, Klein EA, Venn O, Hubbell E, Beausang JF, Gross S, Melton C, Fields AP, Liu Q, Zhang N, Fung ET, Kurtzman KN, Amini H, Betts C, Civallo D, Freese P, Calef R, Davydov K, Fayzullina S, Hou C, Jiang R, Jung B, Tang S, Demas V, Newman J, Sakarya O, Scott E, Shenoy A, Shojaee S, Steffen KK, Nicula V, Chien TC, Bagaria S, Hunkapiller N, Desai M, Dong Z, Richards DA, Yeatman TJ, Cohn AL, Thiel DD, Berry DA, Tummala MK, McIntyre K, Sekeres MA, Bryce A, Aravanis AM, Seiden MV and Swanton C (2022) Evaluation of cell-free DNA approaches for multi-cancer early detection. *Cancer Cell* **40**, 1537-1549
- Ji B and Nielsen J (2015) New insight into the gut microbiome through metagenomics. *Advanced Genomics and Genetics* **5**:77-91
- Jiao X, Fu X, Li Q, Bu J, Liu X, Savolainen O, Huang L, Guo J, Nielsen J and Chen Y (2024) De novo production of protoberberine and benzophenanthridine alkaloids through metabolic engineering of yeast. *Nature Communications* **15**, 8759
- de Jongh WA, Bro C, Ostergaard S, Regenber B, Olsson L and Nielsen J (2008) The roles of galactitol, galactose-1-phosphate and phosphoglucomutase in galactose-induced toxicity in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* **101**:317-326
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589
- Jørgensen H, Nielsen J, Villadsen J and Mølgaard H (1995) Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnology and Bioengineering* **46**:117-131
- Kacser H and Burns JA (1973) The control of flux. *Symposium of Society of Experimental Biology* **27**, 65-104
- Karlsson FH, Nookaew I, Petranovic D and Nielsen J (2011) Prospects for systems biology and modeling of the gut microbiome. *Trends Biotechnology* **29**:251-258
- Karlsson F, Tremaroli V, Nielsen J and Bäckhed F (2013) Assessing the human gut microbiota in metabolic diseases. *Diabetes* **62**:3341-3349
- Karr JR, Sangvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI and Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389-401
- Kell DB, van Dam K and Westerhoff HV (1989) Control analysis of microbial growth and productivity. *Symposium of Society of General Microbiology* **44**, 61-93
- Konzock O and Nielsen J (2024) TRYing to evaluate production costs in microbial biotechnology. *Trends Biotechnology* **42**, 1339-1347
- Kuyper M, Hartog MMP, Toirkens MJ, Almering MJH, Winkler AA, van Dijken FP and Pronk JT (2005) Metabolic engineering of a xylose-isomerase expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation. *FEMS Yeast Research* **5**, 399-409
- Lahtvee PJ, Sanchez BJ, Smialowska A, Kasvandik S, Elsemman I, Gatto F and Nielsen J (2017) Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. *Cell Systems* **4**:495-504
- Lee SB and Bailey JE (1984) Genetically structured models for lac promoter-operator function in the chromosome and in multicopy plasmids: Lac promoter function. *Biotechnology and Bioengineering* **26**, 1383-1389
- Lennen RM, Lim HG, Jensen K, Mohammed ET, Phaneuf PV, Noh MH, Malla S, Börner RA, Chekina K, Özdemir E, Bonde I, Koza A, Maury J, Pedersen LE, Schöning LY, Sonnenschein N, Palsson BO, Nielsen AT, Sommer MOA, Herrgård MJ and Feist AM (2023) Laboratory evolution reveals general and specific tolerance mechanisms for commodity chemicals. *Metabolic Engineering*. **76**, 179-192
- Li F, Chen Y, Qi Q, Wang Y, Yuan L, Huang M, Elsemman IE, Feizi A, Kerkhoven EJ and Nielsen J (2022a) Improving recombinant protein production by yeast through genome-scale modelling using proteome constraints. *Nature Communications* **13**:2969
- Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ and Nielsen J (2022b) Deep learning based  $k_{cat}$  prediction enables improved enzyme constrained model reconstruction. *Nature Catalysis* **5**:662-672
