TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Jenni Seppälä

Application of Computational Methods for Fermentative Hydrogen Production

Julkaisu 1191 • Publication 1191

Tampere 2014

Tampereen teknillinen yliopisto. Julkaisu 1191
Tampere University of Technology. Publication 1191

Jenni Seppälä

Application of Computational Methods for Fermentative Hydrogen Production

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 21st of February 2014, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2014

Custodian Prof. Olli Yli-Harja
Tampere University of Technology
Tampere, Finland

Reviewers Prof. Mauno Vihinen
Lund University
Lund, Sweden

Distinguished Prof. Chieh-Chen Huang
National Chung Hsing University
Taichung, Taiwan

Opponent Prof. Peter Neubauer
Technische Universität Berlin
Berlin, Germany

ISBN 978-952-15-3227-6 (printed)
ISBN 978-952-15-3238-2 (PDF)
ISSN 1459-2045

Abstract

Energy and environment are inseparable, since the production and use of energy always affects the environment. Current energy production relies on nonrenewable energy sources such as oil, coal and natural gas. However, the continuous production of energy from limited resources is not sustainable. This creates an urgent need to develop new methods for the production of energy from renewable sources. One possible solution is fermentative hydrogen (H_2) production. H_2 is seen as a future energy carrier. Fermentative H_2 production has many environmental advantages such as ability to use wastes as the source of energy and possibility to apply ambient temperature and pressure. Drawbacks are rather low yields and slow H_2 production rates. In order to overcome these issues vast amount of research has been conducted.

Under anaerobic conditions, various anaerobic and facultatively anaerobic bacteria utilize organic compounds by fermentation and excrete H_2 as a byproduct. In the nature, bacteria exist as mixed cultures. With appropriate pretreatments and culture conditions, H_2 producing bacteria can be enriched. Microscopy can be used for visual examination of bacterial communities, which can reveal their diversity and dominant bacterial species. Additionally wide range of fluorescent staining methods can be employed in the microscopic analysis of bacterial groups. The manual analysis of the microscopy images is user dependent and laborious. Moreover, the visual quantification of fluorescence intensities and morphological features is impossible. Therefore, automated image analysis methods were developed, e.g., for monitoring culture compositions in the H_2 producing bioreactors.

The highest H_2 production rates have been achieved with undefined mixed cultures, where the role of each bacterium to H_2 production is not exactly known. In this work, the properties of *Escherichia coli* and *Clostridium butyricum* that often coexists in mixed bacterial cultures are described. Additionally the effect of coculture of *E. coli* and *C. butyricum* was investigated and found to enhance the utilization of the given substrate. Moreover, the effects of growth conditions and possibilities of genetic modification to H_2 production by *E. coli* and *C. butyricum* are presented.

The biological approach to the design of experiments often relies on intuition. However, with computational methods higher understanding over fermentative H_2 production can be achieved. Computational methods in this work mostly focus on the modeling of bacterial metabolism and some emphasis is also given to the systematic design of experiments. Metabolic models are interaction based presentations of reactions occurring within metabolic pathways, in which the knowledge of molecules and enzymes taking part to reactions is combined. The largest metabolic models are based on the complete genome of bacteria. Metabolic models can be used to help in designing mutations and cultivation

conditions to enhance bioprocesses. Various approaches, such as flux balance analysis, can be used to simulate and analyze metabolic models. Here, the existing genome-scale metabolic model is utilized with flux balance analysis for analysis and enhancement of fermentative H_2 production.

Increasing amount of knowledge and the need to make the processes as efficient as possible has made the utilization of computational tools inevitable. Therefore, cooperation between experts with biological and computational skills is encouraged. Commonly, the aid of a computational expert is requested when data mining from an overwhelming amount of existing measurements is needed. Actually, the cooperation should start from experimental design to gain most information over the system by applying statistical design-of-experiment methods. This thesis gives an overview of computational methods applied to fermentative H_2 production and describes the use of genome-scale metabolic models to experimental design, analysis and modeling.

Preface

This thesis work has been conducted as a cooperation between Department of Signal Processing and Department of Chemistry and Bioengineering at Tampere University of Technology. The financial support of Tampere University of Technology Graduate School, Tampere Graduate School in Information Science and Engineering (TISE) and Academy of Finland is gratefully acknowledged.

The work was supervised by Prof. Matti Karp, Prof. Olli Yli-Harja and Prof. Jaakko Puhakka to whom I wish to express my gratitude for the faith and support throughout the years. The guidance and help of Adjunct Prof. Ville Santala and Dr. Tommi Aho have been priceless and is deeply appreciated. I am grateful for my co-authors Antti Larjo, Dr. Pekka Ruusuvuori and Dr. Jyrki Selinummi for the pleasant moments over work and coffee. Prof. Mauno Vihinen and Distingushed Professor Chieh-Chen Huang are acknowledged for pre-examination of the thesis and valuable comments.

The work with hydrogen was enabled by a group of welcoming researchers in the HydrogeneE group, to whom I own my greatest gratitude for the guidance and aid in experimental work and mostly for their friendship; Thank you Dr. Annukka Mäkinen, Dr. Katariina Tolvanen, Dr. Nikhil Pratap Pachhandara, Dr. Perttu Koskinen and Dr. Anna Kaksonen. I am also grateful for co-workers Dr. Reija Autio, Dr. Katri Holm, Dr. Anniina Kivistö, Joanna Alanko and Dr. Kaisa-Leena Aho for their friendship and support.

I have been privileged to have wonderful friends within folk music group Väkkärä, who have shared all my joys and sorrows. Thank you for everything. Throughout the years, the endless support of family has been the greatest source of encouragements. Thank you for my parents for the help and unfaltering support in all the areas of life. Warmest thank to my husband Esa for his endless love and encouragement and for lovely home and family we share. Thank you for my children Oskari, Noora and Valtteri for teaching me what really matters in life.

Kangasala, January 2014

Jenni Seppälä

"One important idea is that science is a means whereby learning is achieved, not by mere theoretical speculation on the one hand, nor by the undirected accumulation of practical facts on the other, but rather by a motivated iteration between theory and practice." (George E. P. Box)

Abbreviations

^{13}CMFA	Carbon-13 metabolic flux analysis
ABCP	Algorithm for blocking competing pathways
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CCC	Central composite circumscribed
CCD	Central composite design
CCF	Central composite face-centered
CCI	Central Composite Inscribed
CoA	Coenzyme A
DAPI	4’,6’-diamidino-2-phenylindole
EM	Elementary mode
EMA	Elementary mode analysis
EP	Extreme pathway
FBA	Flux balance analysis
FDH_H	Formate dehydrogenase-H
FDH_N	Formate dehydrogenase-N
FDH_O	Formate dehydrogenase-O
Fd_{ox}	Oxidized ferredoxin
Fd_{red}	Reduced ferredoxin
[FeFe]-	Iron iron
FFD	Full factorial design
FHL	Formate hydrogenlyase
GEM	Genome-scale model
hyd	hydrogenase
MFA	Metabolic flux analysis
NAD^+	Nicotinamide adenine dinucleotide (oxidized form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADP	Nicotinamide adenine dinucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
NFOR	NADH:ferredoxin oxidoreductase
OAA	Oxaloacetate
OD_{600}	Optical density at a wavelength of 600 nm
PFL	Pyruvate formate-lyase
PFOR	pyruvate:ferredoxin oxidoreductase
pH_2	Partial pressure of hydrogen
Pi	Orthophosphate
VFA	Volatile fatty acid

Contents

List of publications

The contents of this thesis provide an introduction to the following publications. The publications are referred as Publication I, Publication II and so on in the text.

I Selinummi, J., Seppälä, J., Yli-Harja, O and Puhakka, J.A. (2005) Software for quantification of labeled bacteria from digital microscope images by automated image analysis. *BioTechniques*, vol. 39, no. 6, pp. 859-862.

II Ruusuvuori, P., Seppälä, J., Erkkilä, T., Lehmussola, A., Puhakka, J. A. and Yli-Harja, O. (2008) Efficient automated method for image-based classification of microbial cells. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008)*, Tampa, Florida, USA, December 7-11, 2008.

III Seppälä, J.J., Puhakka, J.A., Yli-Harja, O., Karp, M.T. and Santala, V. (2011) Fermentative hydrogen production by *Clostridium butyricum* and *Escherichia coli* in pure and cocultures. *International Journal of Hydrogen Energy*, vol. 36, pp. 10701-10708.

IV Seppälä, J.J.[*], Larjo, A.[*], Aho, T., Yli-Harja, O., Karp, M.T. and Santala, V. (2013) Prospecting hydrogen production of *Escherichia coli* by metabolic network modeling. *International Journal of Hydrogen Energy*, vol. 38, pp. 11780-11789.

V Seppälä, J.J., Larjo, A., Aho, T., Kivistö, A., Karp, M.T. and Santala, V. (2013) Modification of the *Escherichia coli* metabolic model iAF1260 based on anaerobic experiments. In *Proceedings of the 10th TICSP Workshop on Computational Systems Biology (WCSB 2013)*, Tampere, Finland, June 10-12, 2013, pp. 79-85.

* Authors have equally contributed to the publication

The author of this thesis contributed to the publications as follows:

In Publication I, the author took part in the software design and result verification and wrote the parts related to biology. Publication I has also been included to the PhD thesis of Dr. Jyrki Selinummi.

In Publication II, the author did the microcopy imaging and wrote the parts related to biology. Publication II has also been included to the PhD thesis of Dr. Pekka Ruusuvuori.

In Publication III, the author designed the experiment, was mainly responsible in conducting the laboratory experiments and mainly wrote the manuscript.

In Publication IV, the author designed the experiments together with A. Larjo and V. Santala. The author conducted calculations applying ABCP and did all experimental preparations. The end measurements were done together with V. Santala. The author analyzed all the results and wrote the biology related parts of the manuscript.

In Publication V, the author designed the study, did the calculations and wrote the manuscript.

Computational work was done under supervision of Professor Olli Yli-Harja and the experimental work was done under supervision of Professor Matti Karp, Professor Jaakko Puhakka and Adjunct Professor Ville Santala.

1 Introduction

1.1 Background and motivation

Since the ancient times, fermentation has been used for food preservation. In the process microbes use the food to be preserved as a substrate for growth and change its composition. For example, milk can be fermented to yogurt. Nowadays, microorganisms are utilized for various tasks, such as bioleaching (Kaksonen et al. 2011), bioremediation (Singh et al. 2008) and as cell factories to manufacture desired products (Porro et al. 2010). Possibilities to use microbes for the production of energy carriers are widely studied (Peralta-Yahya et al. 2012, Quintana et al. 2011). Current energy production mostly relies on nonrenewable energy sources such as oil, coal and natural gas. The continuous production of energy from limited sources is not sustainable. This creates an urgent need to find means for the sustainable production of energy from renewable sources. One possible solution is fermentative hydrogen (H_2) production (Chong, et al. 2009b, Das et al. 2008).

Under anaerobic conditions, various anaerobic and facultatively anaerobic bacteria ferment organic compounds and excrete H_2 as a byproduct. In the nature, bacteria exist as mixed cultures. With appropriate pretreatments and culture conditions, H_2 producing bacterial species can be enriched. Microscopy can be used for visual examination of bacterial communities, which can reveal their diversity and dominant bacterial species. Computational tools can be used for the automatic analysis of experimental results, since the manual analysis of bacterial samples is laborious and user dependent. Therefore, an easy to use image analysis software is developed for analysis of the bright field and fluorescent microscopy images of bacteria (Publication I). In addition, an algorithm to analyze bacterial community composition is formulated (Publication II). These automatic image analysis methods can be used, e.g., for monitoring culture composition in H_2 producing bioreactors.

To enhance H_2 production, various environmental factors and process conditions can be optimized. Those factors include, e.g., the temperature, pH, microorganism, media composition and the reactor set-up (Wang and Wan, 2009). In Publication III, the effect of coculture on the production of H_2 is studied. Along with the increased knowledge of bacterial genome, also tools for altering the bacterial metabolism have evolved. This has created new means to approach the utilization of bacteria. With the aid of metabolic engineering, the native biochemical pathways of bacteria can be altered. This enables the optimization of bacterial metabolism to gain better yields and production rates of the desired compounds, such as H_2.

Metabolic engineering provides various approaches to improve the production of desired goods. The metabolic flow can be directed to desired compounds by removing or adding genes. The gene deletions can be used to remove unwanted pathways and to prevent enzymatic inhibition. On the other hand, by the gene additions the enzymes catalyzing the desired reactions can be overexpressed and new metabolic pathways can be created. The design of bacterial mutations requires in-depth knowledge of biochemistry. Even though bacteria are relatively simple organisms, the functions of all genes are not known to any species (Orth et al. 2011). To better understand the bacterial metabolism, genome-scale reconstructions of metabolic networks of bacteria have been created (McCloskey et al. 2013, Durot et al. 2009). Most extensive genome-scale models have been created of *Escherichia coli* (Orth et al. 2011).

Genome-scale models can be used to simulate the capabilities of an organism and to predict the most probable phenotypic states. As an example in the case of H_2 production by *E. coli*, these models can be used to estimate the most probable composition of the excreted fermentation products and to show the realms of possibilities with various substrates. Additionally, the models can be applied to find answers to biological questions, such as which genes should be removed to increase H_2 production and how the bacterial phenotype changes after the deletion, as described in Publication IV. Moreover, new metabolic pathways can be added to the model to simulate the effect of gene insertions. This enables the systematic design of experiments and means of making new, model driven discoveries towards the increased production of H_2.

The availability of genome-scale models for experimental design has given the possibility to complete the systems biology workflow (Figure 1.1). Here, we have applied the workflow to fermentative H_2 production. Altogether, Publications III-V include experimental design based on the metabolic model, comparison of the experimental results to model prediction and developing the model based on experimental results.

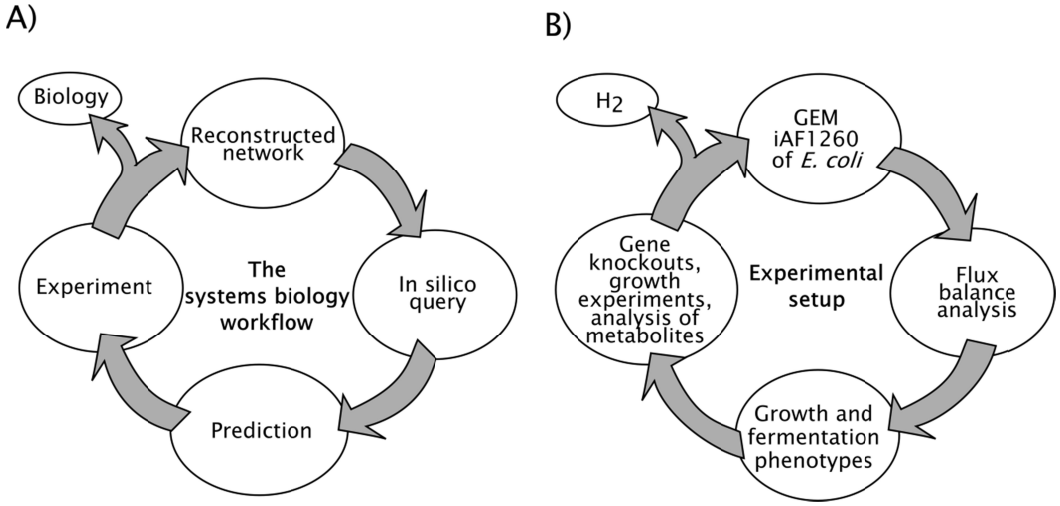


Figure 1.1. The systems biology workflow A) presents the basic workflow that can represent any systems biological experimentation based on reconstructed metabolic networks and B) represents the application of the workflow within Publications IV and V, where reconstructed genome-scale model (GEM) of *Escherichia coli* was simulated with flux balance analysis (FBA) to find mutations enhancing the fermentative H_2 production.

1.2 Objectives of the thesis

This doctoral thesis research is interdisciplinary and combines the fields of microbiology, molecular biotechnology, image processing and computational systems biology. Image processing methods are used to develop automated methods for the analysis of digital microscopy images of bacteria, which can be applied, e.g., for monitoring bacterial populations within bioreactors. Laboratory experiments are conducted to increase the understanding of bacterial behavior in cocultures and genetic regulation behind the fermentative H_2 production. Computational methods are applied to analyze metabolic networks and to computationally describe the regulation of bacterial H_2 production.

Altogether, in this thesis the goal is to apply the computational methods in order enhance H_2 production. Specifically, the aim of this thesis is to:

- Investigate the roles of bacteria in mixed H_2 producing cultures by artificial coculture of two bacterial species.
- Develop methods for automated image analysis to detect changes in bacterial cultures, e.g., during H_2 fermentation.
- Exploitation of existing genome-scale metabolic model for experimental design and analysis of fermentative H_2 production.
- Improve the genome-scale metabolic model to better describe of the mixed acid fermentation by *E. coli*.

1.3 Outline of the thesis

This thesis is organized as follows. In Chapter 2, the extracellular factors affecting to fermentative H_2 production, such as microbial population, reactor type and cultivation conditions, are described and the basic theories behind the fermentative H_2 production are presented. Additionally, Chapter 2 introduces application of image processing and statistical methods to process analysis and design.

Chapter 3 concentrates on analysis of intracellular properties of bacteria. First, metabolic pathways of H_2 production by *E. coli* and *C. butyricum* are described and the possibilities of genetic engineering are presented. Second, theoretical metabolic modeling approaches to analyze bacterial metabolism are introduced. That includes utilization of genome-scale metabolic networks with various methods. The main emphasis is on flux balance analysis. Additionally, challenges and future goals of metabolic modeling are presented. The content of the publications related to the thesis and the methods applied are briefly summarized in Chapter 4. Chapter 5 includes summary over the topics covered within the thesis.

2 Process design and analysis of fermentative hydrogen production

Two main approaches to improve fermentative H_2 production are: 1) the optimization of the extracellular factors (e.g., temperature and media composition) and 2) engineering the intracellular properties of a cell (genetic modification). In this chapter, the extracellular properties are considered as process parameters. Modification and analysis of intracellular properties is discussed in Chapter 3. The key factor affecting to the fermentation process is the microbial population that converts the substrates to H_2. Here, the sources, staining methods and automatic analysis of bacteria are presented. Additionally, cultivation methods and conditions can be varied to enhance the H_2 production. Statistical methods can be used to optimize experimental design and cultivation parameters. Thus, the main design-of-experiment methods applied for fermentative H_2 production are presented.

2.1 Fermentative H_2 production

Under anaerobic conditions, anaerobic and facultatively anaerobic bacteria utilize organic compounds by fermentation. Generalized principle of fermentation is presented in Figure 2.1. During the fermentation, most of the carbon gained from the organic substrate is excreted as fermentation products and some is used for biosynthesis (Figure 2.1.A). For comparison, in aerobic metabolism up to half of the glucose carbon is used for the biosynthesis and the rest is oxidized by the tricarboxylic acid cycle to CO_2 (Causey et al. 2004). The fermentation is anaerobic redox process without an external electron acceptor, such as oxygen. Oxidation is coupled to the reduction of a compound generated from the initial substrate. It provides energy for the adenosine triphosphate (ATP) synthesis by substrate level phosphorylation. The amount of electron carrier nicotinamide adenine dinucleotide (NAD^+) in the cell is limited, thus its generation directs the distribution of end products. Various routes for the reduction of pyruvate exist, but the aim is the same; the reduced nicotinamide adenine dinucleotide hydride (NADH) has to be returned to the oxidized form (NAD^+) to allow the energy yielding reactions of fermentation to continue. The end products of fermentation are excreted (Figure 2.1.B). Fermentation can be applied for the production of biohydrogen. In the contexts of fermentative conversion of organic substrate to H_2, term *dark fermentation* is used (Figure 2.1.C) (Müller, 2001).

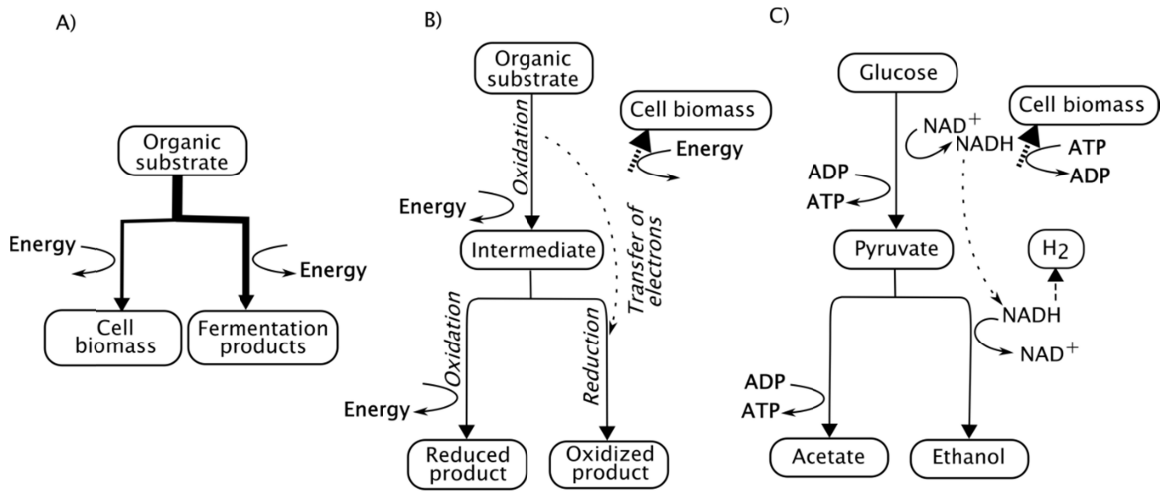


Figure 2.1. Simplified illustrations of fermentation. The distribution of the organic substrate between fermentation products and biosynthesis (A), the basic scheme for utilization of substrates by fermentation (B) and simplified presentation of dark fermentation (C) (modified from Müller, 2001).

Dark fermentation has many environmental advantages over the conventional production of H_2 from coal, oil or natural gas such as the use of wastes as a source of energy and the production of H_2 in ambient temperature and pressure. In addition to H_2, the produced gas includes, e.g., CO_2 and water vapor, which presents technical challenges for its utilization. Other drawbacks of the fermentative production of H_2 are rather low yields and slow production rates.

