An expanded role for microbial physiology in metabolic engineering and functional genomics: moving towards systems biology

Jens Nielsen, Lisbeth Olsson

Center for Process Biotechnology, Building 223, BioCentrum-DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Received 4 October 2001; received in revised form 24 December 2001; accepted 28 December 2001

First published online 6 February 2002

Abstract

Microbial physiology has traditionally played a very important role in both fundamental research and in industrial applications of microorganisms. The classical approach in microbial physiology has been to analyze the role of individual components (genes or proteins) in the overall cell function. With the progress in molecular biology it has become possible to optimize industrial fermentations through introduction of directed genetic modification – an approach referred to as metabolic engineering. Furthermore, as a consequence of large sequencing programs the complete genomic sequence has become available for an increasing number of microorganisms. This has resulted in substantial research efforts in assigning function to all identified open reading frames – referred to as functional genomics. In both metabolic engineering and functional genomics there is a trend towards application of a macroscopic view on cell function, and this leads to an expanded role of the classical approach applied in microbial physiology. With the increased understanding of the molecular mechanisms it is envisaged that in the future it will be possible to describe the interaction between all the components in the system (the cell), also at the quantitative level, and this is the goal of systems biology. Clearly this will have a significant impact on microbial physiology as well as on metabolic engineering.

Keywords: Glucose repression; Mathematical modeling; Pathway reconstruction; Hierarchical control; Aspergillus; Saccharomyces cerevisiae

1. Introduction

Microbial physiology is a classical discipline where the aim is to perform a qualitative and/or quantitative characterization of certain microbial species. Traditionally microbial physiology involved a macroscopic view on cell function, where a given wild-type species was carefully evaluated, e.g. for its growth on different carbon, nitrogen and energy sources. This resulted in an overall identification of key pathways operating in the investigated species. Through screening of a large number of mutants for their phenotype during growth at different conditions it was further possible to map the physical locations of specific mutations in the chromosome(s) using classical genetics and this led to the identification of key genes involved in determining a given phenotype. There are many prominent examples of how this approach resulted in discoveries and identification of key regulatory proteins, and the studies of Arst and co-workers on identification of the regulatory proteins CreA and AreA behind carbon and nitrogen catabolite repression in Aspergillus nidulans form one example [1,2]. With the introduction of genetic engineering, it became possible to study many macroscopic features at the molecular level, resulting in fundamental insight into many processes in different microbial cells. This is well illustrated in the field of glucose repression in Saccharomyces cerevisiae, where different pathways have been shown to exist and many of the components of the different pathways conveying repression have been identified [3,4]. Even though the aim typically is to study the role of specific proteins or pathways in overall cell function, many of these studies represented a move towards reductionism compared with the classical macroscopic studies, as the focus was on the role of specific proteins or pathways and only slowly a global picture appeared. With the advances in the field of genomics, there came a complete

* Corresponding author. Tel.: +45 (45) 25 26 96; Fax: +45 (45) 88 41 48.
E-mail address: jens.nielsen@biocentrum.dtu.dk (J. Nielsen).

1 This paper is dedicated to the memory of Jay Bailey, who contributed significantly to our research field. His ideas and visions have been stimulating for the authors, and they will continue to guide us in the future.
revolves a cyclic operation with a close integration between analysis of the cellular function and genetic engineering [10]. Particularly in connection with metabolite production, where it is desirable to redirect the fluxes through the central metabolic pathways, analysis of cellular function plays a central role in metabolic engineering, and this has resulted in an expanded role for microbial physiology.

Clearly microbial physiology is an important research field, not only in fundamental research on microbial species, but also in all applied aspects of microbiology, i.e., industrial microbiology, environmental microbiology, and medical microbiology. In this paper we discuss the role of microbial physiology in metabolic engineering and functional genomics, and we predict that microbial physiology is moving towards systems biology. As the terms metabolic engineering, functional genomics, microbial physiology and systems biology are extensively used — and often in different contexts — we give our definitions of these terms in Table 1.

### 2. Regulation of metabolism in living cells

The first success stories in metabolic engineering involved introduction of a single heterologous gene or disruption of a single gene, but more advanced objectives can only be reached through coordinated expression (and/or disruption) of several genes [11]. A fundamental problem in attaining more complex metabolic engineering goals is the existence of different levels of control in living cells (Fig. 1): (a) transcriptional control; (b) control of mRNA degradation; (c) translational control; (d) protein activation/inactivation; and (e) allosteric regulation of enzymes. There is a certain degree of hierarchy in the control of cell function, i.e., the genes have an overall control as they define all possible phenotypes, the proteins have a control of the actual phenotype, and clearly signal transduction pathways are placed relatively high in the hierarchy. There are, however, many feedback loops in the hierarchical control structure, and it is therefore difficult...