- Li F, Chen Y, Anton M and Nielsen J (2023) GotEnzymes: An extensive database of enzyme parameter predictions. *Nucleic Acid Research* **51**:D583-D586
- Li X, Wang Y, Lo G, Liu Q, Pereira R, Chen Y and Nielsen J (2021) Metabolic network remodeling enhances yeast's fitness on xylose using aerobic glycolysis. *Nature Catalysis*. **4**:783-796
- Limeta A, Ji B, Levin M, Gatto F and Nielsen J (2020) Meta-analysis of the gut microbiota in predicting response to cancer immunotherapy in metastatic melanoma. *Journal Clinical Investigation Insight* **5**:e140940
- Liu Q, Liu Y, Li G, Savolainen O, Chen Y and Nielsen J (2021) De novo biosynthesis of bioactive isoflavonoids by engineered yeast cell factories. *Nature Communications* **12**:6085

- Lu H, Li F, Sanchez BJ, Zhu Z, Li G, Domenzain I, Marcisauskas S, Anton PM, Lappa D, Lieven C, Beber ME, Sonnenschein N, Kerkhoven EJ and Nielsen J (2019) A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nature Communications* **10**:3586
- Lu H, Li F, Yuan L, Domenzain I, Yu R, Wang H, Li G, Chen Y, Ji B, Kerkhoven EJ and Nielsen J (2021) Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection. *Mol. Systems Biol.* **17**:e10427
- Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O and Goryanin I (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Molecular Systems Biology* **3**, 135
- Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobsen P, Walley AJ, Froguel P, Carlsson LM, Uhlen M and Nielsen J (2013) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Molecular Systems Biology* **9**:649
- Mardinoglu A, Agren R, Kampf K, Asplund A, Uhlen M and Nielsen J (2014) Genome-scale metabolic modeling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications* **5**:3083
- Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, Bäckhed F and Nielsen J (2015) The gut microbiome modulates host amino acid and glutathione metabolism in mice. *Molecular Systems Biology* **11**:834
- Michener JK, Nielsen J and Smolke CD (2012) Identification and treatment of heme depletion attributable to overexpression of a lineage of evolved P450 monooxygenases. *Proceedings of National Academy of Science USA* **109**:19504-19509
- Nakamura CE and Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion Biotechnology* **14**, 454-459
- Nielsen J (2013) Production of biopharmaceutical proteins by yeast. *Bioengineered* **4**, 207-211
- Nielsen J (2017) Systems Biology of Metabolism. *Annual Review of Biochemistry* **86**:245-275
- Nielsen J (2019) Yeast Systems Biology: Model organism and cell factory. *Biotechnology Journal* **14**, 1800421
- Nielsen J and Keasling JD (2016) Engineering Cellular Metabolism. *Cell* **164**:1185-1197
- Nielsen J, Tillegren CB and Petranovic D (2022) Innovation trends in industrial biotechnology. *Trends Biotechnology* **40**,1160-1172
- Nilsson A, Björnson E, Flockhart M, Larsen FJ and Nielsen J (2019) Complex I is bypassed during high intensity exercise. *Nature Communications* **10**:5072
- Nissen TL, Schulze U, Nielsen J and Villadsen J (1997) Flux distributions in anaerobic, glucose limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* **143**:203-218
- Nissen TL, Kielland-Brandt MC, Nielsen J and Villadsen J (2000) Optimisation of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonia assimilation. *Metabolic Engineering* **2**:69-77
- Oliveira AP, Patil KR and Nielsen J (2008) Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Systems Biology* **2**:17
- Ostergaard S, Olsson L, Johnston M, Nielsen J (2000) Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. *Nature Biotechnology* **18**, 1283-1286
- Patil KR and Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of National Academy of Science USA* **102**:2685-2689