Even with the most efficient bacterial strains, the H_2 can not be produced by fermentation in all environmental conditions. Substrates generally used for the basic studies of dark fermentation are simple carbohydrates, such as glucose and sucrose, even though the ultimate aim is the efficient use of organic wastes and sustainably produced materials, such as lignocellulose (Chong et al. 2009b, Guo et al. 2010). Variations in the environmental conditions change the growth and metabolic phenotype of bacteria, e.g., by redirecting the metabolic fluxes. For example, pH, media composition, partial hydrogen pressure (pH_2), temperature and reactor type can be varied to optimize the H_2 production. The basic factors related to dark fermentation are presented in Figure 2.2.

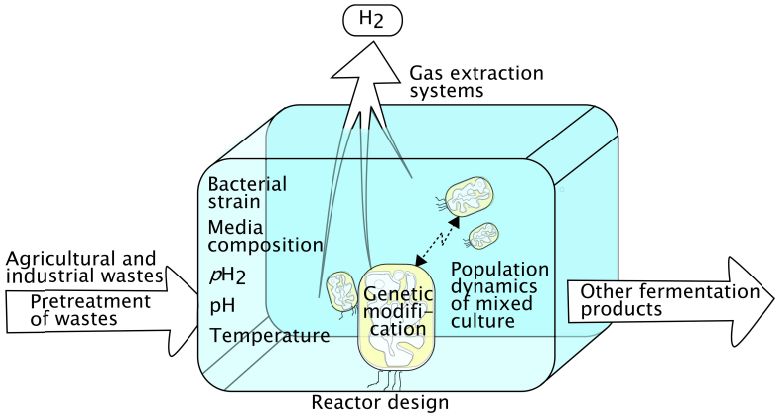


Figure 2.2. The main factors related to fermentative H_2 production.

2.2 Hydrogen producing bacteria

Various facultatively aerobic and strictly anaerobic bacteria produce H_2 by fermentation (Hu et al. 2013, Song, et al. 2012). The experimental design begins from the selection of microorganisms for the fermentative H_2 production. The hydrogen producing bacteria exist widely in the nature, thus there are many sources for the inoculum, such as soil (Selembo et al. 2009), municipal digester sludge (Koskinen et al. 2007), compost (Nissilä et al. 2011a), digested dairy manure (Wo et al. 2013) and hot springs (Koskinen, 2008). Often the microbial culture is pretreated prior the usage for H_2 production, e.g., to inhibit the growth of methanogens (Koskinen et al. 2007). Numerous studies of the H_2 fermentation with mixed cultures exist (Dong et al. 2009, Koskinen et al. 2007, Karadag and Puhakka, 2010, Nissilä, 2013).

Mixed cultures have various positive qualities such as no need for sterilization or use of aseptic techniques, capability to utilize wider range of substrates and better tolerance for oxygen. The better tolerance is due to facultative anaerobes, which utilize oxygen from the media and thus enable the growth of strict anaerobes. Even though the high yields of H_2 can be achieved by utilizing mixed cultures, maintaining the stability of the community composition, and thus the H_2 production, is challenging (Koskinen et al. 2007). Koskinen et al. (2008) achieved better stability of continuous H_2 production using completely stirred tank reactor system (CSTR) with thermophilic bacterial cultures compared to fluidized-bed bioreactor (FBR) with mesophilic cultures (Koskinen et al. 2007). That was due to less diverse bacterial culture in the thermophilic CSTR system. Additionally, the microbial composition of seed can vary and, since mixed cultures are used in open system, new H_2 consuming species may become enriched. Therefore, repetition of the experiments with mixed cultures is difficult.

Mixed cultures can be used as source of isolates for pure cultures to gain more understanding over the roles of each bacterial species in the system. In Publication III, *C. butyricum* was isolated from H_2 producing fluidized-bed reactor, where it and *E. coli* was present throughout the reactor operation (Koskinen et al. 2007). Therefore, H_2 production by *C. butyricum* and *E. coli* was examined separately and in cocultures to study their roles in the reactor (Publicaton III). The experiments revealed that the coculture utilized glucose more efficiently than the pure cultures alone (i.e. less residual glucose existed at the end of the experiment). Since *E. coli* is facultative anaerobic, it can consume the possible traces of O_2 from system, thus enabling the growth and H_2 production of strict anaerobe *C. butyricum*. Similar approach has been presented by Chang et al. (2008) with cocultures of aerobic *Bacillus thermoamylovorans* and *Clostridium beijerinckii* L9. Following, the basics of the H_2 production by of two vastly studied bacteria, *E. coli* (Ordal and Halvorson, 1939, Rosales-Colunga et al. 2012, Publication III-V) and *Clostridium butyricum* (Karube et al. 1976, Yokoi et al. 1997, Chen et al. 2005, Publication III) are presented.

2.2.1 H_2 production of *Escherichia coli* and *Clostridium butyricum*

E. coli and *C. butyricum* have divergent fermentative pathways and methods for the production of hydrogen. As Figure 2.3 illustrates, both species have similar glycolytic steps, where one glucose molecule is degraded to two pyruvate molecules. The main difference between *E. coli* and *C. butyricum* is in the oxidization of pyruvate.

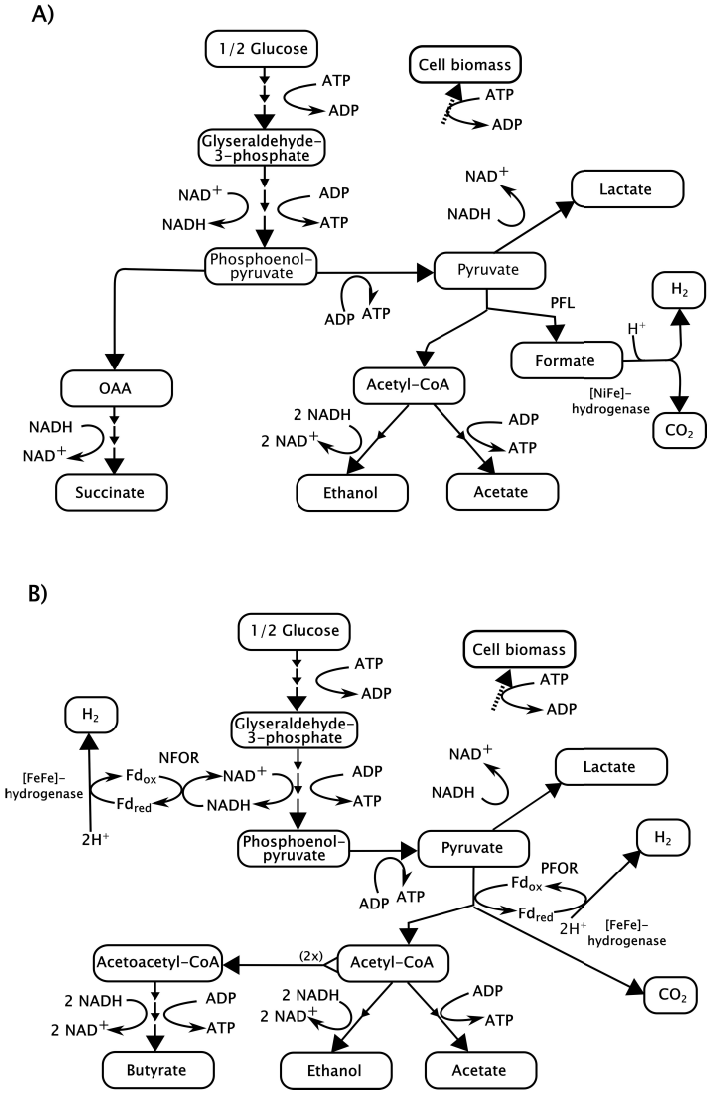


Figure 2.3. Anaerobic degradation pathways of glucose by A) *Escherichia coli* and B) *Clostridium butyricum.* A) *E. coli* uses pyruvate formate lyase (PFL) pathway and [NiFe]-hydrogenase for the H_2 production. B) *C. butyricum* uses pyruvate:ferredoxin oxidoreductase (PFOR) and NADH:ferredoxin oxidoreductase (NFOR) pathways and [FeFe]-hydrogenase for the H_2 production.

E. coli is a Gram-negative facultative aerobic bacterium. In anaerobic conditions, it applies mixed acid fermentation for the utilization of glucose (Figure 2.3 A). The excreted end-products are mainly lactate, succinate, ethanol, acetate, CO_2 and H_2. During the fermentation, pyruvate is broken down to formate and acetyl-CoA in a reaction catalyzed by pyruvate formate-lyase (PFL). Formate hydrogenlyase (FHL) complex catalysis further the division of formate to H_2 and CO_2. If the entire electron flow of *E. coli* is focused on H_2 production, the maximum yield of hydrogen is 2 mole of H_2 per a mole of glucose (Nath

and Das, 2004). In that case, only ethanol, acetate, CO_2 and H_2 are excreted. Degradation of acetyl-CoA to ethanol maintains the NAD^+ balance and degradation to acetate produces energy (ATP). In practice, maximization of H_2 production means that no electrons are directed to the biomass production and the cells cannot maintain growth. In reality, the H_2 production is a method for the bacteria to maintain redox balance by excreting excess H^+. Actually, from energy aspects the production of H_2 is unfavorable, thus normally relatively small amount of hydrogen is produced along with other end-products. The experimentally observed ratio of end-products from glucose in cultivation media is shown in Figure 2.4.A (Publication III and IV). Despite the comparably low maximum yield of H_2 from glucose, utilization of *E. coli* has some advantages; it has simple nutritional requirements, rapid growth rate, ability to grow in aerobic conditions, well established genetics and its metabolic modification is easy (Blattner et al. 1997, Maeda et al. 2008).

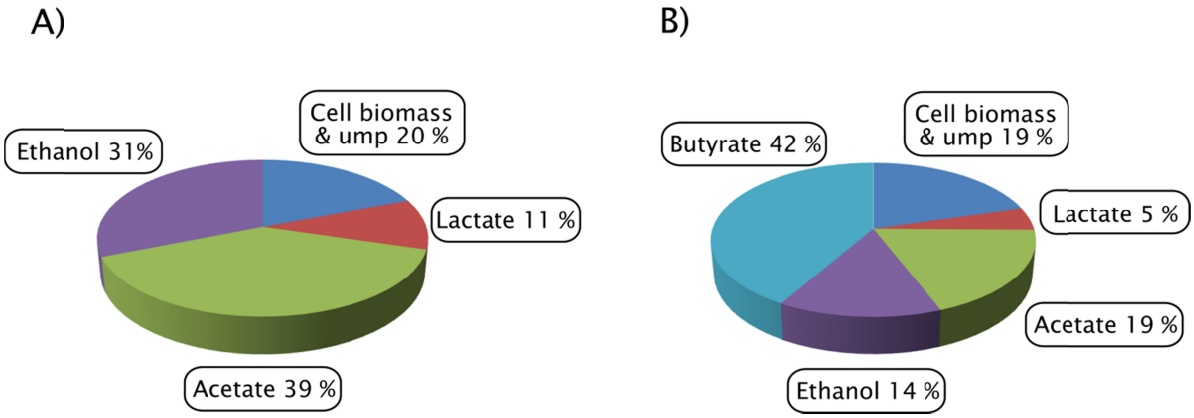


Figure 2.4. Degradation products of glucose by A) *Escherichia coli* and B) *Clostridium butyricum* (Publication III) measured from the cultivation media. The amount of glucose used for the measured metabolites is calculated as follows: n(glucose) = 0.5 n(acetate) + 0.5 n(ethanol) + 0.5 n(lactate) + n(butyrate) (Publication III). The remainder is used for biosynthesis and unmeasured products (ump).

C. butyricum is Gram-positive, spore forming and strictly anaerobic bacterium. It utilizes glucose by butyric acid fermentation. Main end-products of the fermentation are lactate, butyrate, ethanol, acetate, CO_2 and H_2 (Figure 2.4.B) (Saint-Amans et al. 2001). The main difference in H_2 metabolism of *C. butyricum* compared to *E. coli* is that *C. butyricum* directly splits the pyruvate to acetyl-CoA, CO_2 and H_2 with the aid of pyruvate-ferredoxin oxidoreductase, without formate as intermediate. Oxidation of pyruvate to acetyl-CoA requires reduction of ferredoxin ($Fd_{ox} \rightarrow Fd_{red}$). As *E. coli*, also *C. butyricum* produces NADH during glycolysis, but it can apply the reduction of ferredoxin also to convert NADH back to NAD^+, thus the oxidation of NADH is not mandatory reaction at later stages of fermentation. The reduced ferredoxin (Fd_{red}) is oxidized back to oxidized ferredoxin (Fd_{ox}) with the aid of [FeFe]-hydrogenase enzyme, and in the process electrons in the form of molecular H_2 are released. Therefore, it is theoretically possible for *C. butyricum* to have only acetate, CO_2 and H_2 as end products and to achieve the maximum H_2 production, i.e. 4 mole of H_2 per mole of glucose (Levin et al. 2004, Das and Veziroğlu, 2001). At that case, all the NADH is oxidized through ferredoxin reduction. As with *E.*

coli, this does not occur in reality. Theoretical stoichiometric analysis of *C. butyricum* metabolism suggests that the maximum yield of H_2 is 3.26 mol of H_2/mol of glucose (Chen et al. 2006). Experimental results of the end-product distribution are illustrated in Figure 2.4.B, which is derived from data presented in Publication III. Advantages of the usage of *C. butyricum* as H_2 producer is the higher maximum yield compared to *E. coli* and fast growth rate. The drawbacks are the obligatory need for anaerobic conditions and difficulties in genetic engineering. An important breakthrough in genetic engineering was resently done by Cai et al. (2011), who were the first to succeed in transferring a plasmid into *C. butyricum* by conjugation.

2.2.2 Microscopy and staining of bacterial cultures

Antonie van Leeuwenhoek was the first to observe and describe single-celled organisms using his handcrafted microscopes at 1684. Until the beginning of the 20th century, the classification of bacteria was mainly based on morphology, growth requirements and pathogenicity (Schleifer, 2008). Even though nowadays there are other means to analyze the bacterial types in a culture, microscopy is still an important tool for routine analysis. The traditional bright field microscopy offers information over morphological features, such as cell size, shape and aggregation. Examples of cell morphologies are illustrated in Figure 2.5.

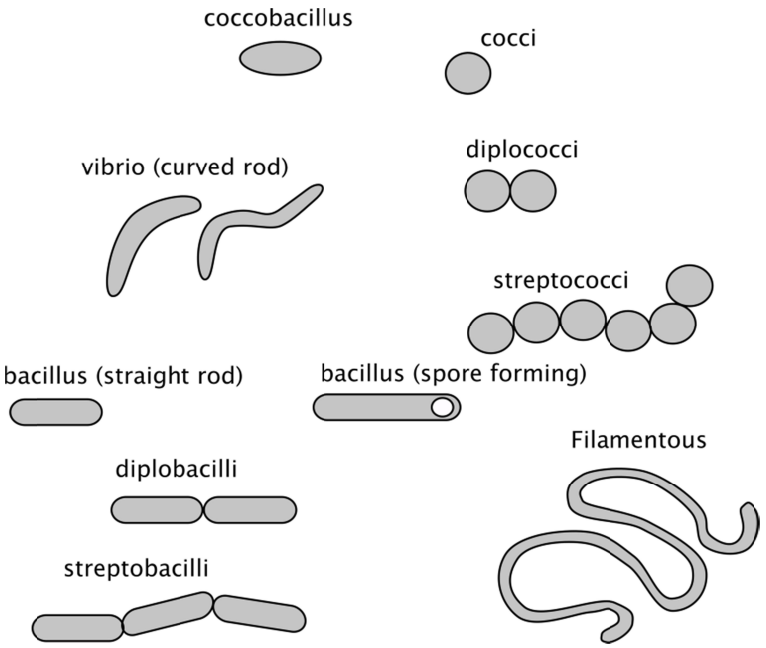


Figure 2.5. Common morphologies found in bacterial samples.

Bright field microscopy can be applied to monitor, e.g., purity of a pure culture, analyze cell morphology and detect changes in mixed bacterial cultures. Composition of bacterial culture is vital, e.g., for the hydrogen production capability of mixed culture. Koskinen et al. (2007) studied mixed bacterial community within fluidized-bed bioreactor (FBR). They

observed that the instability of H_2 production in the FBR was due to rapid changes in the microbial community structure. The fastest method to monitor such changes is the bright field microscopy. Even though it is impossible to identify bacterial species, one can make observations over morphological groups within the system and about possible aggregation. Demirel and Yenigün (2006) studied the behavior of microbial population in an anaerobic reactor based on features such as autofluorescence and cell morphology.

In addition to morphological observations of the culture, it is possible to use, e.g., epifluorescent microscopy with fluorescent markers to enable more detailed analysis of bacterial samples. In Publications I and II, 4',6'-diamidino-2-phenylindole (DAPI) -staining is used for visualization and enumeration of bacterial cells. The DAPI staining is based on its attachment to bacterial DNA, but since the bacterial chromosome is freely distributed within the cytosol it can also be used for approximation of cell morphologies. In Publication I, dual stained bacteria are analyzed, where DAPI staining (blue light) is applied for detecting all cells within the system and fluorescent in situ hybridization (FISH) (e.g. red or green fluorescent label) is used to detect certain group of bacteria based on 16S rRNA. The specificity of the FISH is dependent on probe design. Some applications of fluorescent stains are based on cell wall structure, e.g., live versus dead cells can be detected using DAPI, which penetrated all cells, together with propidium iode (PI), which penetrate only damaged cell walls of dead bacteria. The fluorescent Gram-staining is also based on the penetration differences of two dyes through the cell wall (Mason et al. 1998). In the additional analysis for Publication III (not published), fluorescent Gram-staining was used to enumerate the ratios of *E. coli* and *C. butyricum* within the coculture batch experiments.

Nowadays, microbial cells can be visualized with single-molecule sensitivity. For example, Taniguichi et al. (2010) have constructed yellow fluorescent protein (YFP) fusion library for *E. coli*. There, each of the library strain has YFP translationally fused to the C terminus of a protein in its native chromosomal position, enabling visualization of abundance and cellular localization of the studied protein. An application of similar construction is labeling of nucleoid-associated proteins to detect chromosome organization in live bacteria (Wang et al. 2011).

2.2.3 Automatic analysis of microscopy images

Despite of the application of microscopy, manual analysis of the microscopy images is user dependent and laborious and moreover the visual quantification of fluorescence intensities is impossible. Possibility to take digital images over bright field or fluorescence microscopy views enables the usage of automatic image analysis for these purposes. The basic flow of automated image analysis is presented in Figure 2.6.

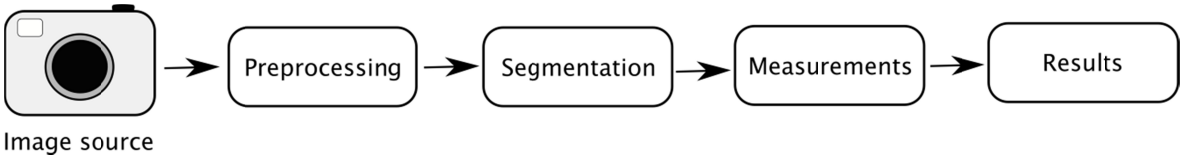


Figure 2.6. Typical stages of image processing. The first step is image acquisition with a digital camera. Then preprocessing is applied in order to minimize errors and artefacts caused by the image acquisition system. The next step is segmentation, in which the objects of interest are separated from the background. The separated objects are measured to gain the final results (Modified form Lehmussola, 2009).

In Publication I, an easy to used image analysis software, called CellC, is introduced and its analysis results are validated. CellC can be applied to enumeration of bacteria and to measure cell length, width and intensity from digital images. It can also be used for comparative analysis of images taken from dual stained samples. CellC has been used among other applications to analyze cell sizes from soil samples (Elazhairi-Ali et al. 2013), to enumerate bacterial cell for enrichment cultures of seawater (Gray et al. 2009), and to qualify spore harvest purity (Harrold et al. 2011).

Karr et al. (2012) have created a whole-cell computational model for small genome bacteria *Mycoplasma genitalium*. The goal of the model is to predict the phenotype from genotype. In this aspect, the phenotype includes, among many other features, the data over cell mass, volume and shape. Since variation caused by a single mutation to these aspects might be small, image processing can aid in enumerating differences that are not observable by naked eye.

Besides enumeration and quantification of the cell properties, image analysis can be used for further downstream analysis. One example is the division of the objects to different classes based on cell morphology. In Publication II, a framework for automated classification of microbial cells based on morphological features, such as cell size, shape and intensity, is presented. Changes in the community structure can be fast and cause drastic effect on the reactor operation. The framework could be used, e.g., for continuous monitoring of H_2 producing bioreactor with mixed culture. The automatic analysis of the population structure from microscopy images is a rapid method to observe upcoming changes and allow quick response by operator if needed. Additionally, the automatic classification could be used to analyze dynamic changes and interspecies interactions in mixed bacterial cultures or cocultures to gain more understanding over bacterial behavior.

Automatic image analysis has many difficulties that should be kept in mind while taking pictures to be analyzed or to be used for training a classifier. The most important factor of the images that are analyzed by automated method, such as CellC, is the quality of the pictures. The software can not differentiate bacterial cells from other particles if both look similar, thus taking those into the imaging field should be avoided. Additionally, samples should be dilutes to proper density before microscopy, since too crowded images are more difficult to analyze. Moreover, the cells should locate in a one focus level. Therefore,

stained cells are often attached to filters to enable analysis at one focus level and to create higher threshold between object and background. The microscope and image acquisition setting should be properly adjusted to avoid background noise. If the images are used for training a classifier, it is important to take vast amount of pictures from each example case.

2.3 Cultivation method or reactor type

The selection of cultivation method is a trade-off between many factors such as time and amount of data. Experimentation with unknown cultures or new mutants is often started with short batch experiments and sometimes the most prominent strains are then studied in continuous cultures. In Figure 2.7, some properties of different cultivation methods are presented.