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microbial physiology</td>
<td>Qualitative and quantitative analysis of the functions of microorganisms and their parts. Typically the overall function is studied through analysis of the cellular response to different environmental conditions. This includes the response of cell—cell interactions (crosstalk between cells), and this is often studied in the context of cell cultures.</td>
</tr>
<tr>
<td>Metabolic engineering</td>
<td>The use of directed genetic modification to improve the properties of a given cell, e.g., improved yield or productivity, expanded substrate range, and production of novel products.</td>
</tr>
<tr>
<td>Functional genomics</td>
<td>The assignment of function to open reading frames (ORFs). This includes assignment of function to ORFs that have been identified but have no known function as well as assignment of additional functions to ORFs with already assigned functions.</td>
</tr>
<tr>
<td>Systems biology</td>
<td>Description of overall cell function through quantitative description of the interaction between all the individual components in the system (the cell), e.g., gene transcription, translation, protein-protein interaction, enzyme catalysis of biochemical reactions, and receptor-metabolite interaction. With a detailed description of the individual molecular events it is also possible to consider cell-cell interactions, and hereby whole cultures can be quantitatively described.</td>
</tr>
</tbody>
</table>
Fig. 1. Hierarchical control and the ‘omes’ of systems biology represent the central dogma of biology: The genes in the genome are transcribed to mRNAs (the transcriptome) that are further translated to proteins (the proteome). The proteins are responsible for determining cellular function, either through protein–protein interactions (the interactome) or through their action as enzymes that catalyze the inter-conversion of small molecules, often called metabolites (the metabolome), best represented in the metabolism of sugars to gain free energy. The metabolites and enzymes interact in a complex metabolic network that ensures that the cell is supplied with the necessary free energy and building blocks for cell growth or maintenance of cellular activity. The activity of the different branches of the metabolic network can be represented by the flows of carbon (and other elementary compounds), often referred to as metabolic fluxes (the fluxome).

3. The impact of genomics

Presently about 80 microbial genomes have been completely sequenced and a large number of microbial sequencing projects are in the works. Furthermore, besides the human genome several genomes of other higher eukaryotes have been fully sequenced and many genomic sequencing programs involving higher eukaryotes (several plants, mouse, pig, etc.) are in progress. Of all the genomes sequenced typically only about 50–70% of the sequenced genes/open reading frames (ORFs) have a known function and therefore much research is done in the field of functional genomics. This has led to the development of many powerful analytical methods for analysis of cellular functions. These tools include:

- DNA arrays, where the expression of all genes in the genome may be measured through analysis of the mRNA pool. Different techniques have been applied for genome-wide transcription analysis, but the basic principles are the same – namely that the mRNA levels are quantified by nucleotide–nucleotide hybridization.
- Proteome analysis (or proteomics), where all the proteins within the cell are analyzed using 2D electrophoresis. Through analysis of the individual spots on the gels the proteins can be analyzed in further detail using mass spectrometry (typically MALDI-TOF MS), e.g. with respect to the degree of phosphorylation or other post-translational modifications.
- Protein–protein interactions, where the interaction between different proteins is analyzed using the yeast two-hybrid system, techniques based on fluorescence resonance energy transfer (FRET), etc. This has even been scaled-up by spotting 6000 yeast transformants on an array, and hereby the possible interaction between 192 S. cerevisiae proteins with all the other possible S. cerevisiae proteins could be investigated (the interactome) [14]. This approach is particularly important for unraveling signal transduction pathways that are believed to play a role in connection with overall regulation of cell function.
- Protein–DNA interactions, where the interaction of proteins with genomic DNA is characterized using DNA arrays. Using a recently developed methodology for the analysis of genome-wide protein–DNA interactions 200 new putative transcription factors were identified in S. cerevisiae [15].
- Metabolome analysis, where a large number of intra-
cellular metabolite concentrations are quantified, typically using GC-MS or LC-MS. Using GC-MS it is possible to measure 150 intracellular metabolites [16], and through the use of this kind of data it is possible to reveal the phenotype of otherwise silent mutations [17].

- Metabolic flux analysis, where the flows of carbon through the different metabolic routes within the cell are quantified. The fluxes can be quantified using simple metabolite balancing, but a more robust estimation of the fluxes is obtained through the use of 13C-labeled substrates followed by analysis of the labeling patterns of intracellular metabolites [18-20].