- Pereira R, Wei Y, Mohamed E, Radi M, Malina C, Herrgård MJ, Feist AM, Nielsen J and Chen Y (2019) Adaptive laboratory evolution of tolerance to dicarboxylic acids in *Saccharomyces cerevisiae*. *Metabolic Engineering* **56**:130-141
- Qin J, Krivoruchko A, Ji B, Chen Y, Kristensen M, Özdemir E, Keasling JD, Jensen MK and Nielsen J (2021) Engineering yeast metabolism for the discovery and production of polyamines and polyamine analogues. *Nature Catalysis* **4**:498-509
- Qin N, Li, L, Ji X, Pereira R, Chen Y, Yin S, Li C, Wan X, Qiu D, Jiang J, Lou H, Zhang Y, Dong G, Zhang Y, Shi S, Jessen HJ, Xia J, Chen Y, Larsson C, Tan T, Liu Z and Nielsen J (2023) Flux regulation through glycolysis and respiration is balanced by inositol pyrophosphates. *Cell* **186**:748-76
- Robinson JL, Feizi A, Uhlen M, Nielsen J (2019) A systematic investigation of the malignant functions and diagnostic potential of the cancer secretome. *Cell Report* **10**:2622-2635
- Robinson JL, Kocabas P, Wang H, Cholly PE, Cook D, Nilsson A, Anton M, Ferreira R, Domenzain I, Billa V, Limeta A, Hedin A, Gustafsson J, Kerkhoven EJ, Svensson LT, Palsson BO, Mardinoglu A, Hansson L, Uhlen M and Nielsen J (2020) An Atlas of Human Metabolism. *Science Signal.* **13**:eaz1482
- Sanchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven E and Nielsen J (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology* **13**:935
- Schilling CH, Covert MW, Famili I, Church GM, Edwards JS and Palsson BO (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *Journal Bacteriology* **184**, 4582-4593
- Schmidt TSB, Raes J and Bork P (2018) The human gut microbiome: From associations to modulation. *Cell* **172**, 1198-1215
- Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S and Nielsen J (2013) Elucidating the interactions between bacteria in the human gut through metabolic modeling. *Science Report* **3**:2532
- Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, de Wouters T, Juste C, Rizkalla S, Chilloux J, Hoyles L, Nicholson JK, MicroObese Consortium ANR, Dore J, Dumas ME, Clement K, Bäckhed F and Nielsen J (2015) Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metabolism* **22**:320-331
- Stephanopoulos G and Vallino JJ (1991) Network rigidity and metabolic engineering in metabolite over-production. *Science* **252**, 1675-81
- Stephanopoulos G, Aristidou A and Nielsen J (1998) *Metabolic Engineering*, Elsevier

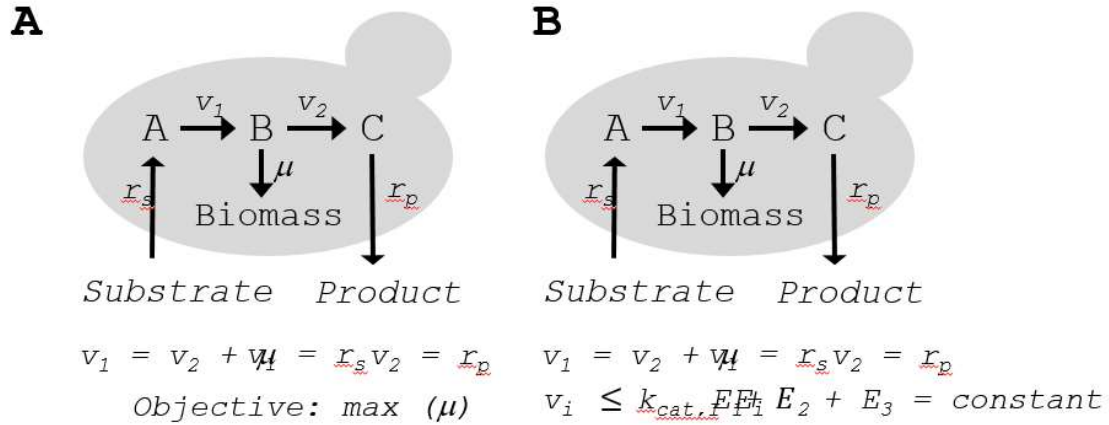
- Vallino JJ and Stephanopoulos G** (1993) Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnology Bioengineering* **41**, 633-646
- Varma A and Palsson BO** (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied Environmental Microbiology* **60**, 3724-3731
- Vongsangnak W, Olsen P, Hansen K, Krogsgaard S and Nielsen J** (2008) Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genomics* **9**:245
- Väremo L, Nielsen J and Nookaew I** (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acid Research* **41**:4378-4391
- Wang H, Robinson JL, Kocabas P, Gustafsson J, Anton M, Cholley PE, Huang S, Gobom J, Svensson T, Uhlen M, Zetterberg H and Nielsen J** (2021) Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proceedings of National Academy of Science USA* **118**:e2102344118
- Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, Wicky BIM, Hanikel N, Pellock SJ, Courbet A, Sheffler W, Wang J Venkatesh P, Sappington I, Torres SV, Lauko A, Bortoli VD, Mathieu E, Ovchinnikov S, Barzilay R, Jaakkola TS, DiMaio F, Baek M and Baker D** (2023) De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100
- Westerhoff HV and Palsson BO** (2004) The evolution of molecular biology into systems biology. *Nature Biotechnology* **22**, 1249-1252