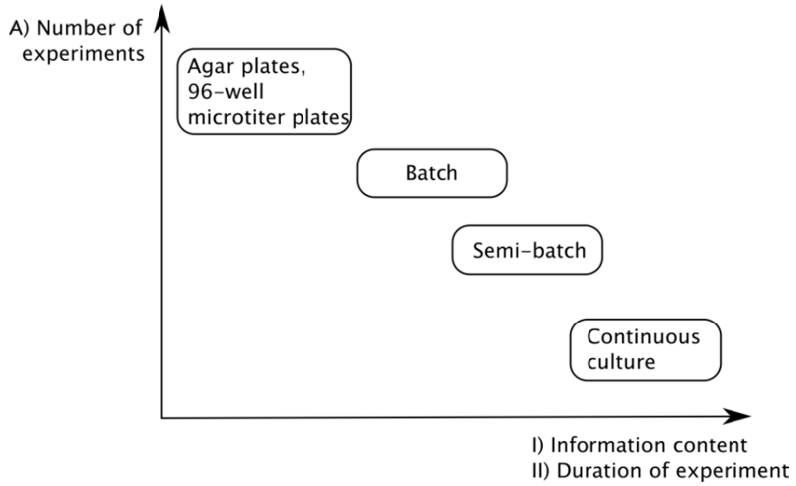


Figure 2.7. Alternative experimental setups. Continuous cocultures are time consuming experimentations, but provide large amount of information over the system. Small scale experimentation are fast and easy to repeat, but less informative.

In order to screen large amount of bacterial samples, fast methods to evaluate H_2 production capabilities are needed, e.g., to select the most prominent strains after the exposure to random mutagenesis. A method to screen the H_2 production is to use chemochromic membranes.

The basic principle of the use of chemochromic membranes is following. Bacterial cultures are grown in well plates or on agar. After the colonies are formed, first a filter paper and then chemochromic membrane is placed over the active cultures. After few minutes, the H_2 production can be seen as bluish spots on the film (Seibert et al. 1999). Chemochromic membranes has been used in protein engineering to find the mutations of FhlA that enhance the H_2 production of *E. coli* (Sanchez-Torres et al. 2009) and to screen the single deletion mutants of *E. coli* gained from KEIO collection. Drawbacks of the method are that the color change is transient and difficulties in the quantification of the results. As a solution to these problems another high-throughput screening assay has been presented by Schrader et al. (2008). The system consists of two 96-well plates used as separate layers;

bacteria are cultivated in lower layer and upper layer, having membrane bottom, includes color indicator and catalyst. The evolved H_2 diffuses up to second layer and causes visually detectable permanent color change that can be analyzed with plate reader. That enables quantitative comparison of the cumulative H_2 production within the sample set.

Despite the existing methods for semi-quantitative analysis for measuring the H_2 production, traditional random mutagenesis is not generally used to create mutants with enhanced properties for H_2 production. That is due to the challenges created by anaerobic conditions and difficulties in the true quantitative screening of H_2 production.

Quantitative screening of samples can be done using small batch experiments with only endpoint measurements, e.g., with anaerobic Hungate tubes (Publication IV). It enables the quantification of the total gas production, measurements of culture density and analysis of metabolites and residual substrates from the liquid at the end of experiment. This method can reveal good candidates for H_2 production, but without additional sampling, can not be used for the analysis of the growth or production rate. With manual sampling the amount of simultaneous experiments is restricted. Additionally, especially in small batch reactors, high partial pressure of H_2 (pH_2) inhibits the H_2 production. Prominent candidates should be subjected to further study in a larger scale reactors to gain more knowledge of the dynamics of growth and gas production.

Larger scale batch experiment system enables more frequent sampling and larger head space for the gas production. E.g., in Publication III, 2.1 liter vessels were used. In this manner, the inhibition of H_2 production by high partial pressure of hydrogen can be decreased. Larger systems can be connected, e.g., to automatic detection of pressure which enables constant monitoring of gas production. Additionally, more frequent manual sampling of gases and liquids can be done. The gas samples and the OD_{600} had to be measured immediately after sampling; therefore, frequent sampling restrict the amount of reactors run simultaneously. From constantly monitored batch system, it is possible to gain data over growth, H_2 and metabolite production and substrate consumption rates. Since large vessels enable larger sample volume without disturbing the growth, also many other parameters, such as hydrogenase activities could be measured. Since the length of the batch experiment is considerably short, it enables repetitions in order to gain more reliable data analysis. Batch experiments are often done under controlled pH, which enables the total utilization of substrates and function of the hydrogenase throughout the experiment. Also semi-batch mode can be used, where media or substrates can be added to the system in the middle of the experiment allowing the continuation of the growth.

In Publication III, batch reactor was used to study the behavior of coculture of *E. coli* and *C. butyricum*. With batch system, information can be gained over what occurs when the both species are in the exponential growth phase. Anyhow, it would be interesting to follow the population dynamics in continuous reactor system. That would enable more

detailed analyses of the effects of growth conditions to the ratio of the species and to H_2 production. Additionally, information over how or does cultures reach stable mixture after certain time could be gained. Nevertheless, setting up and maintaining such a system and keeping it free from contamination would be challenging and time consuming.

2.4 Medium composition, pH and temperature

One aim of fermentative H_2 production is the efficient utilization of waste materials. Large quantities of renewable materials are produces as residues of industrial processes, e.g., cellulosic residues from agricultural, forestry and food processing industry. Cellulose is composed of thousands of linked glucose units and those molecules are bound together by hemicellulose. The hemicellulose is composed of randomly ordered molecules such as glucose, xylose, mannose, galactose, rhamnose and arabinose which can be used as a carbon sources for fermentation (Gírio et al. 2010). Nissilä et al. (2011b) used bacterial culture isolated from cow rumen for production of H_2 from cellulose. Additionally, the H_2 production has been studied with number of other waste materials such as sewage sludge (Chen et al. 2005), cheese whey (Rosales-Colunga et al. 2010) and palm oil mill effluent (Chong et al. 2009a). From modeling point of view, all components of growth media should be known to gain systematic information of the process. Thus, media should be composed on defined substances and pH and temperature should be controlled. On the other hand, better production and growth rates are achieved when complex sources of amino acid and nutrients are added (Mathews et al. 2010). While optimizing the parameters, the goal to apply complex waste materials as a source of H_2 should be kept in mind.

Various bacterial species produce hydrogen and each of those has distinct optimal growth conditions and capabilities to utilize substrates. Here, the main focus is on *E. coli* and *C. butyricum* and their optimal H_2 production conditions. *E. coli* is capable of degrading various sugars, such as fructose, sorbitol, galactose, xylose and glucose, which increases its applicability for degradation of various carbohydrates present in industrial wastes (Ghosh and Hallenbeck 2009). For example, Publication IV presents that wild-type *E. coli* can reach nearly the same yield of H_2 from closely related substrates, galactose and glucose. Anyhow, the growth on galactose is slower and lag time is longer than on glucose, which was also reported in the experiments with cheese whey (Rosales-Colunga et al. 2010). Additionally, the metabolic fluxes of *E. coli* are differently distributed (van Rijsewijk et al. 2011) and responses to gene deletion vary based on which substrate, glucose or galactose, is used (Publication IV). Different substrates have distinct pathways to enter the central metabolism. For example, xylose enters it through pentose phosphate pathway. Therefore, metabolic engineering and experimental design should be done based on the utilized substrates. Moreover, the design of mutants and optimization of experimental conditions should always be started from the beginning when the substrate is changed.

Even though the ultimate goal of fermentative H_2 production is to utilize waste materials, most of the research with *E. coli* has been conducted and highest yields of H_2 are reported using glucose as substrate. High concentrations of glucose can inhibit the bacterial growth and H_2 production (Penfold et al. 2003). The highest H_2 production yields and rates with *E. coli* are achieved using moderate glucose concentrations (~10-18 g/l) (Penfold et al. 2003, Ghosh and Hallenbeck 2010, Yoshida et al. 2006). Similarly, H_2 yield of *C. butyricum* is inhibited by high substrate concentrations and the optimal initial glucose concentration has been reported to be around 5 g/l glucose (Cai et al. 2010, Yokoi et al. 1997). Some of the reported optimal conditions for H_2 production are listed in Table 2.1.

Table 2.1. Reported optimal condition for H_2 production by *Escherichia coli* and *Clostridium butyricum.* The range tested is given in the parenthesis.

Optimized environmental conditions for H_2 production by *E. coli*

Strain	DJT135	W3110	(W3110) SR15
H_2 yield (mol/mol glucose)	1.69	1.08	1.82
Glucose (g/l)	13.5 (4.5-22.5)	10.8	10.8
pH (controlled batch)	6.5 (4.5-8.5)	6 (5-7.5)	6
Temperature (°C)	35 (25-45)	37	37
Reference	Ghosh and Hallenbeck, 2010 and 2009	Yoshida et al. 2006	Yoshida et al. 2006

Optimized environmental conditions for H_2 production by *C. butyricum*

Strain	CWBI1009	W5	CGS5
H_2 yield (mol/mol glucose or sucrose)	1.7	1.375	2.78
Glucose (g/l)	5	4 (4-10)	-
Sucrose (g/l)			17.8 (4.3-26.4)
pH (controlled batch)	5.2 (4.7-7.3)	6.5 (6.0-7.0)	5.5 (5-6.5)
Temperature (ºC)	30	39	37
Reference	Masset et al. 2010	Cai et al. 2010	Chen et al. 2005

In addition to the carbon source also other factors of the culture media, such as nitrogen, sulfate, phosphate and trace metal concentrations, affect to the H_2 yields. With *E. coli* limitations on glucose increases the H_2 yield, whereas limitations on phosphate, sulfate and nitrogen decreases the H_2 yields (Turcot et al. 2008). Addition of yeast extract increased the total H_2 production by *C. butyricum* and shortened the lag time after incubation with an active culture (Chong et al. 2009b). In the pretests related to Publication III, various yeast extract concentrations were tested and the smallest concentration (0.3 g/l) allowing the growth of *C. butyricum* was used. Utilization of complex media (e.g., with yeast extract) causes problems with reproducibility and difficulties to determine the medium composition. Thus, especially for modeling purposes, rich defined media can be used in

laboratory-scale experimentation (Mathews et al. 2010). With *E. coli,* usage of rich defined media, including amino acids, has been found to increase the growth rate, hydrogen production rate and acetate accumulation in comparison to minimal media. Additionally, with rich media, there was no lag time after the incubation from active *E. coli* culture (Mathews et al. 2010).

Mixed acid fermentation is a type of metabolism, which facultative anaerobes use to survive under various environmental conditions. For example, based on pH and available substrates, *E. coli* can redirect the metabolite production to maintain the growth conditions favorable. With *Enterobacter aerogenes*, which uses mixed acid fermentation, the increase of pH in chemostat culture from 5.0 to 7.5 increased the ethanol and acetate production and decreased the lactate production. For *E. aerogenes,* the optimum pHs for H_2 production and growth were 5.8 and 7.0 respectively (Nakashimada et al. 2002). In addition to end product composition, also enzyme functions are affected by pH. For example, the degradation of formate is regulated by pH and hydrogenase 3 is most active at low pH (below pH 7) (Bagramyan et al. 2002). Yoshida et al. (2006) showed that between pH 5.5 and 6.0 the H_2 production yields of *E. coli* were highest. Anyhow, some H_2 was also produced at as low as pH 5.0 and at as high as pH 7.5 (Yoshida, et al. 2006).

Cultivation of *C. butyricum* in pH controlled batch at various pH levels indicated that the maximum H_2 production yields are achieved at pH 5.2-6 and the production rates are highest in slightly higher pH values (6-6.5) (Chen et al. 2005, Cai et al. 2010, Masset et al. 2010). Also with *C. butyricum,* the distribution of metabolites varies based on pH. When the pH is decrease from 7.0 to 6.0, more butyrate and less lactate and ethanol is produced (Cai et al. 2010). Additionally, the lag phase of the culture have been shown to be shorter at pH 6.5 than at pH 6 (Cai et al. 2010).

Box-Behnken experimental design coupled to response surface methodology has been used to optimize fermentation conditions (Ghosh and Hallenbeck, 2010 Rosales-Colunga et al. 2010). Ghosh and Hallenbeck (2010) applied it to find optimal combination of glucose, pH and temperature for H_2 production by metabolically engineered *E. coli* (DJT135) and found them to be 75 mM of glucose, 35ºC and pH 6.5. This experimentation was conducted at controlled pH. Anyhow, batch experimentations are often done in uncontrolled pH, thus pH decreases during the fermentation. An example of this is shown in Publication III, where the fermentation of glucose ended when the pH decreased to 4.4. The response surface methodology has also been applied for *E. coli* cultured on cheese whey (byproduct of cheese production) in batch reactor without pH control. There, the optimal initial conditions for H_2 production were pH 7.5, 37 ºC and 20 g/L cheese whey. In these conditions, yield of 1.37 mol of H_2/mol of hexose was reached (Rosales-Colunga et al. 2010). Thus, the higher initial pH increased the H_2 yield when the pH was not controlled. Additionally, the optimization methods have been used to find best H_2 production conditions from palm oil mill effluent for *C. butyricum*. Optimized conditions

were pH 6.05, 36°C and 94 g COD/l (Chong et al, 2009a). In Figure 2.8, the basic factors affecting H_2 fermentation are summarized. Following, the statistical methods for experimental design are discussed in more detail.

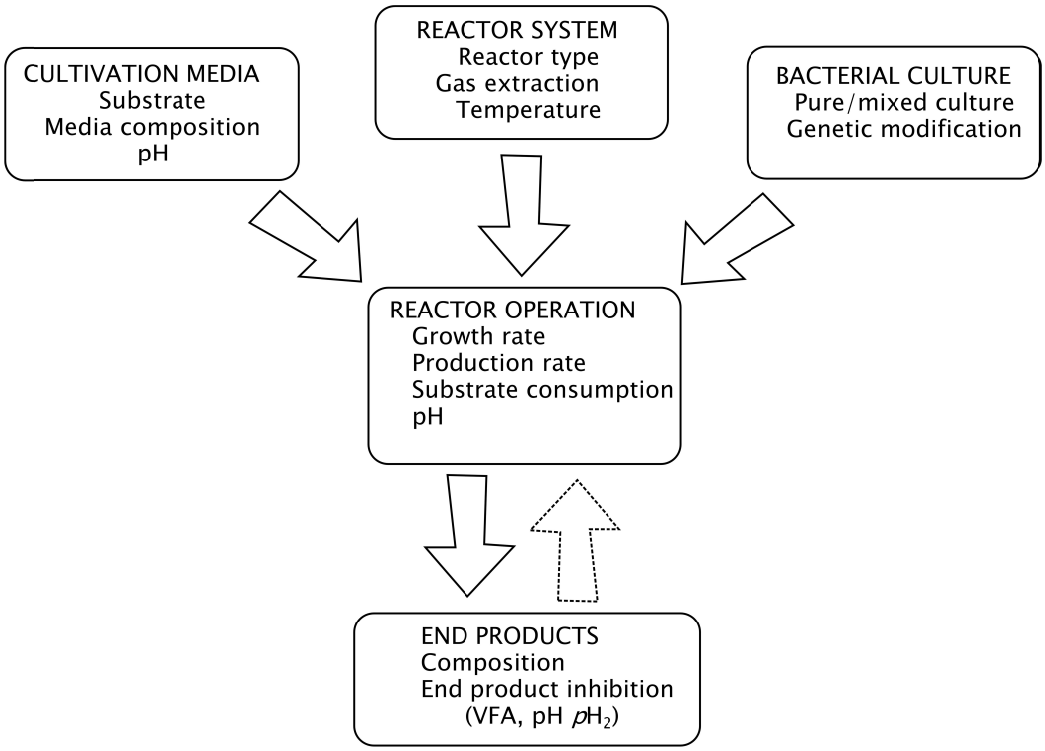


Figure 2.8. Summary of factors affecting fermentative H_2 production.

2.4.1 Statistical methods for optimization of experimental set-up

Optimization of cultivation conditions is important. It can reveal the true potential of the studied wild-type, mutant or mixed culture for the fermentative H_2 production. If the culture is only capable of producing H_2 under narrow conditions, e.g., in low glucose concentrations, its usage for real life applications is difficult. Traditionally, optimal culture conditions are sought by testing the effect of one variable (factor) at a time. This method can lead to large number of experiments, with never finding the optimal conditions since it does not take into account the interdependence of the variables. Various statistical approaches can simultaneously result in finding the optimal conditions and decreasing the number of experiments needed. Figure 2.9 presents the basic terms used with in the bioprocess optimization.

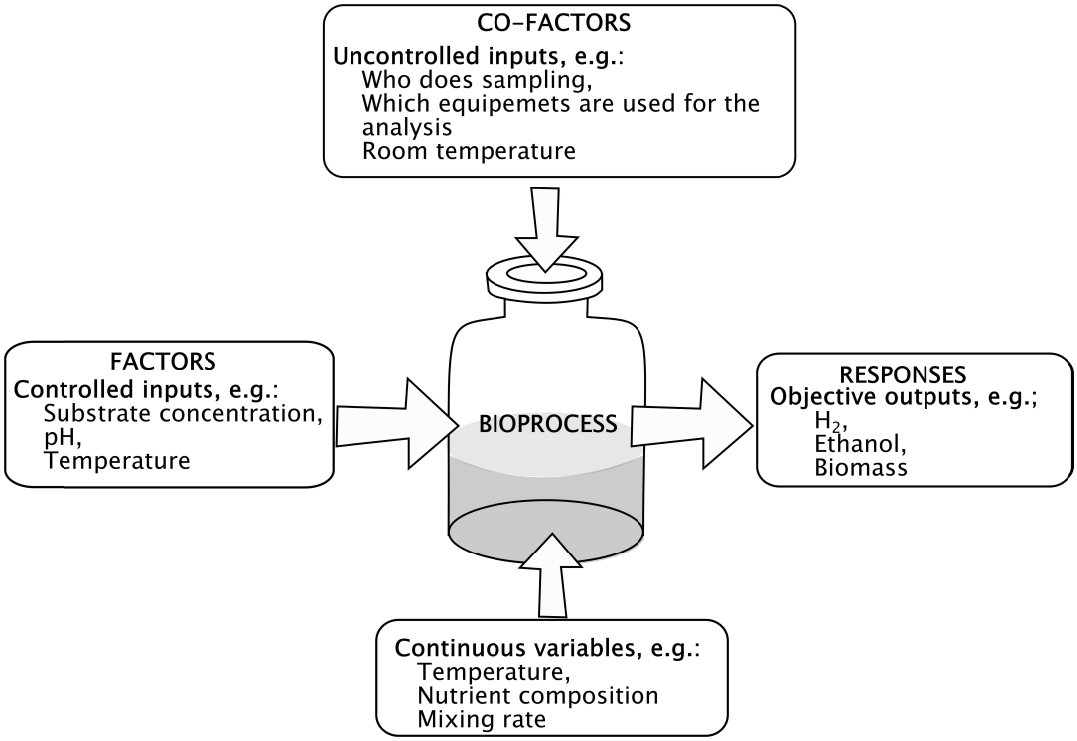


Figure 2.9. Basic terms used for statistical design of experiments. Factors: the inputs that are varied with in the optimization experiment. Response: the objective output (outputs) which production is optimized. Continuous variables: possible variables which are kept as constant as possible during the experiments. Co-factors: those variables that not thought to affect the experimental results, but actually often does.

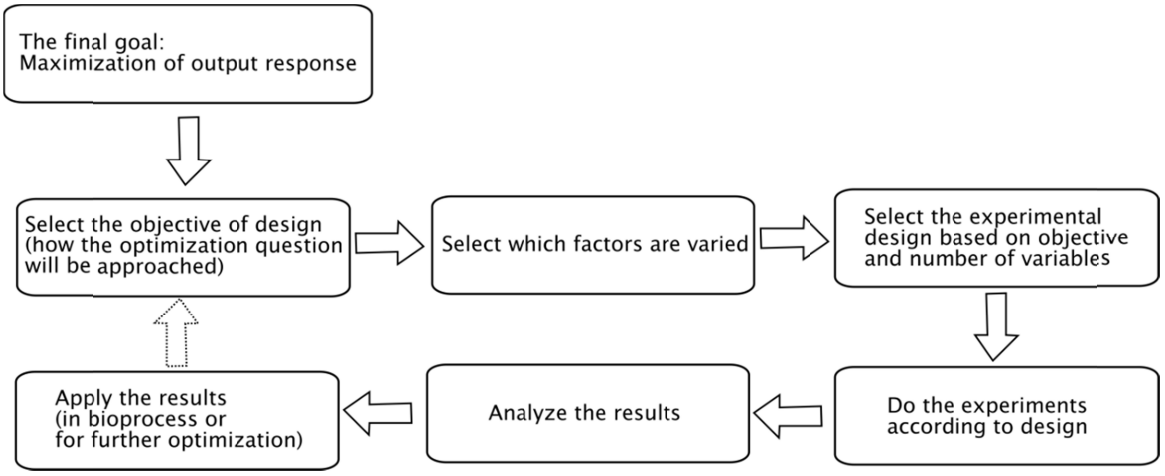


Figure 2.10. The basic scheme for optimization experiments. Often more than one cycle is needed, e.g., first the most important factors are screened and then, during second cycle, those are optimized.

The flow of optimization process is illustrated in Figure 2.10. Extensive handbooks of engineering statistics are available (Nist/Sematech, 2013). Thus, only the main methods that have previously been applied to fermentative hydrogen production are introduced. The approaches are classified based on their *objective* as follows:

1) Comparative objective: The goal is to determine whether the factor of interest is significant among the group of several factors that could be affected. For example, does addition of iron to the cultivation media have an effect on the H_2 production.
2) Screening objective: The goal is to find the most important factors among large group of factors that could be varied. For example, what are the most critical constituents within the cultivation media for efficient H_2 production.
3) Response surface objective: The goal is to find improved or optimal process settings, to troubleshoot process problems and weak points, and increase the robustness of the process. For example, it can be used to find optimal combination of pH, temperature and glucose concentration for maximization of fermentative H_2 production.
4) Mixture design objective: The goal is to find best proportion of factors within a mixture. For example, what is the best bacterial composition within artificially mixed culture to increase the H_2 production (Nist/Sematech, 2013).