A common feature of these tools is that they give information about cellular processes at the global level, i.e. the complete mRNA pool is measured or a large fraction of the total protein pool is measured. This information supplies information of all the molecular interactions in the cell, but obviously it is difficult to reconstruct these interactions from the raw data. Advanced computational techniques therefore play an important role in connection with this type of analysis, and bioinformatics is consequently positioned at the focal point of functional genomics. The problem of reconstructing signal transduction pathways and metabolic functions from these global data can be seen as an inverse problem: the many different types of interactions overlay each other, and it is therefore difficult to reconstruct pathways directly from the raw data. To solve inverse problems mathematical models are very useful as they allow for analysis of interactions between the different components in the system at the quantitative level, and through comparison of model simulations and experimental data it is possible to evaluate the performance of the model (or the proposed hypothesis) (Fig. 2). With the appearance of large datasets the use of mathematical models to unravel the role of all the different components of the system will therefore play an increasing role in microbial physiology. As this approach is based on analysis of the complete system the term systems biology has been coined. In the phase of building models of complete pathways or overall cell function, it is very important to analyze at different levels in the hierarchical control structure as illustrated in a recent analysis of the GAL-regulon of S. cerevisiae, where transcriptome analysis of a large number of mutant strains was combined with proteome analysis of the reference strain grown at different conditions [21]. Building complete mathematical models that describe all cellular processes as well as the overall cell function is clearly an ambitious task that can only be achieved through increased insight into biological phenomena and more powerful computers. Despite this there

![Fig. 2. The approach to quantitative study of the interactions between the different molecular events in living cells applied in systems biology. Based on empirical data and knowledge of cellular function a mathematical model is proposed. The model is used to simulate the overall cell function, and model simulations are compared with experimental data. If there is a good fit between experimental data and model simulations the model is likely to be a good representation of the system, and the pathway can be considered reconstructed or the system properly described. In case there is a poor fit the model needs to be revised, and often the discrepancy between model simulations and the experimental data will point to where the model needs to be revised.](image)
has already been some progress on setting up detailed models of more simple systems [6].

Besides confirming an overall hypothesis, models reconstructed from empirical information like genome data, transcriptional data, proteome data and metabolite profiles, can play a very important role in analyzing the interplay between the system components. Based on genomic information, flux balance models have been set up for Helicobacter pylori, Haemophilus influenzae and E. coli [22–24]. The E. coli model, which is the most complex of the three models, consists of 720 reactions and 436 metabolites, and using the model structural information could be obtained, e.g. which are the most currently used metabolites and what is the overall yield of synthesis of different metabolites from glucose. Furthermore, the role of specific genes could be analyzed in silico, e.g. which genes are essential for biomass synthesis [23]. Recently, a model was reconstructed from genomic information on S. cerevisiae. This model is the first genome-scale model for a eukaryotic cell and it comprises more than 1200 reactions and more than 700 metabolites [25,26]. Using this model, it was possible to gain insight into complex mechanisms like respiration, and the role of compartmentation in biomass synthesis could be evaluated. These models only contain structural information about the metabolism, but through comparison of calculated fluxes and transcript profiles new regulatory structures may be identified, which may lead to incorporation of regulatory features in these models, i.e. the model is revised according to the scheme in Fig. 2.

4. Comparison of species

In classical microbial physiology comparison of species has frequently been used in order to draw some general conclusions on overall mechanisms. Thus, lessons concerning the Crabtree effect have been derived from comparison of S. cerevisiae and the Crabtree-negative yeast Kluyveromyces lactis [27]. Furthermore, in fungal physiology there are lessons from yeast physiology, which due to the ease of genetic manipulation of S. cerevisiae has progressed more rapidly in later years than e.g. our understanding of the physiology of A. nidulans. Even though there are several homologous regulatory proteins in S. cerevisiae and A. nidulans, it is not always simple to transfer information directly from one organism to the other. This can be illustrated by the mechanisms of glucose repression in these two organisms. Glucose repression has been studied extensively in S. cerevisiae [3], and there are several pathways conveying information on the presence of a high glucose concentration to DNA-binding proteins that effect gene transcription. Fig. 3 depicts two of these pathways. One pathway involves the Snf1 protein kinase complex and the DNA-binding protein Mig1, and probably hexokinase 2 (Hxk2) is involved in conveying the signal from glucose to the Snf1 complex. Mig1 has been demonstrated to repress the expression of enzymes involved in degradation of alternative carbon sources like sucrose (SUC-genes), maltose (MAL-genes) and galactose (GAL-genes), but Mig1 has also been shown to have binding sites to genes encoding enzymes and proteins playing a role throughout the central carbon metabolism [28]. However, disruption of MIG1 only results in relieved expression of SUC-, MAL- and GAL-genes, whereas there is no effect on the fluxes through the central carbon metabolism [18]. The other pathway is believed to act upon sensing of glucose in the extracellular medium by the glucose sensors Rgt2 and Snf3, and then a signal is forwarded to the transcriptional regulator Rgt1, probably via Grr1 [4]. Rgt1 is believed to primarily play a role in regulation of the hexose transporters (HXT-genes). As Mig1 is also influencing the regulation of some of the HXT-genes [4], there is clearly interaction between the pathways, and furthermore regulation of the hexose transporters may affect the glycolytic flux and hereby via Hxk2 influence the other pathway. Even the simplified illustration of glucose repression in Fig. 3 therefore illustrates how there may be indirect effects when the different signal transduction pathways are analyzed, and it is therefore necessary as discussed earlier to analyze at different levels in the complex control structure in order to reconstruct the details of the pathways.