- Wisselink HW, Toirkens MJ, Berriel MRF, Winkler AA, van Dijken JP, Pronk JT and van Maris AJA** (2007) Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose. *Applied and Environmental Microbiology* **73**, 4881-4891
- Wu K, Liu H, Sun M, Mao R, Jiang Y, Kerkhoven EJ, Chen Y, Nielsen J and Li F** (2024) Yeast-MetaTwin for systematically exploring yeast metabolism through retrobiosynthesis and deep learning. *Nature Catalysis*, in review
- Xia J, Sanchez BJ, Chen Y, Campbell K, Kasvandik S and Nielsen J** (2022) Proteome allocation change linearly with specific growth rate of *Saccharomyces cerevisiae* under glucose-limitation. *Nature Communications* **13**:2819
- Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ and Van Dien S** (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology* **7**, 445-452
- Yizhak K, Gabay O, Cohen H and Ruppin E** (2013) Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nature Communications* **4**, 2632
- Yu R, Campbell K, Pereira R, Björkeröth J, Qi Q, Vorontsov E, Sihlbom C and Nielsen J** (2020) Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. *Nature Communications* **11**:1881
- Yu T, Zhou Y, Huang M, Liu Q, Pereira R, David F and Nielsen J** (2018) Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. *Cell* **174**:1-10
- Yulug B, Alta O, Li X, Hanoglu L, Cankaya S, Lam S, Velioglu HA, Yang H, Coskun E, Idil E, Nogaylar R, Ozsimsek A, Bayram C, Bolat I, Oner S, Tozlu OO, Arslan ME, Hacimuftuoglu A, Yildirim S, Arif M, Shoaie S, Zhang C, Nielsen J, Turkez H, Borén J, Uhlén M and Mardinoglu A** (2023) Combined metabolic activators improve cognitive functions in Alzheimer's disease patients: a randomised double-blinded, placebo-controlled phase-II trial. *Translational Neurodegeneration* **12**:4
- Zeybel M, Yang H, Altay O, Arif M, Fredolini C, Akyildiz M, Saglam B, Gonenli MG, Ural D, Kim W, Li X, Schwenk JM, Zhang C, Shoaie S, Nielsen J, Uhlen M, Boren J and Mardinoglu A** (2021) Combined metabolic cofactor supplementation ameliorates liver fat in NAFLD. *Molecular Systems Biology* **17**:e10459
- Zhang C, Sanchez BJ, Li F, Quan ECW, Farias SM, Liebal U, Blank LM, Mengers H, Anton M, Range AT, Zhang L, J. Nielsen, Lu H and Kerkhoven EJ** (2024) Yeast9: A yeast metabolic enables quantitative analysis of cellular metabolism by incorporating big data. *Molecular Systems Biology* **20**, 1134-1150
- Zhang J, Hansen LG, Gudich O, Viehrig K, Lassen LMM, Schrübbers L, Adhikari KB, Rubaszka P, Carrasquer-Alvarez E, Chen L, D'Ambrosio V, Lehka B, Haidar AK, Nallapareddy S, Giannakou K, Laloux M, Arsovska D, Jørgensen MAK, Chan LJG, Kristensen M, Christensen HB, Sudarsan S, Stander EA, Baidoo E, Petzold CJ, Wulff T, O'Connor SE, Coudavault V, Jensen MK and Keasling JD** (2022) A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* **609**, 341-347
- Zhao Y, Hansen NL, Duan YT, Prasad M, Motawia MS, Møller BL, Pateraki I, Staerk D, Bak S, Miettinen K and Kampranis SC** (2023) Biosynthesis and biotechnological production of the anti-obesity agent celastrol. *Nature Chemistry* **15**, 1236-1246
- Zhou Y, Buijs NA, Zhu Z, Qin J, Siewers V and Nielsen J** (2016) Production of fatty acid derived oleochemicals and biofuels by synthetic yeast cell factories. *Nature Communications* **7**, 11709
- Zhu W, Winter MG, Byndloss MX, Spiga L, Duerkop BA, Hughes ER, Büttner L, Romao EL, Behrendt CL, Lopez CA, Sifuentes-Dominguez L, Huff-Hardy K, Wilson RP, Gillis CC, Tükel C, Koh AY, Burstein E, Hooper LV, Bäuml AJ and Winter SE** (2018) Precision editing of the gut microbiota ameliorates colitis. *Nature* **553**, 208-211