The *experimental design* method is selected based on the objective. Examples of experimental design methods are given in Table 2.2.

Table 2.2. Examples of design-of-experiment methods used for experimental design with the given objective (Nist/Sematech, 2013).

Number of Factors	1) Comparative objective	2) Screening Objective	3) Response surface objective	4) Mixture design objective
1	1-factor completely randomized design	-	-	-
2-4	Randomized block design	Full or fractional factorial	Central composite or Box-Behnken	Simplex-Lattice or Simplex-Centroid design
5 or more	Randomized block design	Fractional factorial or Plackett-Burman	*Screen first to reduce the number of factors*	

Experimental design methods

The most important objectives applied in H_2 production experiments are screening and response surface objective. The experimental design methods used to achieve those objectives are presented here (Figure 2.11). The *number of levels* within each method refers to how many different values of each factorial need to be tested. The proper range is selected by the user, thus vast understanding of the system is needed to make the solution space such that the optimal conditions will be within the reasonably selected range. The selected experimental values are usually given as coded within the experimental design. Coded value x_i of the *i*th independent variable is calculated as following:

$$x_i = \frac{X_i - X_i^*}{\Delta X_i}, \tag{2.1}$$

where X_i is the uncoded value of the *i*th independent variable; X_i^* is the uncoded value of the *i*th independent variable at the center point and ΔX_i is the step change value (Jo et al. 2008). As an example, if temperatures 20, 30 and 40 would be tested and ΔX_i is 10, the corresponding coded values are -1, 0 and 1.

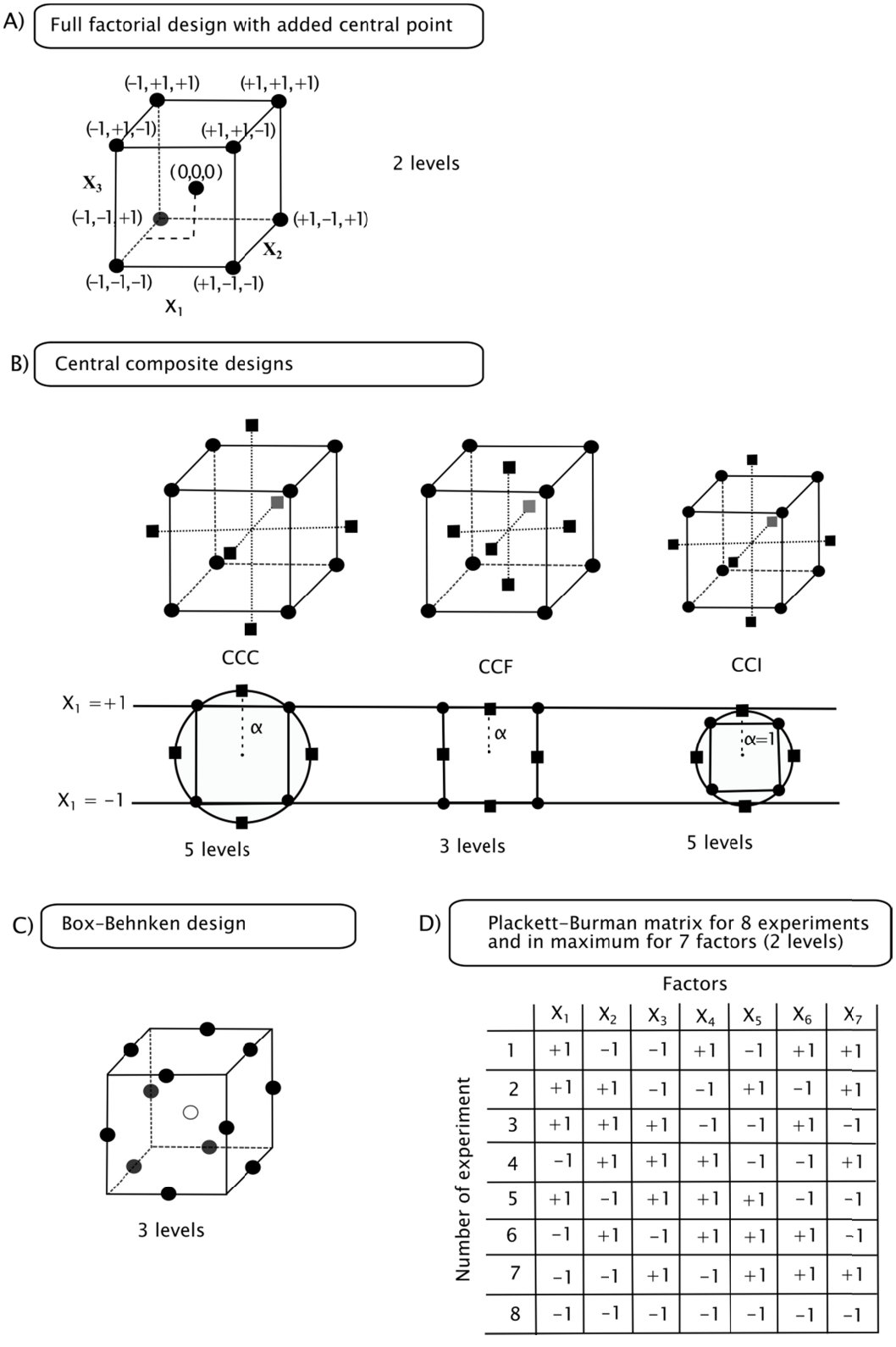
In *full factorial design*, every combination of each factor level is experimentally tested. With this method, mostly two level designs are applied, i.e. all input factors are set at two levels, low and high. Experiments are done such that all possible high/low combinations of all the input factors are tested. E.g., if the effect of *k* factors is studied for each at two levels, a full factorial design has 2^k separate experiments. When the number of factors (*k*) increases, number of experiments becomes excessive and only fraction of the experimental setting specified by the full factorial can be tested. In that case, the method is called *fractional factorial design* (Nist/Sematech, 2013).

Center point experiments (all factors have coded value 0) provide a method to analyze both, process stability and possible curvature of the factors. Thus, 3-5 center points should be added to a full or fractional factorial design (Box and Bhenken, 1960). The center point experiments are not randomized since those are used as controls against process instability, which is more likely detected when the operation is examined regularly (Nist/Sematech, 2013).

Central composite designs (CCD) (also called Box-Wilson designs) (Box and Wilson, 1951) are based on a full or fractional factorial designs with additional center points and two axial points on the axis of each design variable at the distance α from the center (Figure 2.11 B). Based on how axial points are selected, CCD can be divided to *circumscribed* (CCC), *inscribed* (CCI) and *face centered* (CCF) design. The CCC design provides good predictions over the entire design space, but require factor values outside the range of the selected high/low values, thus coded values are higher/smaller than ±1. The CCI design uses only values within the originally specified levels, resulting in lower quality of prediction than CCC. CCF designs provide relatively good predictions and no values outside the original factor range are used. However, with CCF the estimation of pure quadratic coefficients is not as precise (Nist/Sematech, 2013).

Plackett-Burman design (Plackett and Burman, 1946) is approach that only offers information of the effect of single factors, but not on interactions between factors (Figure 2.11 D). It is efficient for screening when only the main effects are of interest. It is mostly applied to detect the most important factors early in the experimentation phase when complete knowledge about the system is usually unavailable and large set of factors are to be screened. All Plackett-Burman designs include 4*n* experiments and with each design the maximum amount of factors within an experiment is 4*n*-1. Thus, if one wants to study effect of five factors, minimum *n* to be selected is 2, the number of experiments will be 8 and number of factors included is 7. The extra two factors can be used as dummy factors to estimate the random measurement errors. In this design, predefined matrices are used to select the experimental combinations (Analytical methods committee, 2013).

In *the Box-Behnken design* (Box and Behnken, 1960), the factor combinations are at the midpoints of the edges and at the center of the process space (Figure 2.11 C). This design requires 3 levels of each factor. The absence of corner values is useful when combined factor extremes should be avoided. Box-Behnken designs require fewer experimental combinations than CCD in the cases involving 3 or 4 factors.



	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	+1	-1	-1	+1	-1	+1	+1
2	+1	+1	-1	-1	+1	-1	+1
3	+1	+1	+1	-1	-1	+1	-1
4	-1	+1	+1	+1	-1	-1	+1
5	+1	-1	+1	+1	+1	-1	-1
6	-1	+1	-1	+1	+1	+1	-1
7	-1	-1	+1	-1	+1	+1	+1
8	-1	-1	-1	-1	-1	-1	-1

Figure 2.11. Methods for statistical design of experiments. A-C present the experimental design for experiments with 3 factors (k=3). A) Two level full factorial design where all the combinations of low and high level are measured. Also the center point is marked. B) Central composite designs. C) 3 factorial Box-Behnken design. D) An example matrix for Plackett-Burman design which can be used for up to seven factors. Abbreviations: CCC: Central Composite Circumscribed, CCF: Central Composite Face-centered, CCI: Central Composite Inscribed.

As presented in Table 2.2, the full or fractional factorial designs and the Plackett-Burman design are often used to screen important factors affecting to the response. These designs can be used for creating first-order (linear) models. When center point measurements are added to full or fractional factorial designs, those can provide information on the existence of second-order (quadratic) effects i.e. curvature. Even though these methods can detect the curvature, the data can not be used to fit the curvature to quadratic model. The first-order model can be used to find the path along which the increase in the values is fastest, i.e., path of steepest ascent, by calculating the gradient for the polynomial model (arrow in Figure 2.12) (Box and Wilson, 1951).

When the experiment has the response surface objective, commonly central composite designs or Box-Bhenken design is used to enable the fitting of the data. Three factorial experiment data can be fitted to quadratic model

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2, \quad (2.2)$$

where Y is the predicted response, b_0 is constant, $b_{1,2,3}$ the linear coefficients, $b_{11,22,33}$ the squared coefficients and $b_{12,\ 13,\ 23}$ the cross-product coefficients. Generally, software is applied to regression and graphical analysis of the obtained experimental data (Figure 2.12). After the data have been fitted to a selected equation, it is important to test the significance of the fit (Mandenius and Brundin, 2008). This can be done, e.g., by analysis of variance (ANOVA).

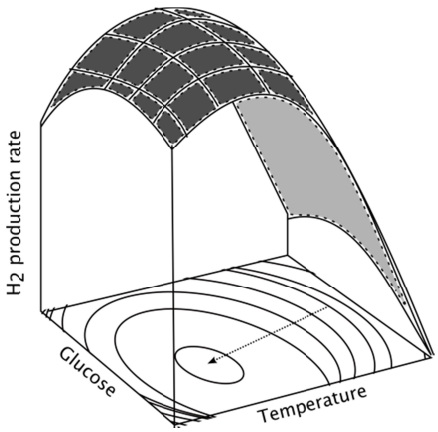


Figure 2.12. An example of three-dimensional surface graph, with data of two factors (glucose and temperature) and the response (H_2 production rate). At the bottom lays two-dimensional contour plot. The arrow represents the direction of steepest ascent, which can lead to local maximum or minimum. If experiment has more than two factors, usually all are drown separately as combination of two factors.

For three factors, the Box-Behnken design offers some advantage since it requires a fewer number of experiments. For 4 or more factors, this advantage disappears. Generally 3 factorial experiments are enough, since those can be repeated with different factors to gain more information of the system. A Box-Behnken and CCI design contains areas with lower prediction quality than CCC. Additionally, Box-Behnken design has limited capabilities for orthogonal blocking compared to the central composite designs (CCD). For example, in

the case of CCD, orthogonal blocking can stand for the division of the design to two sub-designs such that other includes the points of full factorial design with additional center points the other the points form axial levels from central composite design and center points (squares in Figure 2.11 B). Since for proper fitting the use of only 3 levels is not adequate, utilization of CCF design is not advisable. Examples of how the presented methods have been applied for H_2 production and media optimization are presented in Table 2.3.

Various things need to be taken into consideration while planning and proceeding with the optimization experiments. Experimental design should be kept plain with no unnecessary extra variables. Possibilities of growth in the test conditions should be verified (literature, short pretests) and all the measurement equipment should be calibrated prior the experiments. The experimental conditions (continuous inputs) should be kept the same, e.g., avoid removing bottles from the warm cabin during the experiment, and all the events during the experiment should be recorded (who, when, how, the original results and all the unexpected events). Systematic experimentation and complete notes of experiment allow more reliable statistical analysis. Additionally, the possibility to compare the data with other studies increases.

Another method to computationally analyze fermentative H_2 production is to use metabolic models. Those can be applied to describe microbial metabolism, e.g., by analyzing the metabolic models with flux balance analysis. That is discussed in detail in the following chapter. Metabolic modeling creates additional challenges to the experimental design. Often, when the metabolism of microbial cell is modeled, stabile conditions are assumed. The stable conditions consist of constant supply of consumed substances, stable growth rate and constant removal of excreted metabolites from the system. Therefore, the flow rates and composition of growth media should be precisely optimized. In practice, exponential phases of batch fermentations are often used to describe stable conditions since a lot of data is needed to estimate and improve the accuracy of metabolic models.

Table 2.3. Applications of desing-of-experiment methods. **Objectives**: response surface objective (RS), screening objective (S), mixture desing (MD). **Designs**: Box-Behnken desing (BB), Plackett-Burman desing (PB), central composite face-centered design (CCF), fractional factorial desing (FFD), central composite circumscribed desing (CCC), full factorial design (FD). **Others**: Palm oli mill effluent (POME), chemical oxygen demand (COD), *C. beijerinckii* L9 (L9), *C. butyricum* M1 (M1), *B. thermoamylovorans* (B5), not given (n.g.), * (the equation is limited with X1 +X2 +X3 = 1)

Step	Objective	Factors	Design	Equation type	Response	Ref. / (Note)
1/1	RS	Glucose concentration, T, pH	BB	Quadratic	H_2 production rate of *C. tyrobutyricum*	Jo et al. 2008a
1/1	RS	Glucose concentration, T, pH	BB	Quadratic	H_2 production rate of *E. aerogenes*	Jo et al. 2008b
1/1	RS	Glucose concentration, T, pH		Quadratic	H_2 yield of modified *E. coli* DJT135	Ghosh ans Hallenbeck. 2009
1/2	S	8 media components	PB	Linear	H_2 production potential of clostridium sp.	(3 most important selected)
2/2	RS	Glucose concentration, vitamin solution, phosphate buffer	BB	Quadratic	(same as previous)	Pan et al. 2008
1/2	RS	pH, T, COD	n.g.	Quadratic	H_2 production yield of *C. butyricum* from POME	(combination of the results)
2/2	RS	pH, T, COD	n.g.	Quadratic	H_2 production rate of C: butyricum with POME	Chong et al. 2009b
1/1	RS	pH, T	CCF	Quadratic	H_2 production (ml) by acidogenic fermentation	Infantes et al. 2011
1/2	S	7 media components and pH	FFD	Linear	Optical density of *C. butyricum*	(3 most important selected)
2/2	RS	Soyabeancake extract, corn steep flour, NaHCO3	CCC	Quadratic	(same as previous)	He et al. 2004
1/3	RS	ml of L9 and B5 in coculture	FD	Quadratic	Separately 1) H_2 production rate, 2) H_2 prod. potential and 3) lack time of coculture	(Comibnation of data from 1/3 and 2/3 to select limits for 3/3)
2/3	RS	ml of M1 and B5 in coculture	FD	Quadratic	(same as previous)	
3/3	MD	ml of L9, B5 and M1 in coculture	MD	Quadratic*	(same as previous)	Chou et al. 2011
1/1	RS	pH, hydraulic retention time, total solids	CCC	Quadratic	Separately 1) Biogas prod. rate, 2) H_2 prod. rate, 3) H_2 content and 4) H_2 yield from co-fermenting molasses with swin manure	Wu et al. 2013

3 Metabolic engineering and modeling of fermentative H_2 production

The environmental need for sustainable manufacturing methods has driven research towards optimization of metabolic capabilities of bacteria. Mostly, high yields and production rates of desired end products are aspired. By genetic engineering, the metabolism of bacteria can be directed towards higher production of the desired compounds. For example, with gene deletions, unwanted pathways and regulators can be removed and with gene additions, the enzymes catalyzing the desired product can be overexpressed and new metabolic pathways can be created. Traditional approach for the metabolic engineering is to manually analyze metabolic pathways and regulatory systems and based on that, design the mutants. This requires in-depth knowledge of biochemistry.

The basic mathematical approach of microbiologist is to analyze microbial processes with kinetic models. Application of kinetic models for microbial growth, substrate utilization and product formation in the case of H_2 production was recently reviewed by Nath and Das (2011) and will not be discussed here. Another approach is to apply metabolic models to theoretically estimate the metabolic capabilities of an organism. As an example, in the case of H_2 production by *E. coli*, metabolic models can be used to estimate the most probable composition of the excreted fermentation products and to show the realms of possibilities with various different substrates. Additionally, metabolic models can be applied to study biological questions, such as which genes should be removed to result in increased H_2 production and how the bacterial phenotype changes after the deletion (Publication IV). Moreover, new metabolic pathways can be added to the model to simulate the effect of gene insertions. This enables the systematic design of experiments and means of making new, model driven discoveries, towards increased production of H_2.

Small metabolic models have been created to various different microbial species. However larger genome-scale reconstructions of metabolic networks of bacteria have been generated to create better understanding of the bacterial metabolism (McCloskey et al. 2013, Durot et al. 2009). Nowadays, sequencing of bacterial genome is considerably cheap, thus more annotated genomes has become available. This has increased the construction of new genome-scale metabolic models. Updated list of currently available predictive genome-scale metabolic network reconstructions can be found from Supplement 1 of Feist et al. (2009) (http://gcrg.ucsd.edu/InSilicoOrganisms/ OtherOrganisms#bacteria). At the time of writing, the list contained genome-scale network reconstructions for 42 different bacteria and 5 archaea. Most extensive genome-scale models have been created for *E. coli* (Orth et al. 2011).

Following, the genetic regulation and engineering of H_2 production by *E. coli* and *C. butyricum* is discussed in detail. Subsequently, the use of metabolic models for the analysis of bacterial metabolism is introduced.

3.1 Genetic regulation and engineering of H_2 production by *E. coli*

Vast amount of research has been carried out over the metabolism of *E. coli* and means for its genetic modification are well known. Publications over the mechanism of H_2 production in bacterial cells have appeared since the beginning of the twentieth century, e.g., the comparison of the H_2 production by evolutionary variant strains of *E. coli* was published at 1939 (Ordal and Halvorson, 1939). Since the end of the twentieth century, genetic engineering methods to improve fermentative H_2 production has been widely reported (e.g., Sode et al. 1999, Penfold et al. 2003, Penfold et al. 2006, Redwood et al. 2008, Maeda et al. 2007, Maeda et al. 2008, Kim et al. 2009, Fan et al. 2009). In Figure 3.1, genes related to the fermentation pathways of *E. coli* are presented. In the current metabolic models only the enzymes catalyzing certain reactions in the metabolic pathways are included. In reality, also the environmental and enzymatic regulation of transcription, protein maturation and activation and inhibition of the existing enzymes has an effect to the H_2 production rates.

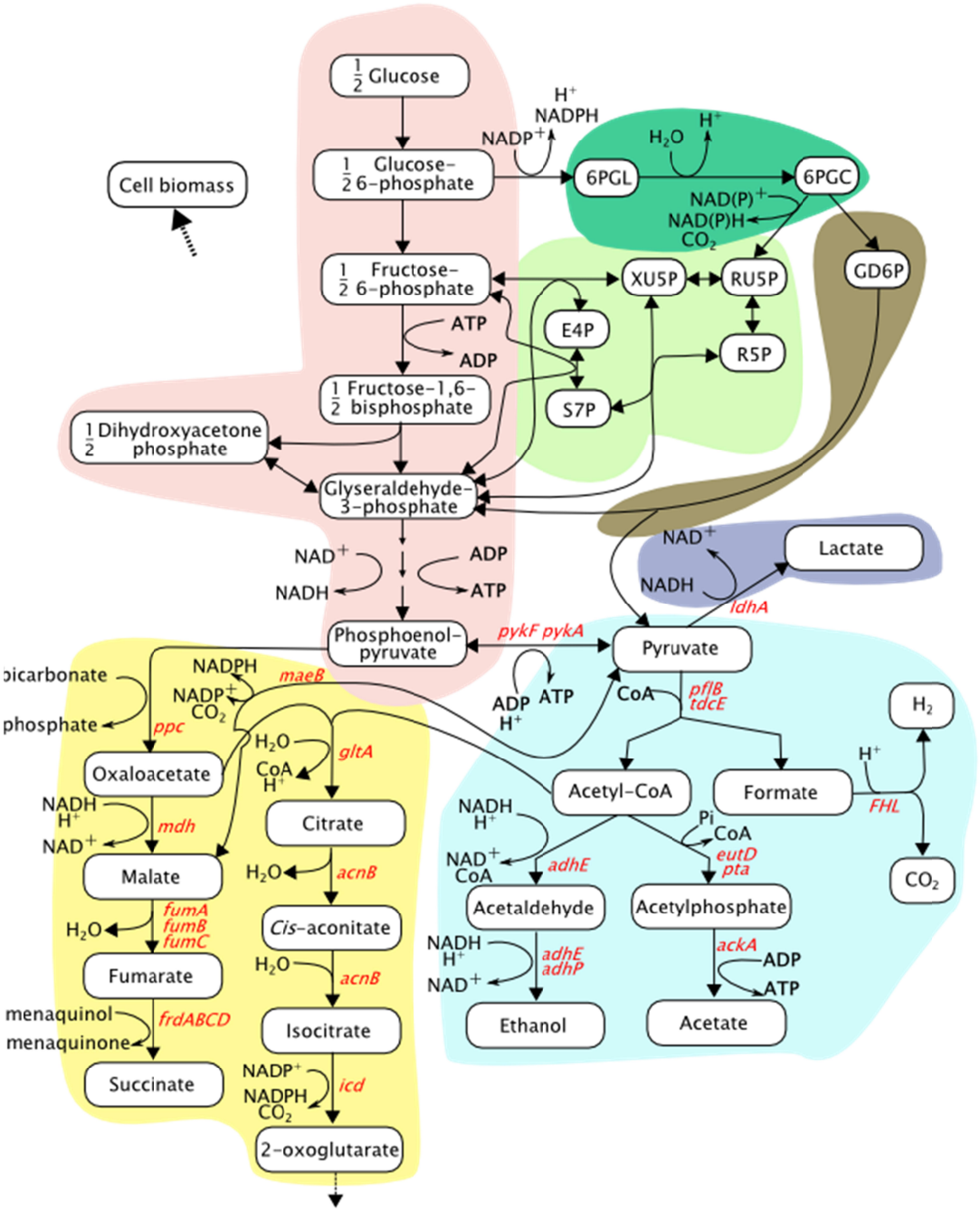


Figure 3.1. Metabolic pathways of degradation of glucose by mixed acid fermentation by *E. coli.* Colored areas mark following metabolic pathways: pink – glycolysis, yellow – branched citrate cycle, green – oxidative pentose phosphate pathway, light green – non-oxidative pentose-phosphate pathway, purple - lactate production, light blue - main pathways for H_2 production.