In A. nidulans much less is known about glucose repression, but there is a homolog of Mig1 called CreA [29,30]. CreA has been shown to be involved in carbon catabolite repression on many different genes, particularly genes involved in metabolism of less favorable carbon sources like starch and ethanol [31]. Despite the similarities in function between Mig1 and CreA – both are involved in carbon catabolite repression of genes encoding enzymes involved in metabolism of less favorable carbon sources – there are some large differences. Whereas Mig1 is certainly important in S. cerevisiae, there are many redundant pathways (there is even a redundant protein Mig2 that also represses the SUC- and MAL-genes), and disruption of MIG1 has no serious effects on the overall cell function. On the other hand disruption of CRE4 in A. nidulans has severe growth effects. The specific growth rate is reduced and the morphology is significantly altered [31,32]. The more signifi-
cant role of CreA in *A. nidulans* as compared to that of Mig1 in *S. cerevisiae*, is likely to be due to the different environments the two organisms are exposed to – *S. cerevisiae* needs redundant pathways for glucose repression as it is naturally exposed to very high glucose concentrations, whereas *A. nidulans* is only rarely exposed to high glucose concentrations and can therefore rely on a more simple system.

With the sequencing of a large number of microbial genomes it has become possible to perform comparative genomics – on the analogy of what has been done in microbial physiology. Many lessons have been drawn from this, and in the medical field genes involved in pathogenesis – so-called pathogenicity islets – have been identified in pathogenic microorganisms by comparing sequenced genomes of different species or different strains (pathogens and non-pathogens) [33]. Furthermore, with the increased information on annotated genes from different genomes, it has become much easier to assign function to genes in a sequenced genome, as compared with the situation when the genomic sequence of *H. influenzae* was published as the first completely sequenced genome. This allows for relatively fast reconstruction of metabolic maps for sequenced microorganisms, something that is of substantial value in connection with sequencing of microorganisms used for industrial-scale metabolite production. The example on glucose repression does, however, illustrate that even though alignment of genes may reveal something about their function, e.g. that both CreA and Mig1 are involved in carbon catabolite repression, there still needs to be performed a detailed physiological characterization in order to identify the proper function of the specific genes.

5. The importance of quantitative studies

An important element of microbial physiology is quantitative analysis. In many cases qualitative information is sufficient to gain information on the role of a specific gene or on the capacity of the microorganism of interest to grow on a given substrate or not. However, quantitative analysis is a requirement for setting up detailed models in systems biology and hereby to gain insight into the details of specific mechanisms prevailing in the cell and their relative importance. In the 1960’s Tempest and others introduced chemostat cultures as a very powerful tool to perform quantitative analysis [34]. In chemostat cultures it is possible to study the growth under nutrient limitations and the specific growth rate of the cells can be precisely controlled. This has been demonstrated in many different studies both in the field of physiology and of metabolic engineering [35–40]. In recent years the use of chemostat cultures to analyze the role of specific mutations has been demonstrated (see e.g. [41,42]). Clearly the analysis of different mutants under tightly defined cultivation conditions is going to have a significant impact on functional genomics in the future, particularly as it may allow a far more detailed evaluation of e.g. gene transcription using DNA arrays, as compared with the results obtained under poorly defined growth conditions in complex media, often applied in shake-flask cultures.

6. Conclusion

Based on the above discussion we conclude that microbial physiology – even though it is a classical discipline – still plays a very important role in both application-driven research (metabolic engineering) and in fundamental biological research (functional genomics), as in many cases it is required to consider the effect of specific genetic modifications at the macroscopic level. In fact, both in metabolic engineering and functional genomics there is a move towards analyzing the complete system, and here the approach of systems biology offers great promise. Returning to our introductory definition of microbial physiology it is, however, also clear that the aim of systems biology – to describe at the quantitative level the interaction between all the components in the system – is very similar to that of microbial physiology.

Acknowledgements

We would like to acknowledge the contribution of our graduate students and research colleagues at Center for Process Biotechnology. The research environment at the center acts as continuous inspiration for our work. Particularly we would like to acknowledge the following who made comments to this paper: John Villadsen, Jochen Förster, Gerald Hofmann, Christoffer Bro, Mikkel Nordkvist and Raghevendran Vijayendran.

References