**Fig. 1** Metabolic networks and how their fluxes can be regulated. **A.** Illustration of a typical hairball metabolic network. Green dots are enzymes and the dots with different colors are metabolites interacting with the enzymes. The metabolites are color coded according to their cellular compartment. Most metabolites are in the cytosol (yellow dots) and the mitochondria (red dots). **B.** Simple representation of the different layers of regulation of flux through a reaction that converts metabolite A to B. The reaction is catalyzed by the enzyme  $E_i$  and the flux is a function of the catalytic capacity (turnover number) of this enzyme, i.e.  $k_{cat,i}$ , the concentration of the enzyme ( $E_i$ ), and a function of the different metabolites in the network ( $f$ ). In the simple model there is feedback inhibition of the enzyme by metabolite C, and the flux is therefore determined by the concentration of the three metabolites A, B and C. The enzyme concentration is determined by transcriptional regulation of the corresponding gene and by translational regulation of the corresponding mRNA. **C.** Simple illustration of the concept of flux balancing. In this simple network the three fluxes are constrained by a simple mass balance around the metabolite B. For most intracellular metabolites the turnover is so high that an assumption of steady state, resulting in the simple algebraic constraint equation, is reasonable. If the cells are experiencing a significant environmental change there will be a short period of time where the steady state assumption does not apply, but as the characteristic time constant for most metabolite concentrations, i.e. the concentration of the metabolite divided by the flux through the metabolite, is in the order of seconds (rarely minutes), a new steady state level of the metabolites will rapidly be obtained. Therefore, the simple balance equation is in practice always valid.





**Fig. 2** Constraints of GEMs and how they impact flux estimation. **A.** In simple flux balance analysis where model simulation is based on the mass balances illustrated in Fig. 1C it is necessary to constrain either one input or one output flux combined with an objective function, here illustrated by maximizing the specific growth rate  $\mu$ . This is often one of two options: 1) the substrate uptake rate is defined and there is optimized for growth rate, or 2) the growth rate is defined and there is minimized for substrate uptake rate. If  $r_s$  is given and there is maximized for  $\mu$  then the model will obviously not predict any product formation, i.e.  $r_p$  is zero. **B.** By constraining the flux through each of the reactions by the enzyme turnover number ( $k_{cat}$ ) and the enzyme concentration ( $E_i$ ) it is not necessary to constrain any input or output flux and the model therefore has better predictive strength. Either the enzyme concentrations can be given as input, or the sum of all enzyme concentrations is capped at a constant value that is consistent with experimental measurements.



**Fig. 3** Use of GEMs for integrative analysis of omics data. **A.** Using the graphical structure of GEMs it is possible to identify Reporter Metabolites, which are metabolites in the metabolic network around which there are significant changes in transcript level or protein level. The lines represent enzyme catalyzed reactions and the circles are metabolites. The thickness of the lines indicates the changes in enzyme levels, measured by transcripts or protein levels. Metabolites around which there are large changes in enzyme levels become reporter metabolites and the significance is marked by the greyness, with dark grey being very significant and light grey less significant. **B.** The graphical structure of GEMs can enable identification of Reporter Networks, which are sub-networks where there are significant changes in the transcription level or protein level. Two sub-networks are marked with light grey circles.