Final stage of hydrogen production is the biosynthesis of H_2. The reaction $2H^+ + 2e^- \leftrightarrow H_2$ (g), is catalyzed by hydrogenases (hyd). Most hydrogenases can be placed to three classes based on the metals at their active site: [FeFe]-hydrogenases, [NiFe]-hydrogenases and [Fe]-hydrogenases (Shima2008). *E. coli* has [NiFe]-hydrogenases (Maeda et al. 2007). Currently, four hydrogenases, named hydrogenase 1, 2, 3 and 4 (hyd-1, hyd-2, hyd-3 and hyd-4) are found in *E. coli*. Hyd-1 and hyd-2 are mostly responsible for the uptake of

hydrogen. Single deletion of *hybC* (subunit of hyd-2), causes slight increase in H_2 yield (Publication IV) and simultaneous removal of both of the enzymes can increase H_2 production by 32% on glucose (Maeda et al. 2007, Fan et al. 2009). Hyd-4 is thought to be silent (Self et al. 2004, Maeda et al. 2007). Hyd-3 catalyzes reversible reaction of H_2 production, but the rate of synthesis is greater than uptake rate. Thus based on current knowledge of hydrogenases, mainly hyd-3 is responsible for production of H_2 in *E. coli* (Maeda et al. 2007).

Formate hydrogen lyase (FHL) consists of formate dehydrogenase-H (FDH_H) (encoded by *fdhF*) and hyd-3 (encoded by *hycBCDEFG*). It is the main enzyme complex related to H_2 production by *E. coli*. In anaerobic conditions, it catalyzes the oxidation of formate to H_2 and CO_2 (Sawers, 2005). Expression of FHL is inhibited, e.g., by electron acceptors such as oxygen or nitrate, and its synthesis is controlled by the formate concentration and pH (Wang and Gunsalus, 2003, Bagramyan and Trchounian, 2003). Additionally, the expression of FHL is under genetic control of FhlA and HycA. FhlA promotes the transcription of FHL complex enhancing the H_2 production. HycA may interact directly with the FhlA protein or prevent the binding of FhlA to activator sequences and thus to repress the transcription of FHL (the mechanism is not yet known). Single deletion of *hycA* does not increase the H_2 yield (Kim et al. 2009, Publication IV), but simultaneous deletion of *hycA* and genes of hyd-1 and hyd-2 increases the H_2 yield from glucose (Kim et al. 2009, Maeda et al. 2007). Volumetric production rate from formate was increased two orders of magnitude and the specific hydrogen production rate by 2.8 folds by simultaneous disruption of *hycA* and overexpression of *fhlA* (Yoshida, et al. 2005). However, when the same double mutant utilized glucose instead of formate as substrate, the mutation had no effect on H_2 production (Yoshida, et al. 2006). It indicated that activity of FHL complex is not a limiting factor for hydrogen production from glucose (Yoshida, et al. 2006).

In addition to FDH_H, two other formate dehydrogenases, formate dehydrogenase-N (FDH_N: *fdnG*) and formate dehydrogenase-O (FDH_O: *fdoG*) are synthetized in *E. coli*. Both can metabolize formate to CO_2 connected with reduction of nitrate to nitrite by nitrate reductase A (Wang and Gunsalus, 2003). Removal of *fdoG* with the simultaneous deletions of *hyaB*, *hybC* and *hycA* at the background, increased H_2 production 1.7-fold compared to wild-type (Maeda et al. 2007). With the same background, deletion of either *fdnG* or *focA* was not effective for increasing the H_2 production (Maeda et al. 2007).

FocA is a formate transporter and in some experimental setting its inactivation can increase H_2 production (Fan et al. 2009), even though the positive effect of single deletion was not observed in Publication IV. Simultaneous inactivation of hyd-1, hyd-2, FDH_O and FDO_N can be achieved by inactivation of twin-arginine translocation (Tat) protein translocation system, that is responsible in transporting theses enzymes (among others) from periplasm to cytoplasm (Berks et al. 2003). FHL_H and hyd-3 are located to cytoplasm, thus the

deletion of Tat translocase (coded by *tatABCE*) does not affect to their function (Penfold et al. 2006). Deletion of *tatC* was enough to cause same effect than the removal of all subunits of Tat translocase and doubled the H_2 production rate from 0.88 to 1.70 mL H_2 mg dry weight^{-1} L culture^{-1} (Penfold et al. 2006).

Among other functions, global transcriptional regulators, Fnr and narL, can effect on H_2 production. Fnr is a transcriptional regulator that mediates the transition from aerobic to anaerobic growth through the regulation of hundreds of genes (Salmon et al. 2003). It activates genes involved in anaerobic metabolism and represses genes involved in aerobic metabolism (Salmon et al. 2003). The overexpression of *fnr* increased and its deletion along with the deletions of *hyaB*, *hybC* and *hycA* drastically decreased the H_2 production (Fan et al. 2009, Maeda et al. 2007). NarL phosphate, binds at the *fdhF* promoter to suppress *fdhF* gene expression thus inhibiting the function of FHL (Wang and Gunsalus, 2003). By deletion of *narL*, increase in H_2 yield from glucose has been achieved (Fan et al. 2009). Perhaps due to differences in the cultivation media, the positive effect of *narL* deletion was not seen in Publication IV. Figure 3.2 presents a summary of the regulation of formate metabolism.

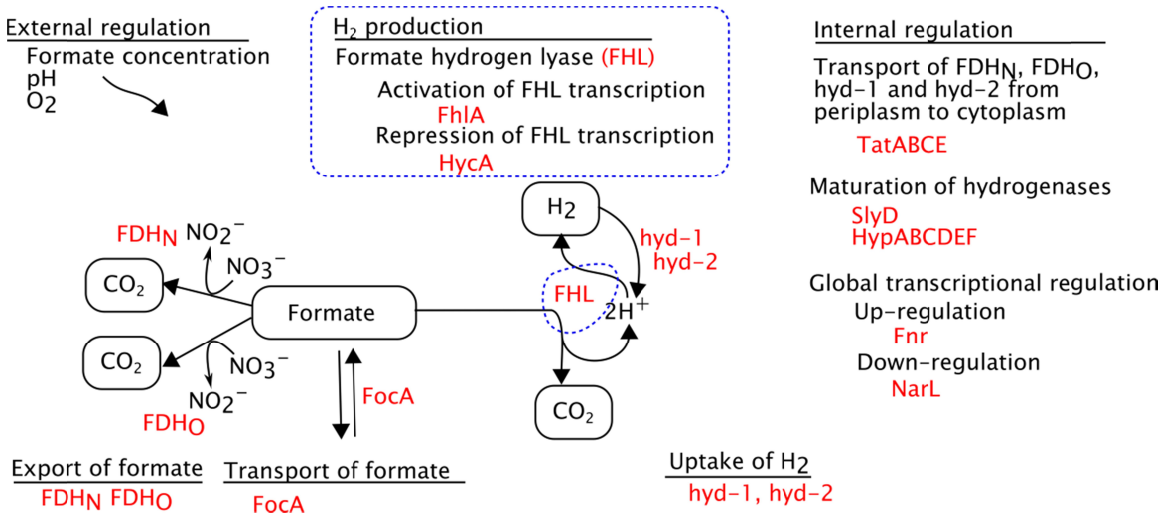


Figure 3.2. Main factors affecting to formate metabolism of *E. coli*. The responsible proteins are shown in red. (FDH –formate dehydrogenase, hyd –hydrogenase)

Pyruvate is a key branch point in the central carbon metabolism and its consumption is mainly regulated by the redox state of the cell. During fermentative metabolism greater than 95% of pyruvate is used as terminal electron acceptor for NADH oxidation and less than 5% is used for biosynthesis (Figure 2.1.A) (Causey et al. 2004). In *E. coli*, pyruvate formate-lyase (PFL) catalyzes the production of formate from pyruvate. Its function is inactivated by oxygen and some small-molecule thiols, but its enzymatic inactivation is not unquestionably shown (Nnyepi et al. 2007). Thiols, such as ethanethiol and cysteine, are observed to inactivate PFL. Ethanethiol, inactivates PFL rapidly, while cysteine causes much slower inactivation. In contrast, larger thiols such as glutathione do not cause

detectable inactivation (Nnyepi et al. 2007). PepD is one of the four enzymes known to catalyze the degradation of glutathione to L-cysteine and glycine in *E. coli* (Suzuki et al. 2001). The deletion of *pepD* increases the H_2 yield from glucose, which can be due to the higher activity of PFL caused by the mutation (Publication IV). Main factors affecting to the pyruvate metabolism are shown in Figure 3.3.

In addition to increasing the activity of the reactions producing H_2, metabolic fluxes have been redirected towards H_2 by removing pathways that utilize phosphoenolpyruvate (pep) or pyruvate for other purposes (Figure 3.3). The H_2 production yield (mol of H_2 per mol of glucose) has been improved by single deletion of *ldhA*, but this mutation alone has not been found to have effect on production rates (Sode et al. 1999). Additionally, single mutation of *frdB, frdC* and *frdD* increased the H_2 yield from glucose (Publication IV). Simultaneous deletion of *ldhA* and *frdBC* enhanced the hydrogen yield from wild-type strain W3110 1.08 mol/mol glucose to 1.82 mol/mol glucose and increased the specific H_2 production rate 1.4 times higher than wild-type strain (Yoshida, et al. 2006). In addition, the deletion of *ppc*, which product catalyzes the production of oxaloacetate from phosphoenolpyruvate, can increase H_2 production (Fan et al. 2009), even though this was not observed in Publication IV. Furthermore, the single deletion of *maeB* and *mdh* increased the H_2 yield on glucose (Publication IV) (Figures 3.1. and 3.3). Overexpression of *maeB* increased the succinate production, thus its deletion might direct metabolic flow towards formate production (Kwon et al. 2007).

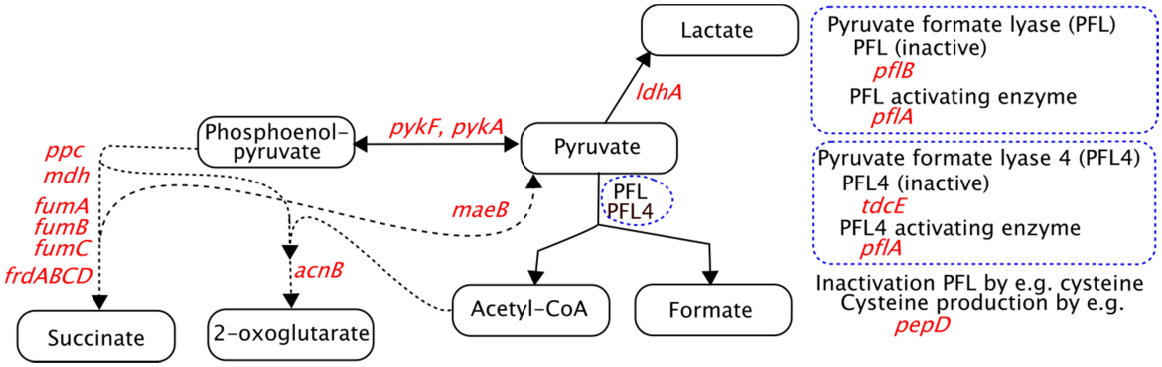


Figure 3.3. Main factors affecting to pyruvate metabolism of *E .coli*. The genes encoding the related proteins are shown in red.

The best improvements in H_2 production have been achieved by multiple deletions, including deletions related to pathway engineering and enzymatic regulation. For instance, the H_2 yield of strain BW25113 was increased 1.5 times by simultaneous deletions of *hycA hyaAB hybBC ldhA frdAB* (Kim et al. 2009) and *hyaB hybC hycA fdoG ldhA* (Maeda et al. 2007). When the genetic alterations are combined with improved environmental conditions, e.g., with low partial H_2 pressure (pH_2), even larger improvements have been achieved (Kim et al. 2009, Maeda et al. 2007).

Even though the metabolism of *E. coli* is vastly studied, there are still many unexplored metabolic interactions. Related to *E. coli* hydrogen production, this issue has been tackled by utilizing the Keio collection, which is a complete set of single-gene deletions of all non-essential genes of *E. coli* K-12 (Baba et al. 2006). The effects of all single deletions on H_2 production have been screened by chemochromic membranes (Mohd Yusoff et al. 2012). In Publication IV, the effects of 81 separate gene deletions on *E. coli* mutants (gained from the Keio collection) are studied by batch experiments. These methods were discussed under Section 2.3.

In addition to applying the possibilities of increasing H_2 production by modification of intrinsic characteristics of *E. coli* by under and overexpression, its proteins can be altered to increase their activity. For example, the H_2 yield from formate has been increased 9-fold by random protein engineering of HycE or FhlA (Maeda et al. 2008, Sanchez-Torres et al. 2009). Moreover, a lot of research has been conducted to create oxygen tolerant hydrogenases (Liebgott et al. 2011, Hamdan et al. 2012).

E. coli is often used as host to genes gained from other organisms. Recombinant strains of *E. coli* have been created, e.g., by transferring foreign [FeFe]-hydrogenase genes from other bacteria. For example, transferring [FeFe]-hydrogenase of *Enterobacter cloacae* (Mishra et al. 2002, Chittibabu et al. 2006) or *Clostridium butyricum* (Subudhi and Lal 2011) into *E. coli* has enabled production of more than 3 mol of H_2 per mol of glucose. Additionally, cyanobacterial [NiFe]-hydrogenase has been successfully expressed in *E. coli* (Wells et al. 2011).

In addition to the insertions of single enzymes, completely new synthetic pathways can also be created. As previously presented (Fig 2.2), *E. coli* can produce H_2 only through formate, whereas clostridial species can use ferredoxin for H_2 production. [FeFe]-hydrogenase encoding gene (*hydA)* and maturation genes of *Clostridium acetobutylicum* (*hydF-hydE-hydG)* and [4Fe-4S]-ferredoxin encoding gene from *Clostridium pasterianum* were converged onto a single plasmid and expressed in *E. coli* lacking the native H_2 production pathway (Akhtar and Jones, 2008b). The expression of this synthetic pathway with simultaneous expression of *E. coli ydbK* (predicted pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase (Serres et al. 2001)) and deletion of *iscR* (negative transcriptional regulator of *isc*,"iron-sulfur cluster" operon) increased the H_2 yield from 0.01 to 1.46 moles of H_2 per mole of glucose, demonstrating the possibilities of the synthetic pathways (Akhtar and Jones, 2008a, 2008b, 2009, Veit et al. 2008). The function of synthetic H_2 production pathways can be improved by insulating the electron transfer circuit from existing cellular metabolism. This can be done, e.g., by deletion of competing reactions, optimization of binding surfaces, direct protein-protein fusion and localization to a synthetic protein scaffold (Agapakis et al. 2010).

Genetic modification has also been used to enable the use of alternative substrates, e.g., addition of AmyE of *Bacillus subtilis* to *E. coli* enabled starch-dependent H_2 synthesis in minimal media (Akhtar and Jones, 2009). Another example is the addition of genes necessary for sucrose transport into *E. coli*, thus enabling the sucrose utilization from nougat waste water (Penfold and Macaskie, 2004). Summary of the methods used for genetic manipulation are shown in Table 3.1.

Table 3.1. Methods of genetic engineering to enhance H_2 production

Technique	Target	Notes	References
Gene deletion	Disturbing competing pathways and enzymes repressing the H_2 production	E.g. directing of pyruvate towards formate in *E. coli* and prevention of consumption of NADH for other purposes than H_2 production. Removal of enzymes inhibiting the H_2 production.	Yoshida et al. (2006)
Overexpression of native genes	Enhancing the activity of native genes.	E. g., overexpression of hydrogenases or activators of desired reactions.	Yoshida et al. (2005)
Modification of the native proteins	Enhancing the function of a protein.	E.g., improving the O_2 tolerance of hydrogenases.	Liebgott et al. (2011)
Construction of synthetic pathways	Addition of foreign genes, restricted from better H_2 producers which genetic manipulation is more challenging. Construction of new pathways for H_2 production.	E.g., insertion of ferredoxin related hydrogen production pathway to *E. coli*.	Akhtar and Jones (2008a)

3.2 Genetic regulation and engineering of H_2 production by *Clostridium butyricum*

Method for genetic modification of *Clostridium butyricum* was recently discovered (Cai et al. 2011). Until then the main hypotheses of how to increase the H_2 production of *C. butyricum* was the removal of the pathways producing butyrate or ethanol, thus possibly increasing the metabolite flow towards acetate production. However, separate deletions of the pathways revealed that the H_2 production was not increased without additional changes in the cultivation conditions. Instead, after the blockage of the ethanol pathway the flow of metabolites was directed to butyrate production. Moreover, the blockage of ethanol pathway increased the metabolite flow towards lactate production (Cai et al. 2011, Cai et al. 2013). These issues might be solved by multiple deletions. Metabolic pathway map of *C. butyricum* is shown in Figure 3.4.

Metabolism of *C. acetobutylicum* is better known than metabolism of *C. butyricum*. Anyhow, poor understanding of central metabolic pathways of all clostridia hinders the metabolic engineering (Crown et al. 2011). In Figure 3.4, non-oxidative path of the pentose-phosphate pathway and the split citric acid cycle are presented based on pathways existing in *C. acetobutylicum* (Crown et al. 2011). Based on gene expression analysis of *C. acetobutylicum*, the expression of genes related to the citric acid cycle is high and of genes related to non-oxidative path of pentose-phosphate pathway is low. Oxidative path of pentose-phosphate pathways is found to be silent in *C. acetobutylicum* (Crown et al. 2011).

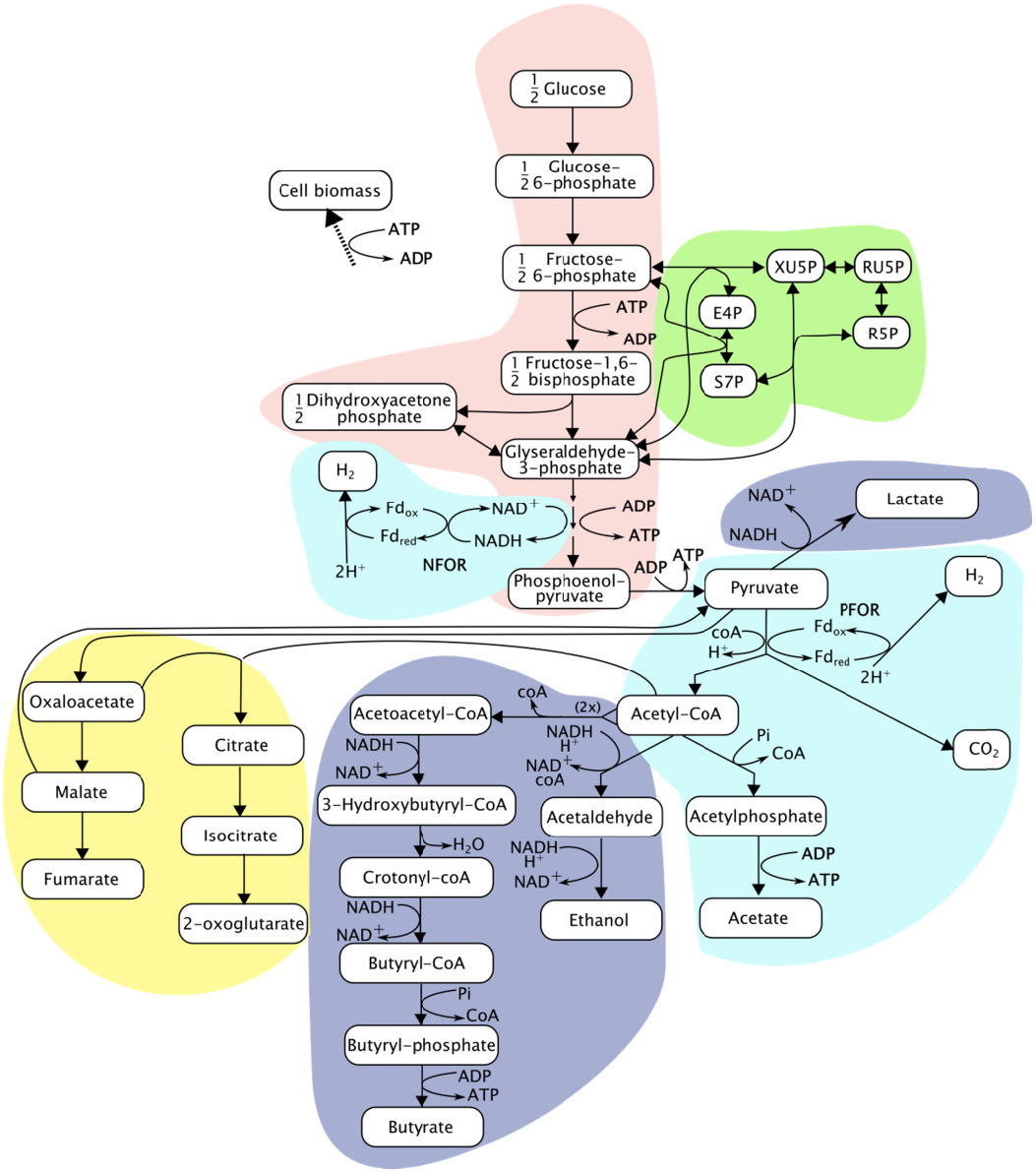


Figure 3.4. Potential metabolic pathways of glucose degradation by *C. butyricum*. Colored areas indicate the following metabolic pathways: pink – glycolysis, yellow – branched citrate cycle, light green – non-oxidative pentose-phosphate pathway, purple- lactate, butyrate and ethanol production, light blue- main pathways for H_2 production. Yellow and green areas are drawn based on metabolism of *C. acetobutylicum*.

C. butyricum is strictly anaerobic, thus its cultivation is more challenging than of *E. coli*. Exposure to oxygen causes irreversible changes to most [FeFe]-hydrogenases found in

bacteria. Anyhow, some species of Thermotogales have [FeFe]-hydrogenases which are tolerant to small amounts of O_2 (Tosatto, et al. 2008). The knowledge of these enzymes could be used in designing mutations for clostridial hydrogenases in order to increase the O_2 tolerance. Additionally, it has been found that deletion of *perR* in *C. acetobutylicum* causes increased resistance to oxidative stress (Hillmann et al. 2009). Creation of less oxygen sensitive clostridial species could simplify the genetic engineering and use of clostridia for H_2 production.

Since the genetic manipulation of *C. butyricum* has only recently been discovered, just a few reports are available (Cai et al. 2011, Cai et al. 2013). Since tools for altering the genome of *C. butyricum* now exists, it is possible to gain much new information over its metabolism. A hypothetical presentation of the possibilities of combining the H_2 production mechanisms of *E. coli* and *C. butyricum* is illustrated in Figure 3.5.

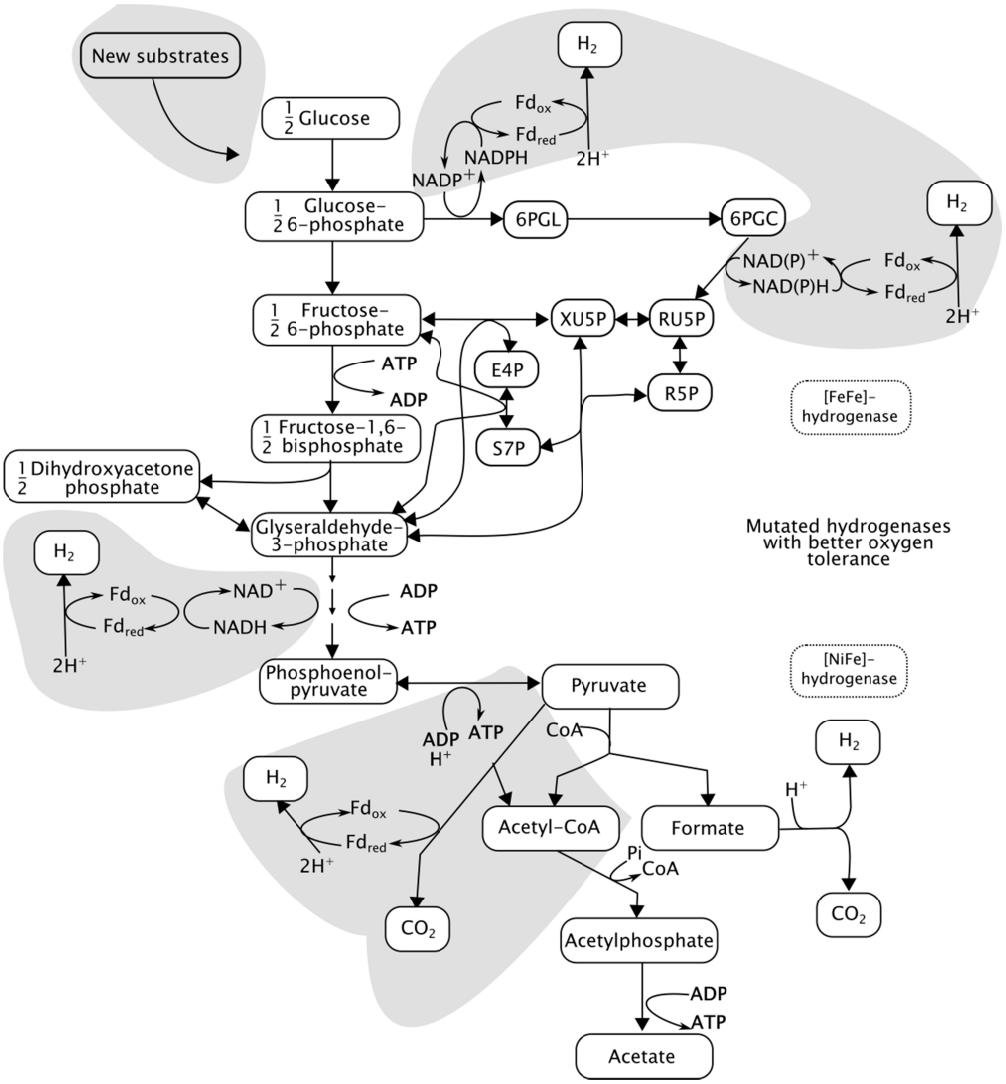


Figure 3.5. Hypothetical pathway of how the metabolic engineering approaches presented could be applied to *E. coli* to enhance H_2 production. Light gray areas present the inserted pathways.

3.3 Metabolic models

Metabolic models are topological (interaction based) presentations of reactions occurring within metabolic pathways. Creation of even small, realistic kinetic model of microbial metabolism requires vast knowledge of, e.g., enzymes, reaction rates, cofactors and inhibitors related to the system. The type of data added to the model is dependent on what kind of analyzes it will be subjected to. Even with the most studied microorganism, *E. coli*, majority of the reactions in the metabolic network lack the detailed knowledge of reaction kinetics. Therefore, constraint based metabolic models are often used to gain more understanding of bacterial metabolism. In current metabolic models, the emphasis is on combining genetic information with reactions behind the growth and metabolite production, rather than on exact knowledge of the regulation within each reaction. By this means, alternatively one optimal solution for flux distribution or a set of possible solutions that satisfy the terms given in the metabolic model can be simulated. Metabolic models can be used, e.g., to estimate the bacterial phenotype based on the given substrate and nutrients. As an example, the possible end product distributions of mixed acid fermentation can be simulated or effect of gene deletions on the whole metabolic system can be estimated (Publication IV and V). Here, we focus on metabolite level models. Following, the metabolic reconstruction required for the constraint-based analyses is described.

In metabolic models, metabolic pathways are presented as a set of stoichiometric reactions. The reactions can be divided into two groups, internal reactions occurring inside the system and exchange reactions transferring metabolites through the cell membrane. The exchange reactions can be constrained based on cultivation media, creating a group of uptake reactions. Additionally a biomass reaction, which is a group of, e.g., small molecules, nucleotides, amino acids and cell wall components, is formulated in order to determine the components that have to be produced to enable growth. Basic components of a metabolic model are presented in Figure 3.6.

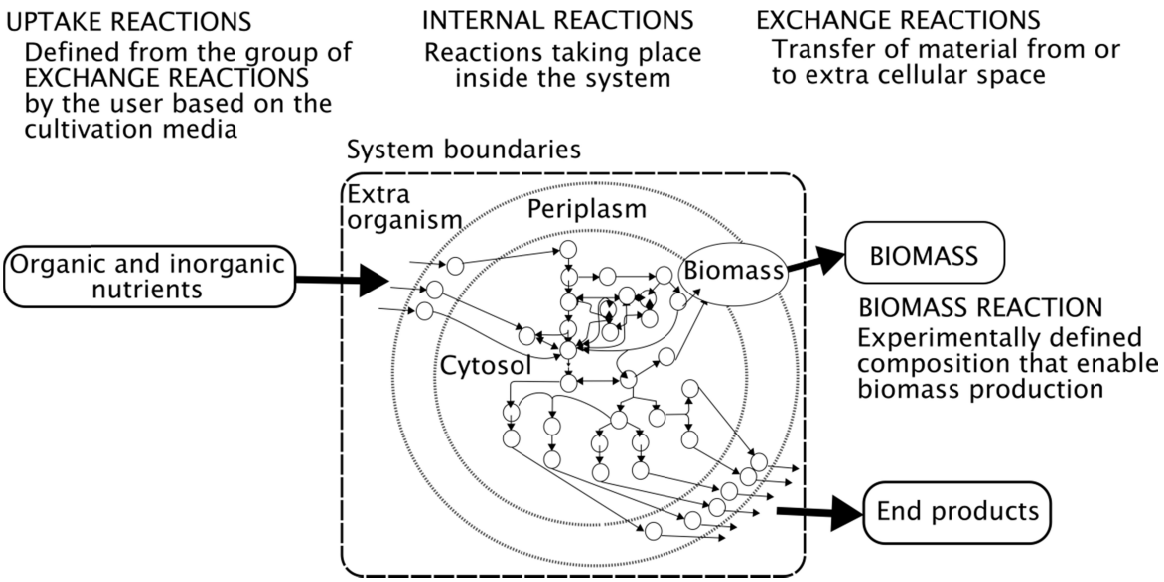


Figure 3.6. While constructing the metabolic model, system boundaries must be defined. Reactions occurring within the system are internal reactions and reactions transferring metabolites over the system boundaries are exchange reactions. Modified from Durot et al. (2009).

While analysis of complete genomes of bacteria has become easier and less expensive, more genome-scale models (GEMs) have been constructed. Extensive protocols of how to create genome-scale reconstruction of bacterial metabolism has been released (Thiele and Palsson, 2010, Durot et al. 2009). Largest GEM has been reconstructed for *E. coli* K-12 MG1655, including 1366 genes, 2251 metabolic reactions and 1136 unique metabolites (Orth et al. 2011). Earlier version of this GEM (Feist et al. 2007) was applied in Publications IV and V. GEMs connect the information of genome to metabolic reactions. From annotated genome, relation of genes to proteins and proteins to reactions are defined. All metabolites involved in reactions are presented and balanced. Additionally, rates and directionalities of reactions can be constrained based on experimental data and thermodynamic calculations. As an example, information related to a reaction in the model is shown in Figure 3.7.

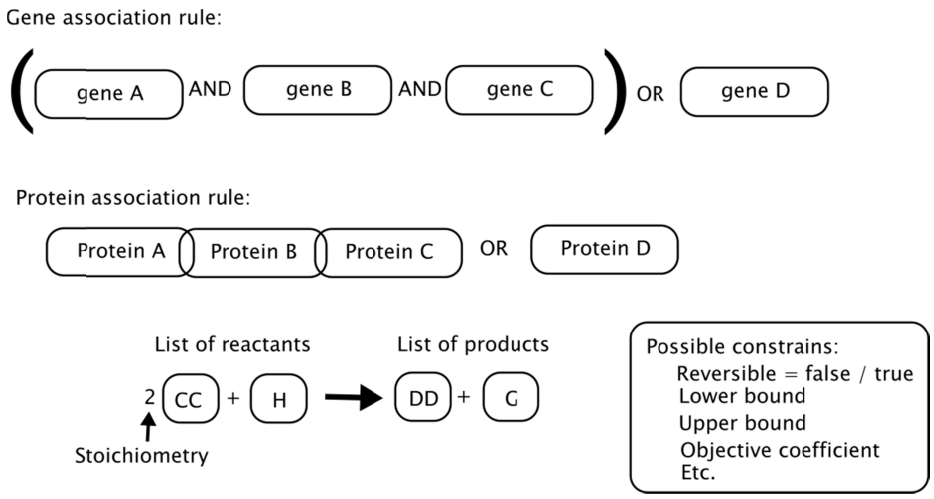


Figure 3.7. Schematic presentation of the components present in the GEM. The gene association rule declares, which genes are essential for reaction to occur. In this example, removal of only one gene has no effect on the given reaction. If gene D along with either gene A, gene B or gene C is removed also the reaction is removed from the pathway. The thermodynamic information i.e. direction of the reaction is given. Limits for the production rates can be determined. Additionally, if the metabolite is necessary for biomass production, objective coefficient can be defined. Similar description is given to all reactions in the model.

The aim of creating models of metabolism is the understanding of underlying mechanism behind the observed phenomena, e.g., why cell produces certain metabolites at the given conditions. The reconstructed metabolic model as such does not give that information, thus computational tools are needed to find answers hidden in the network. The type of analysis used is based on the question in hand; is it question of fundamental phenomena of metabolism in general or detailed question of production of some metabolite. Various approaches to analyze cellular networks exist. Stelling (2004) has dived network models into three categories: 1) interaction-based models, 2) constraint based models and 3) mechanism-based models (also called kinetic or dynamic models) (Figure 3.8).

Interaction based models are static models that describe network topology, i.e. the pattern of interactions between the components. Most often those are not dependent on reaction stoichiometry or rates (Stelling, 2004, Wagner and Fell, 2001). Interaction based models, such as graph theoretical models can be applied, e.g., to describe the large-scale organization of metabolic networks. Additionally, those can be used to divide the metabolism to modules, each responsible for distinct function, and to find highly connected substrates that can provide connections between those modules (Jeong et al. 2000). This can increase general understanding of the metabolism. Jeong et al. (2000) found that the metabolism of microorganism appears to be a robust and error-tolerant scale-free network (Barabási2009). Based on graph theoretical model, Wagner and Fell (2001) presented the central role of citric acid cycle in *E. coli*. Additionally, they suggested that metabolic network of *E. coli* is a small-world graph, which may allow a metabolism to react rapidly to perturbations. It has been proposed that the microbial metabolic state is a trade-off between two principles: optimality under given condition and minimal adjustment between conditions (Szhuets2012). These theories support each other.

As interaction based models, also constraint-based models are static models, but those contain also information over reaction stoichiometry. Additionally, other constrains, such as limitation of metabolite production rates, can be defined to make the predictions more accurate. Constrain-based modeling has been applied in Publication IV and V, and will be discussed in detailed in the following section.

Mechanism-based models are dynamic models. Those include information over reaction stoichiometry and kinetics. The outcome of mechanism-based genome-scale model, including information over signaling and gene regulation would be detailed, quantitative prediction of cellular dynamics (Stelling, 2004). The lack of information restrains the creation of such models. So far, mechanism-based models have been applied for modeling kinetics of smaller subunits of metabolism, such as modeling the function of *lac* operon (Vilar et al. 2003) and biosynthesis of valine and leusine in *C. glutamicum* (Dräger et al. 2009). Even though construction of genome-scale dynamic models would be challenging, proposals of how these models should be approached has been presented (Steuer, 2011, Jamshidi and Palsson, 2008).

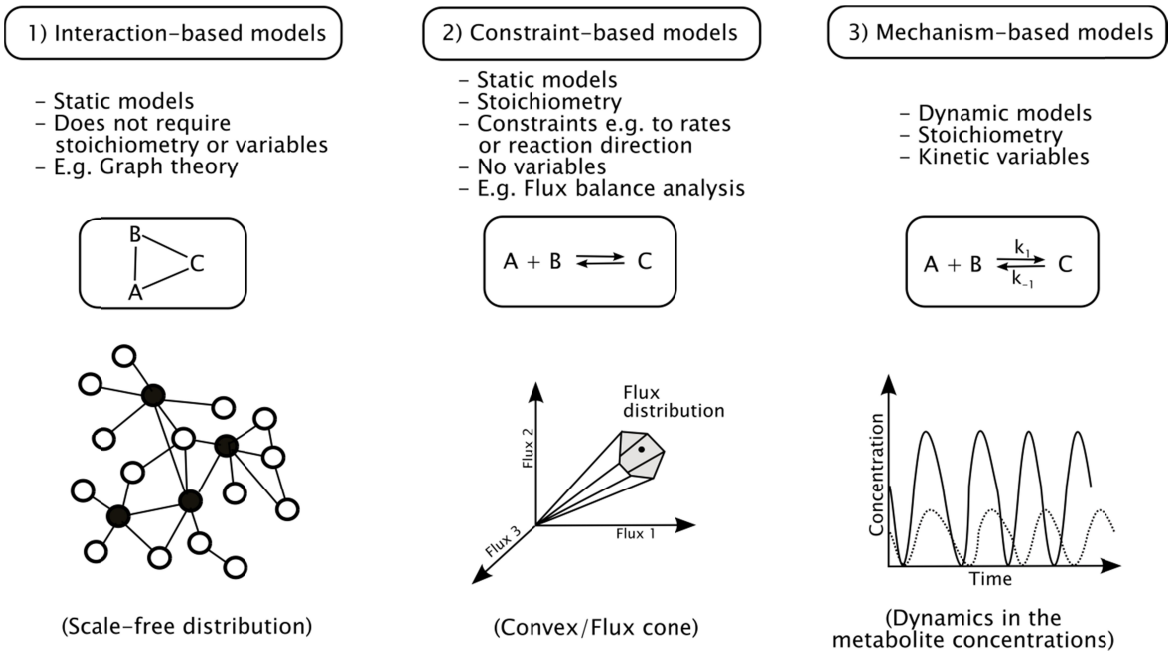


Figure 3.8. Division of network models into three categories: 1) interaction-based models, 2) constraint based models and 3) mechanism-based models (also called kinetic or dynamic models). Modified from Stelling (2004).

3.4 Constraint based models

The selection of the method used to analyze the metabolic network is dependent on the goal of the analysis. Based on the goals, Trinh et al. (2009) has divided constraint based modeling tools to three different groups: 1) flux balance analysis, 2) metabolic flux analysis, and 3) metabolic pathway analysis. Theory behind all three methods is based on

principle of mass conservation. Equation to describe the mass conservation of metabolites in a system of defined volume can be written as

$$\frac{d\boldsymbol{x}}{dt} = \mathrm{S} \cdot \boldsymbol{v} - \mu \cdot \mathbf{x} \quad , \tag{3.1}$$

where **x** is vector of metabolite concentrations, S is stoichiometric matrix, $\boldsymbol{v}$ is vector containing reaction rates i.e. flux distribution and μ is the specific dilution rate associated with the change in volume of the system. Since the contribution of volume change to the concentration changes of metabolites within the system is considered to be insignificant and at steady state there is no accumulation of internal metabolites, equation (3.1) can be simplified to

$$\mathrm{S} \cdot \boldsymbol{v} = \mathbf{0}. \tag{3.2}$$

Since all reactions are not reversible, thermodynamic constrains are given to determine the limits to reaction rates. If the reaction v_i is irreversible it can have only positive values, thus for all *i*,

$$v_i \geq 0 \ . \tag{3.3}$$

Small example of how metabolic networks are converted to a format, form which reaction rates can be solved is illustrated in Figure 3.9.

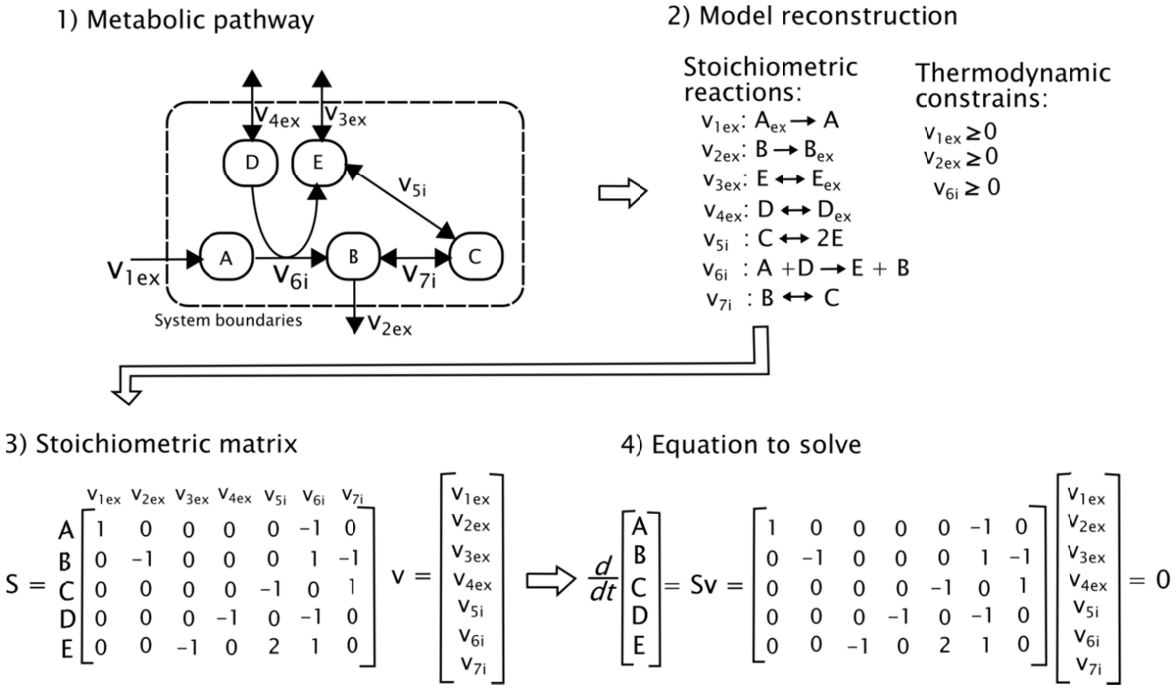


Figure 3.9. 1) Small theoretical metabolic network. 2) Based on stoichiometry of reactions within the network and the thermodynamics (i.e. for irreversible reaction rate $v_i \geq 0$) a stoichiometric matrix (3) is constructed, where horizontal lines present the substrates and vertical lines reactions. 4) The matrix is used to find solution for Equation 3.2.

The solution to Equation 3.2 is most often underdetermined, i.e. amount of metabolites is less than amount of reactions to be solved. The previously mentioned division of constraint based methods to metabolic flux balance analysis, flux analysis and metabolic pathway analysis is made based on means to solve the Eq. 3.2.

3.4.1 Flux balance analysis

Flux balance analysis (FBA) has been developed to increase quantitative understanding of metabolism in order to design and optimize bioprocesses (Varma and Palsson, 1994). It is widely used approach for analyzing biochemical networks, especially in analysis of GEM (Orth et al. 2010). FBA can be used to calculate the flow of metabolites through network, without need of experimental data. The method is based on knowledge of 1) stoichiometric matrix , 2) constraints and 3) objective function (Fig. 3.10). Since Eq. 3.2 can have infinite number of solutions, constraints need to be given. The most important constraints are thermodynamic constraints that state whether the reaction is reversible or irreversible (Eq. 3.3). Moreover, the constraints are used to define the growth media, i.e. uptake reactions from the group of the exchange reactions (Figure 3.1). Additional constraints can be given based on the application of the FBA. Based on the constraints, the amount of possible solutions can be limited giving the allowable solution space. To gain only one optimal solution for flux distribution, an objective function is defined. The most used objective function is maximization of biomass, even though also other objectives, such as minimization of nutrient utilization or maximization of desired product, can be utilized (Raman and Chandra 2009). Simulation results have been noted to be most consistent with experimental data when combination of maximization of ATP and biomass yield together with minimization of sum of absolute fluxes achieved is used as an objective. In other words, flux states evolve under the trade-off between optimality under given condition and minimal adjustment between conditions (Schuetz et al. 2012). The outline of FBA is shown in Figure 3.10.

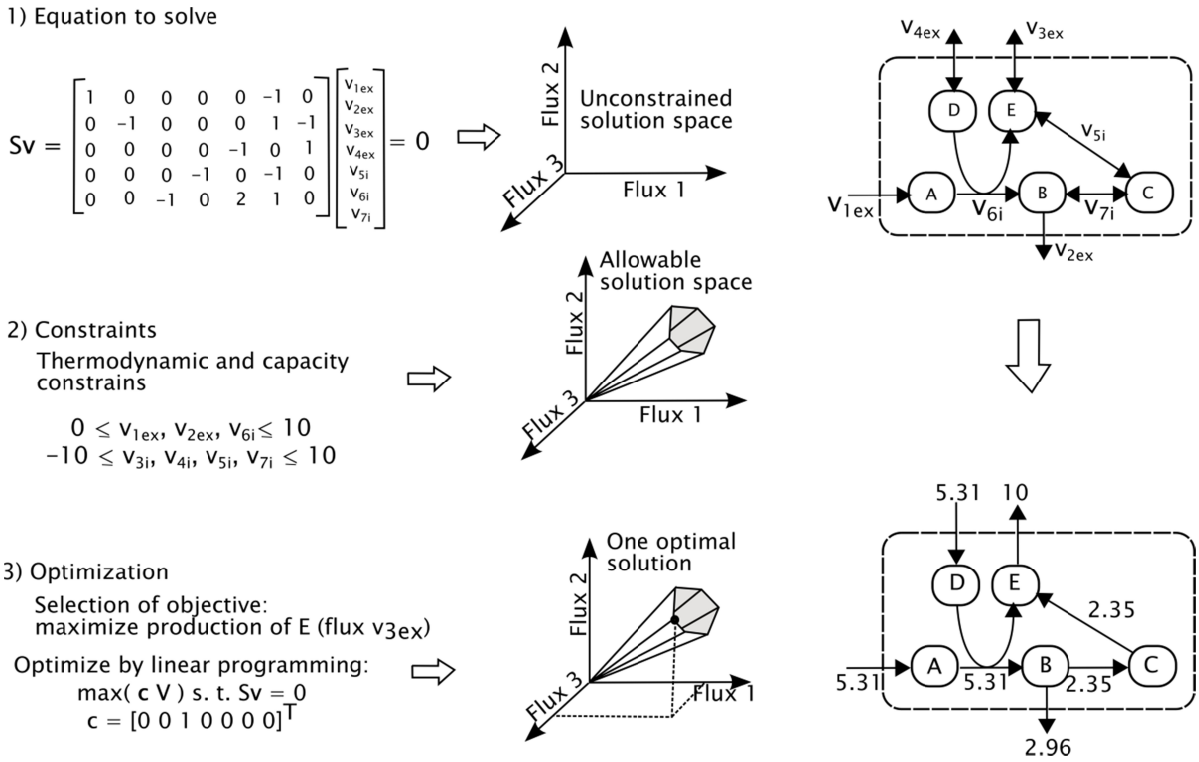


Figure 3.10. The basics of FBA analysis. 1) Stoichiometric presentation of metabolic network within Eq. 3.2 and the unconstrained solution space describing the infinite amount of solutions for Eq. 3.2. 2) Addition of constrains creates a cone that includes all accepted solutions for Eq. 3.2. 3) Production of **E** is selected as objective function (e.g. biomass in real application) and its production is optimized by linear programming. The values gained by FBA are presented within the network. (Partly based on Orth et al. 2010)

Applications of flux balance analysis

FBA have been used to solve various questions based on metabolic networks. Small scale metabolic networks have been created to analyze hydrogen production by *E. coli* (Manish, et al. 2007, Kim et al. 2009), *C. butyricum* (Cai et al. 2010) and mixed communities (Chaganti et al. 2011). Manish et al. (2007) constructed metabolic network of 27 reactions, that were related to glycolysis and mixed acid fermentation and Kim et al. (2009) presented the same more detailed with 66 reactions. Manish et al. (2007) applied the FBA for studying the effect of gene deletions within the small network and to find feasible operating space i.e. end product distribution for H_2 production. Kim et al. (2009) constructed mutant with eight deletions related to H_2 production. They studied the mutant in batch experiments with various glucose concentrations. Afterwards, the results were compared to simulations with FBA to gain more understanding over the changes caused to internal fluxes by the mutation and glucose level. Experimentally measured acetate and formate values were used as constrains to estimate the H_2 production. Additionally, they made gene additions to the model to make predictions over the effects of additions to the H_2 production. Cai et al. (2010) constructed a metabolic network model with 34 reactions for glucose fermentation by *C. butyricum* W5. Their focus was to study the effect of

changes on glucose concentration and pH to internal fluxes. They experimentally measured the concentrations of excreted metabolites at different conditions and used those as constrains for FBA analysis. They found that the change of pH had a greater effect on internal fluxes than the glucose concentration. Even though these studies are made with the focus of H_2 production, the main outcome is not merely higher yield of H_2 but increased understanding over the microbial metabolism. Above mentioned, Manish et al. 2007, Kim et al. 2009 and Cai et al. 2010 said that they applied metabolic flux analysis (MFA) for *in silico* analysis. Since they used linear optimization with objective function, they actually applied FBA.

An interesting application of FBA is metabolic modeling of microbial communities. Chaganti et al. (2011) created a small scale model of mixed microbial community by reconstructing metabolic model of universal organism that produces all the metabolites which are observed during anaerobic fermentation. Since some bacterial species, such as methanogens, consume H_2 during fermentation, they studied the effect of linoleic acid and pH on fermentative H_2 production by mixed microbial culture. Linoleic acid is known to act as inhibitor towards many H_2 consuming species, but its addition or pH variation solely was not found to be applicable method to increase H_2 production in mixed bacterial communities. Stolyar et al. (2007) had more accurate approach for analyzing a mutualistic microbial coculture of two bacterial species with sequenced genome. Their approach was based on two separate models of central metabolism, which were treated as separate compartments that were connected by transfer reactions. They demonstrated that this method can be applied for modeling even more complex microbial communities. Zomorrodi and Maranas (2012) presented a multi-level optimization framework based on FBA for the metabolic modeling and analysis of microbial communities to describe trade-offs between individual vs. community level viability of bacteria.

Small scale models can offer detailed information over the studied subsystem. Application of genome scale models, however, can further increase the understanding of the metabolic system of a microorganism and to help to find new approaches for genetic engineering. The GEM of *E. coli* has been used and under active development over a decade and it has attained good predictions at aerobic conditions (Orth et al. 2011). In aerobic conditions, comparison between experimental and simulated result has been conducted, but less emphasis has been given to anaerobic conditions (Edwards et al. 2001). Therefore, its applicability to fermentative H_2 production was studied in Publication IV and V.

In Publication IV, genome-scale model of *E. coli* (Feist et al. 2007) was applied to predict which single gene deletions could be useful for enhancing the fermentative H_2 production. Alternatively, glucose or galactose was used as substrate. Biomass production was selected as an objective function. FBA was done separately to wild-type and all possible one gene deletions. Rates of H_2 production by mutants were compared to wild-type *E. coli* and based on the results, mutants to be tested in wet-lab were selected. According to FBA, the

utilization pathway of glucose and galactose differs only in the point of uptake and it is then degraded through same glycolytic routes. Therefore, based on the model, both have similar distribution of metabolites and growth rate. Anyhow, in *in vivo* studies the growth on galactose was much slower than growth on glucose and the response to gene deletion was different (Publication IV). Additionally, lactate production with galactose was notably lower than with glucose, even though acetate and ethanol production remained at the same level (unpublished data related to Publication IV). This has been observed earlier based on large-scale ^{13}C-flux analysis over aerobic culture of *E. coli.* It employs substantially different distribution of metabolites with glucose than galactose. Additionally, *E. coli* actively represses the uptake of galactose causing slower growth on galactose than on glucose (van Rijsewijk et al. 2011). This presents one issue of topological network models; they do not take into account the effect of internal regulation. For example, model applied in Publication IV does not contain information of genes that code regulatory enzymes. Therefore, those effects are not taken into account while simulating the H_2 production. Experimentally deletion of regulatory enzymes has been shown to increase the H_2 production (e.g., Fan et al. 2009). This and other issues of topological models are discussed in Section 3.5.

External fluxes (excreted products) can be given as constrains to FBA, enabling the estimation of internal fluxes based on measured data. Additionally, in order to unify the results of FBA and laboratory experiments, original constrains can be changed or new constraints can be added. E.g., McAnulty et al. (2012) analyzed GEM of *C. acetobutylicum* by FBA and used additional flux ratio constraints to increase the correlation of simulation to experimental data. In Publication V, *in vivo* and *in silico* data of mixed acid fermentation with glucose as substrate was critically compared and changes to GEM of *E. coli* were made to improve the simulated results. The main factor to be compared while analyzing the capabilities of the model is the biomass formation: If the model does not produce biomass with the given gene deletion, the deletion is considered to be lethal. Sometimes mutant is predicted to grow (i.e. mutation is non-lethal), but no growth is observed in laboratory experiments. That is often due to fact that based on model, several gene products can catalyze the same reaction. In reality, the activity of the enzymes is related to environmental conditions and all genes affecting to a reaction are not active simultaneously (i.e. aerobic vs. anaerobic conditions). Since in Publication V, anaerobic conditions were applied, changes to reaction directionalities and removal of genes mainly functioning in aerobic conditions were done based on literature survey and experimental data. Additionally, lower limit to lactate production was set to force its production during maximum growth. The modified model presented in Publication V was better suited for simulating the mixed acid fermentation in the conditions applied during the experiments in Publications III and IV than the original model. To demonstrate the improvement, both models, iAF1260 (Feist et al. 2007) and the modification of iAF1260 presented in Publication V, were used to simulate H_2 production using experimental results of acetate,

lactate and ethanol production as constrains and biomass production as objective function. The simulated and experimental results are shown in Figure 3.11.

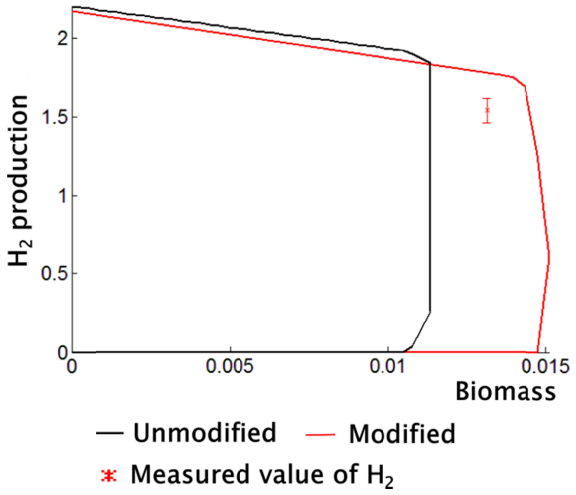


Figure 3.11. Byproduct secretion envelope of H_2. The area within drawn lines represent the allowable solution space for H_2 production based on FBA with the iAF1260 (Feist, 2007) (black line) and the modification of iAF1260 (Publication V) (red line). Experimental results of acetate, lactate and ethanol production are used as constrains and biomass production as objective function. The measured value of H_2 and biomass production fits better inside envelope created by modified model.

FBA has been applied for various situations, e.g., to gain more information over normal behavior of cell, to simulate effects of single and multiple gene deletions and gene additions, to estimate the capability of the cell to apply different substrates and to estimate end product excretions. In addition, FBA has been used to simulate internal fluxes based on measured data on excreted products in order to estimate the effects of environmental conditions, such as nutrients and pH, to the metabolism. It is a powerful tool to gain more understanding over the internal reactions of cell that are difficult to measure experimentally.

Identical cellular phenotype (i.e. same excreted end products) may be achieved by various combinations of internal fluxes, even though the FBA offers only one optimal solution. Thus one should be critical while analyzing the results. An example of this situation is given in Publication IV where, according to simulations, an alternative pathways to bypass some steps of glycolysis is used. Nevertheless, the same composition of end products results, when usage of this alternative pathway is blocked. This issue is taken into account when *metabolic pathway analysis* method is used to simulate the system behavior (Section 3.3.3). If enough measurement data exists to make system over determined, metabolic flux analysis (MFA) can be applied to analyze internal fluxes.

3.4.2 Metabolic flux analysis

Metabolic flux analysis is based on large amount of experimental data. In order to use MFA, number of measured reaction rates has to be at least the same as the number of

metabolites in the network. Solution to Eq. 3.2 is calculated by dividing the stoichiometric matrix S into two sub matrices: stoichiometric matrix S_m of reactions which rate has been measured and matrix S_u of reactions with unmeasured reaction rates. Corresponding rate vectors are $\mathbf{v}_m$ with measured reaction rates and $\mathbf{v}_u$ with unmeasured reaction rates. Based on measured values, rates for unmeasured fluxes can be calculated. Therefore, in MFA, an objective function is not used. As a result of analysis, only one optimal solution is gained. The basic concept of MFA is presented in Figure 3.12.

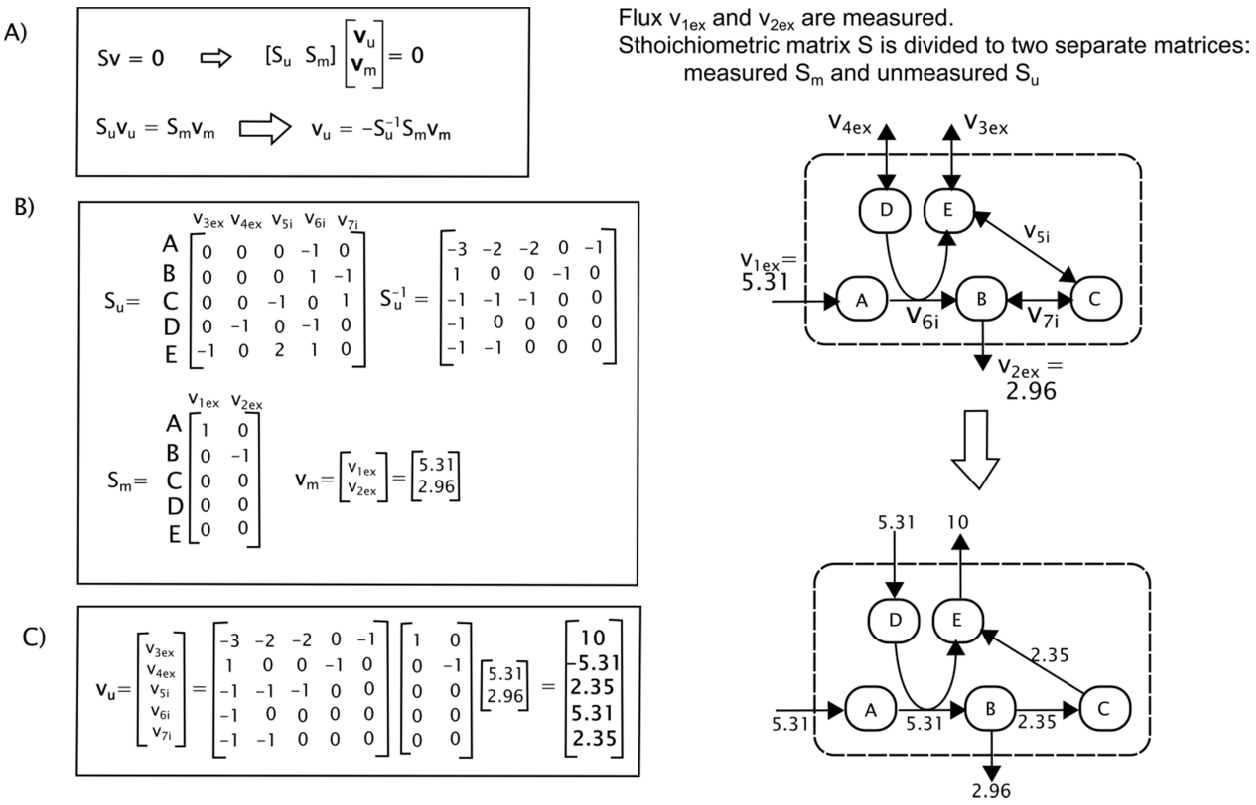


Figure 3.12. The basic method of metabolic flux analysis. A) The Equation 3.2 is divided into two sub matrices, measured (S_m) and unmeasured (S_u). The corresponding division is done within the rate vector *v* to $\mathbf{v}_m$ and $\mathbf{v}_u$. After the division is made, equation to calculate $\mathbf{v}_m$ can be given. B) The invert matrix S_u^{-1} of S_u and the measured matrices. C) Insertion of the matrices in B to equation in A, to solve the values for $\mathbf{v}_u$. Upper network presentation illustrates the initial network with two measured valued and the lower the end result of MFA.

In reality, metabolic networks are much larger than in Figure 3.12. Therefore, the amount of experimental data is often not enough to calculate the unmeasured fluxes. According to Wiechert (2001), the main problems of methods where the internal fluxes are simulated only based on reaction stoichiometry are (Figure 3.13): 1) Flux distribution of two parallel pathways or 2) of metabolic cycles can not be solved, if none of the reactions is coupled to measurable variable. 3) Simultaneously occurring bidirectional reaction steps can not be described and 4) correct calculations of how flow of metabolites split between pathways, demand correctly defined network. E.g. to combine subunits of pyruvate metabolism to

glycolysis and pentose phosphate pathway, the flow of energy metabolites such as ATP and NAHD should be exactly known. Due to the issues mentioned, measuring only uptake and excretion of metabolites is rarely enough.

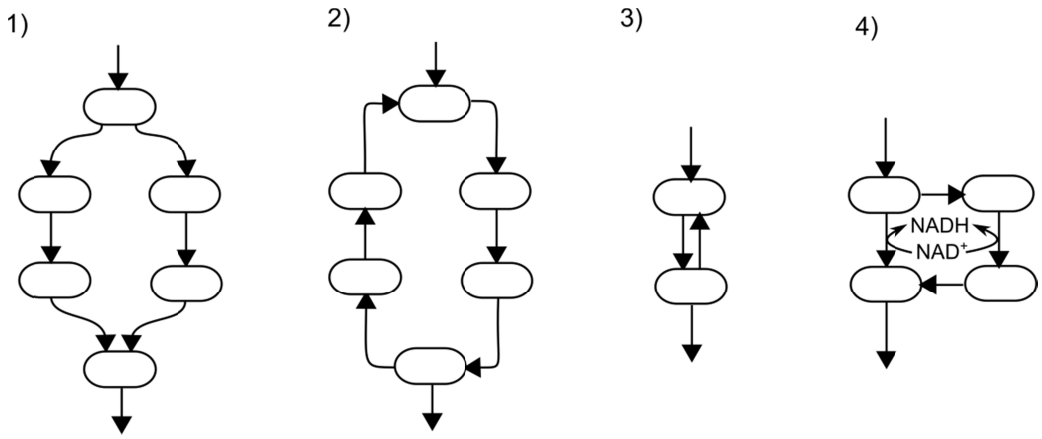


Figure 3.13. Situations that can not be solved, without measurements of internal fluxes. 1) Two parallel pathways to same end product, 2) cyclic reaction routes, 3) bidirectional reactions occurring to both directions at the same time, 4) correct split of fluxes between two pathways, if the cofactors of each reaction are not exactly known (Wiechert, 2001).

To overcome the issue of lacking of data for traditional MFA, carbon-13 MFA (^{13}CMFA) method has been developed to determine internal flux ratios by ^{13}C labeling (Szyperski et al. 1996, Fischer and Sauer 2003, Zamboni, 2011, Zamboni and Sauer, 2009). By ^{13}C labeling of the substrate, part of internal fluxes can be measured and enough data can be produced to find unambiguous solution of internal flux distribution (Bonarius et al. 1997, Fischer et al. 2004). However, the calculations are challenging and the *in vivo* experimentation is rather costly. Applications, such as FiatFlux, can be used to transform of carbon-labeling experiment data to flux rates within metabolic networks (Zamboni et al. 2005).

Metabolic fluxes within *E. coli* have been studied by ^{13}CMFA, e.g., to detect the effects of genetic and environmental modifications to central carbon metabolism, to study the differences in the respiratory and fermentative metabolism and to compare flux distributions with different nutrition sources (Fischer and Sauer, 2003, Sauer et al. 1999, van Rijsewijk et al. 2011). The gathered information of the flux distributions at different conditions can, e.g., guide the design of new strains or experimental setups to enhance fermentative H_2 production.

The term MFA is often used, even though the FBA with objective function has actually been applied. Chen et al. (2011) has compared the consistence of the results gained with ^{13}CMFA and FBA and noticed that the results differed significantly from each other. This supports the combinational usage of FBA and MFA to gain board and accurate view of internal fluxes. MFA presents a flux distribution that describes the cell metabolism at the given conditions. Thus, the measured values used for calculations change, when the growth

conditions, such as pH, are varied. In that case, the fluxes are re-measured and -calculated (Fischer and Sauer, 2003, Cai et al. 2010).

3.4.3 Metabolic pathway analysis

Metabolic pathway analysis can identify all metabolic flux vectors, i.e. sub networks, that exist in a metabolic network. No knowledge of flux rates or objective functions is needed. The group of possible sub networks form an admissible solution space (Trinh et al. 2009). Since the number of solutions by this method is infinite, constraints can be added to decrease the level of complexity. Here two related methods of metabolic pathway analysis with additional constraints are presented: elementary mode analysis (EMA) and extreme pathways (EP).

Elementary mode analysis is based on an idea that theoretical investigation should be started from studying structural properties of network, since kinetic properties, such as reaction rates, are subject to frequent changes (Chuster and Hilgetag, 1994). In the elementary mode analysis, the solutions are found from the admissible solution base defined by Eq. 3.2 by additional constrains of thermodynamics (Eq. 3.3) and non-decomposability constrain. In practice, this means that elementary mode (EM) includes all flux vectors ($\mathbf{em}_1$, $\mathbf{em}_2$…) that can be formed from the given network such that all the reactions within the sub-network are essential for it to be functional. Exact formulation of the EMA is reported by Chuster and Hilgetag (1994). An example of EMA is shown in Figure 3.14. In EMA, it is especially defined that if the sum of flux vectors, where the fluxes have opposite signs for same reactions, forms another flux vector, that is not excluded (Chuster and Hilgetag, 1994). In practice, this means that EMA counts bidirectional pathways as two separate pathways. On the contrary, extreme pathways have an additional condition, that none of the extreme pathways can be expressed as a sum of two or more other extreme pathways, thus such pathways are removed. Therefore, EPs are a subset of EM (Figure 3.9) (Klamt and Stelling, 2003).

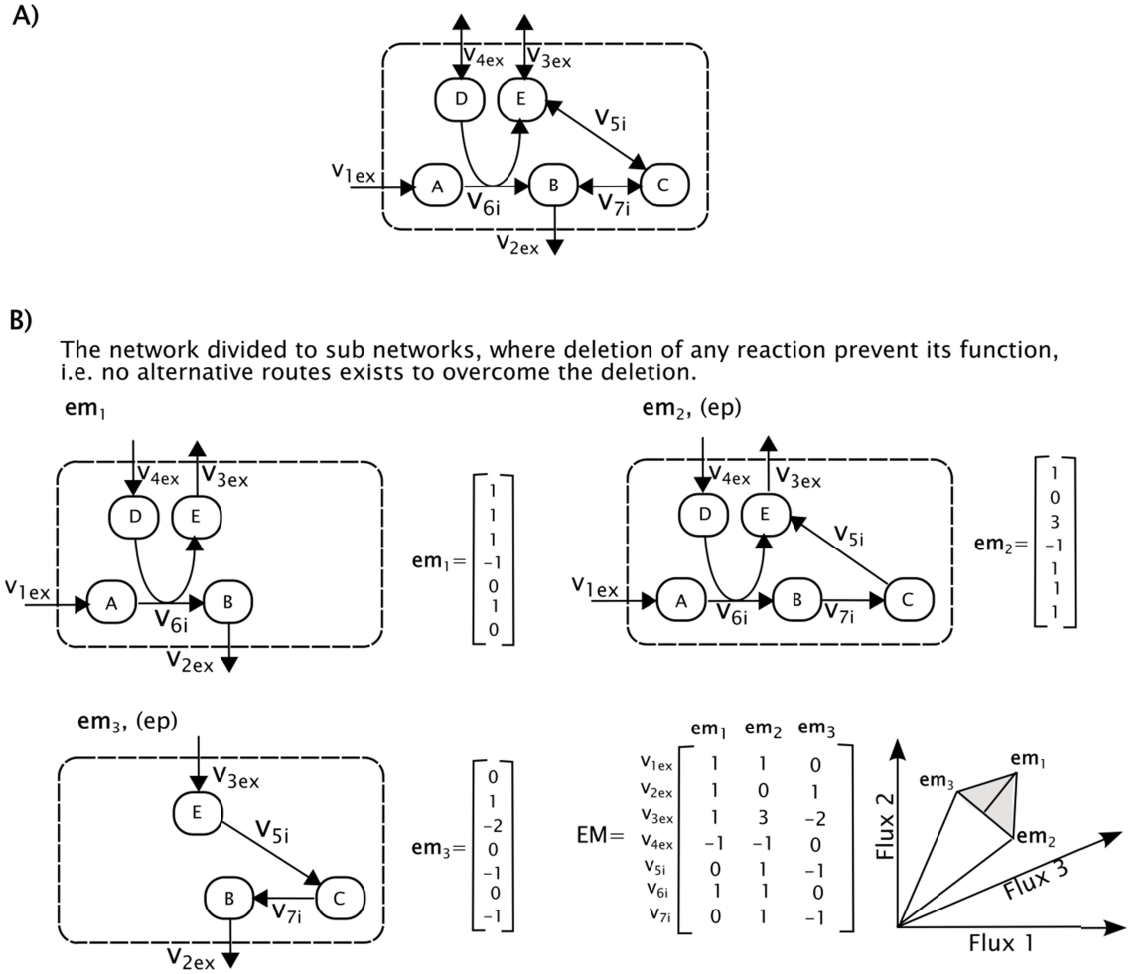


Figure 3.14. Basic concepts of elementary mode analysis (EMA). A) The studied network. B) The original network is divided to elementary modes, such that there is no accumulation of any product within the system boundaries (Eq. 3.2). All reactions that are irreversible in the original network have always positive sign (Eq. 3.3) and none of the fluxes within the elementary mode can be constrained to zero in order to the sub-network to be functional. The elementary modes create boundaries of allowable solutions space, drown as a cone. C) Properties of EMA: C1) With a complete set of the elementary modes, all metabolic flux vectors can be presented as a sum of two or more elementary modes (**em**) multiplied by positive scalar. E.g., the results gained by FBA (Fig. 3.10) can be presented as a sum of $\mathbf{em}_2$ and $\mathbf{em}_3$. C2) Decision whether a pathway is an EP or not, is made by comparing the pathways i.e. if one of the **em**s can be presented as a sum of two others, the **em** is not EP. Extreme pathways allow summary of pathways including same reactions with opposite sing. Therefore, $\mathbf{em}_1$ is EM but not EP. C3) EM matrix can, e.g., reveal pathways that are never used.

The applications of EMA and EP are related to fundamental understanding of network (such as network structure, robustness, regulation and fragility (Trinh et al. 2009)) and discovery of meaningful routes within metabolic networks (Stelling, et al. 2002). EMA has been applied, e.g., to find the most efficient pathways of *E. coli* for biomass and energy production (Carlson and Srienc, 2004). In addition to the exploring the capacities of existing pathways, it can be used for design and analysis of genetic engineering. Unrean et al. (2010) analyzed the carotenoid production network of recombinant *E. coli*. With the added carotenoid pathway, the simplified metabolic network resulted in 29532 different elementary modes. When those were divided to subsets based on utilization of O_2, the number of anaerobic subsets was 5377. This number was further decreased by restricting selected fluxed to zero, thus presenting a gene deletion. Similar approach has been applied for engineering *E. coli* to produce isobutanol (Trinh et al. 2011). The same could be applied to H_2 production for example, for designing mutants and understanding the changes caused to metabolic system by genetic engineering and environmental variations (e.g. aerobic/anaerobic, glucose/galactose as substrate).

Schilling et al. (2001) presented a combinatory use of EMA (or EP) and FBA. The simplified idea was that each possible elementary mode is subjected to FBA with same constrained input and objective function such as biomass. Then the maximum amounts of desired products from each EM can be compared, thus narrowing down the set of applicable EMs for further analysis. All applications of EMA presented in this section were conducted with small metabolic networks, and even then the number of possible solutions was high. The number of elementary modes grows in combinatorial fashion as the size of metabolic network increases, thus approaches to deal with this issue need to be developed (Rezola et al. 2011).

3.5 Future developments of metabolic modeling

Current metabolic models create rather straight forward presentations of the microbial metabolism and various approaches to apply those models exist. Applications of GEMs has been recently reviewed and divided to six groups: 1) metabolic engineering, 2) model-driven discovery, 3) prediction of cellular phenotypes, 4) analysis of biological network properties, 5) studies of evolutionary processes and 6) models of interspecies interactions (McCloskey et al, 2013). The currently used models are suitable for many applications. However, issues that restrain the accuracy of the modeling results remain. The correctness of analysis is greatly affected by the underlying metabolic model (Carlson and Srienc, 2004). Current models have issues such as the enzyme catalyzing the reaction is unknown or vice versa, or enzymes has unknown functions (Zamboni and Sauer, 2009). Presentation of the metabolism merely based on metabolite interactions, provides only a static description of potentially occurring interactions. In reality, activity of many pathways is

based on environmental conditions. Further, current models do not take into account translational regulation and detailed enzyme kinetics.

The actual situation inside bacterial cell is very complex sum of various factors. For example, during hydrogen production by *E. coli* (Figure 3.2.), the flux to formate is regulated by formate concentration, i.e. the metabolic system has feedback regulation. In addition, degradation of formate to H_2 and CO_2 is controlled at least by activation of FhlA and repression of HycA. The rate of reaction is regulated by the properties of reaction catalyzing enzyme complex, formate hydrogen lyase. At the genome level, transcription, translation and maturation of proteins have their own regulators, thus adding the system complexity. An example of the regulation of transcription of FdhF, that codes a subunit of formate hydrogen lyase, is illustrated in Figure 3.15.

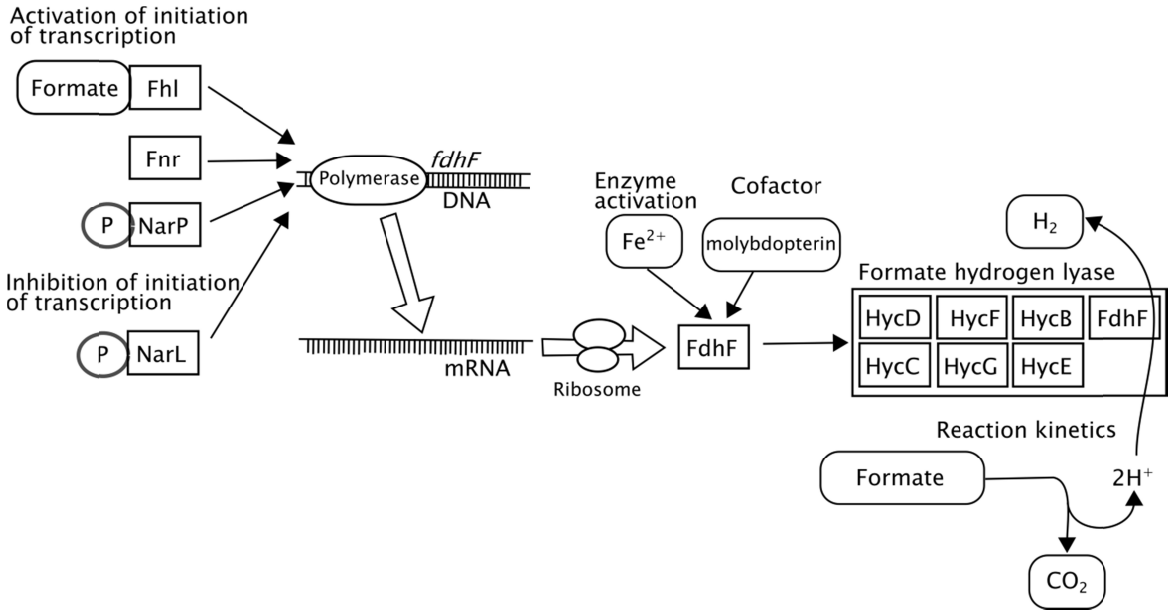


Figure 3.15. The regulation of *fdhF* synthesis. (Modified from Ecocyc, (Keseler et al. 2011))

In the future, more factors affecting to fluxes will be added to models. The machinery of transcription and translations has already been partly integrated to a genome-scale metabolic model of *E. coli* (Thiele et al. 2012). It has been done by integrating sequence information of 303 genes to the model and that is used to describe which metabolites are needed in order to translate a certain gene to an enzyme. Thus, in the model, enzymes take part to the reactions as reactants and are consumed for biomass production. However, the integration of many other biochemical processes, as shown in Figure 3.10, is still needed in order to gain full mechanistic (kinetic) presentation of microbial cell. If such model is created, higher computational accuracy and power will be needed (Thiele et al. 2012, Heinemann and Sauer, 2010). Karr et al. (2012) has already taken a step foward by creating whole-cell computational model for *Mycoplasma genitalium*, which has small

genome consisting of 525 genes. The goal of the model is to predict phenotype from genotype. In their approach, cellular functions are divided to 27 submodels. For example, metabolism is a submodel that is described by FBA. All submodels have a specific computational approach. The submodels are integrated to become a whole-cell model. The created model is described as a first draft that could be extended for more complex organisms which would allow experimental verification of the results (Karr et al. 2012).

4 Summary of publications and methods

Following the Publications included to the thesis and the method used are briefly summarized. Publications I and II, “Software for quantification of labeled bacteria from digital microscope images by automated image analysis” and “Efficient automated method for image-based classification of microbial cells” present approaches for automated analysis of digital microscopy images from bacterial cultures. Manual enumeration and morphological analysis of bacteria with microscope is laborious and depends on the experience of the examiner. Therefore, in Publication I, an easy to use automatic image analysis application, CellC, was developed for quantification of bacterial cells from digital microscope images. The user interface was designed such that the usage of software does not require vast knowledge of image processing. The software can analyze single images or compare two differently labeled images taken from the same position. If two images are used, the first image should include all the cells (e.g., DAPI stained) and the second a subgroup of cells in the first image (e.g., FISH stained cells). This improves the accuracy of enumeration of bacteria in the second image, since the location of bacteria is determined by the first image. The enumeration results of CellC were validated by comparing those to the results acquired by manual analysis. After the objects are segmented and enumerated from the original image, the software gives numerical estimates (such as cell width, length) over cell morphology and intensity of the light. CellC is currently available at https://sites.google.com/site/cellcsoftware/ and it has been applied to various purposes (Elazhairi-Ali et al. 2013, Gray et al. 2009, Harrold et al. 2011).

Publication II presents a framework that can be applied for division of bacteria to groups based on their morphological features. For the training and validation of the results, DAPI stained bacteria were viewed under microscope and digital images of cell groups with different sizes and shapes were taken. Separate images of pure cultures of *C. butyricum*, *E. coli* and *Staphylococcus* were used for training and testing the classifier. Purpose of staining was to achieve high threshold between the cells and background. Additionally, simulated images of bacteria were used for result validation. The presented framework could be used, e.g., for continuous monitoring of H_2 producing bioreactor with mixed culture. Changes in the community structure can be fast and cause drastic effects on the reactor operation (as seen in Koskinen et al. 2007). The automatic analysis of the population structure from microscopy images is a rapid method to observe upcoming changes and could be used to monitor the reactor system and allow quickly response by operator if needed. Additionally, the automatic classification can be used to analyze dynamic changes and interspecies interactions in mixed bacterial cultures or in artificial cocultures to gain more understanding over bacterial behavior.

Publication III was purely biological study of fermentative hydrogen production and is titled “Fermentative hydrogen production by *Clostridium butyricum* and *Escherichia coli* in pure and cocultures”. The study was motivated by the research of Koskinen et al. (2007) where the predominant bacterial species detected in H_2 producing enrichments cultures were closely related to strictly anaerobic *C. butyricum* and facultative anaerobic *E. coli*. It is known that *C. butyricum* is a good H_2 producer but the role of *E. coli* related to H_2 production in mixed culture was unclear. The uncertainty was due to the fact that wild-type *C. butyricum* can produce 4 mole of H_2 per mole of glucose whereas *E. coli* is capable of producing only 2 mole of H_2 per mole of glucose, thus it could have negative impact on the H_2 production capabilities of mixed culture.

The experimentation was done with medium scale batch experiments. That allowed the media volume to be large enough for frequent sampling and comparably large gas phase for maintaining the low partial pressure of hydrogen (pH_2) without gas extraction. The experimental set-up included six batch bottles with attached pH, pressure and temperature sensors that collected the data in one minute interval. The experimentation included parallel experiments with the pure culture of *E. coli* and *C. butyricum* and their coculture. The online monitoring of pressure enabled the observation of the exact moment for the initialization and end of the gas production. Frequent measurement of the gas composition combined with gas pressure data was used to calculate the H_2 production rates. Simultaneously with gas samples, also liquid samples were taken. Those were analyzed for production of biomass (OD_{600}), volatile fatty acids (VFA), ethanol and degradation of glucose.

Hydrogen production of *C. butyricum* was strongly associated with the growth, whereas *E. coli* produced H_2 long after the maximum biomass was reached. Additionally, *E. coli* continued the consumption of glucose after the end of exponential growth. The coculture was found to have lower yield of H_2 from glucose than the pure culture of *C. butyricum*. Nevertheless, the coculture was able to utilize the glucose better, i.e. there was less residual glucose in the media at the end of the experiment than in the pure culture experiments, resulting to higher total amount of H_2 produced. The research did not give clear answer to the initial question about the usefulness of the *E. coli* in the coculture. Anyhow, it showed that coculture of the species and competition over substrate can improve the efficiency of substrate utilization. The coculture of facultative anaerobic and strict anaerobic bacteria is useful in applications, were the creation of strictly anaerobic conditions is challenging; facultative anaerobes could be used to consume the remaining O_2 from the cultivation system allowing the growth of strictly anaerobic bacteria.

Publication IV was a combination of computational and biological methods, concentrated on studying the metabolism of *E. coli* under title, “Prospecting hydrogen production of *Escherichia coli* by metabolic network modeling”. The objective of this study was to apply genome scale metabolic model of *E. coli* (iAF1260) to predict, which single gene deletions

could be useful for enhancing the H_2 production with glucose or galactose as substrate. Simulated results were compared against wet lab experiments. Two different substrates were used in order to study the effect of mutations with alternative substrates. Additionally, the applicability of batch experimentation for screening large amount of samples and for validation of simulated results was evaluated.

The metabolic model was used with two approaches, both based on flux balance analysis. Biomass production was used as an objective function and either glucose or galactose as substrate. In the first selection method, gene deletions were simulated one by one, followed by comparison of the values of H_2 production to the ones of the wild-type. Those mutants that had higher H_2 production than wild-type were selected. The second approach was semiautomatic and based on knowledge of metabolic pathways, i.e. the pathways that utilized formate or pyruvate for other purposes that H_2 production were manually defined. Then, the analysis was done to find the single gene deletions that could block the usage of the defined pathways. The method was named ABCP (algorithm for blocking competing pathways). In addition, manual selection based on pathways maps and databases was done to choose mutants for batch experiments study. Almost all selected mutations were available in KEIO collection and altogether, 81 mutants and a wild-type of *E. coli* were cultivated. In order to evaluate the H_2 production capacity of the mutants, cultivations were done as parallel small scale batch experiments in sealed test tubes (10 ml liquid and total volume 27.5 ml). The cultivations were done in minimal M9 media with added casamino acids and glucose or galactose as substrate. Endpoint measurements of gas production volume, gas composition and biomass (OD_{600}) were done after two days of cultivation. Long cultivation time was due to the slow growth of *E. coli* on galactose.

Two factors were used to compare the model and experimental results: 1) higher or lower production of H_2 compared to the wild-type and 2) the essentiality of a gene (growth or no growth). Several gene deletions increasing the total H_2 production were found, but none of the selection methods was superior to other. Thus, based on the existing knowledge, the use of metabolic models is complementary approach used for planning of experiments and can not replace the traditional experimental design. Batch cultivation is a simple and straightforward experimental method to screen improvements in H_2 production. However, the ability of FBA to predict the H_2 production rates can not be evaluated by batch experiments with a single measurement point at the end of the experiment. Correctness of the gene essentiality predictions by FBA was rather good. Use of metabolic network models was found to be a good method for gaining broader understanding over the complicated metabolic system, and to be useful in prospecting suitable gene deletions for enhancing H_2 production. Many mutants were found that experimentally produced more H_2 compared to wild-type. More detailed experimentations should be done in order to verify the results. Differences between the effects of mutation, when either glucose or galactose was used as substrate, were found in wet lab experiments even though the FBA simulation

suggest that all mutants (expect with deletions in the glucose/galactose uptake reactions) have similar response to the mutation. This is due to the fact that the current metabolic models do not include translational regulation, detailed enzyme kinetics or metabolic feedback regulation.

Since in Publication IV, some inconsistent results between the estimation of the essentiality of genes were found, the data related to genes affecting in the main degradation pathways of glucose was used to modify the existing model to better describe the observed data. The work is presented in Publication V: "Modification of the *Escherichia coli* metabolic model iAF1260 based on anaerobic experiments". Since, the same model can be used in aerobic and anaerobic conditions by constraining intake of O_2 to zero, it is expected that some reactions that occur in aerobic conditions should not take place in the absence of oxygen. Therefore, most of the changes made were based on the removal of inactive genes or changing the reaction directionalities. All the changes were done after thorough study over those effects on the function of the model and based on knowledge gained from experimentation, literature and databases. The model results were compared to the metabolite data produced in Publications III and to essentiality and H_2 production data produced in publication IV. As a result, a model that better described the given conditions was achieved. The main materials and methods used in the Publications I-V are summarized in Table 4.1.

Table 4.1. Summary of materials and methods

Method /materials	**Analysis / source**	**Publication**
Bacterial cultures	*C. butyricum* (Isolated from Koskinen et al. (2007) by plating)	II, III
	E. coli K-12 MG1655	II, III, V
	E. coli K-12 BW25113	IV, V
	81 *E. coli* one gene deletion mutants (KEIO-collection)	IV, V
Substrates	Glucose	III, V
	Glucose and galactose	IV
Cultivation methods	Plating and incubation in anaerobic glove chamber	III
	Batch cultivations in 2100 ml anaerobic jars in 250 ml liquid media. Ports for manual sampling of gas and liquid and automatic sampling of pressure, temperature and pH.	III, IV
	Batch cultivations in 27.5 ml modified Hungate tubes having butyl rubber stopper and an aluminum seal finish in 10 ml liquid media.	IV, V
Cell staining	DAPI and FISH-staining	I, II
Cell morphology	Phase-contrast and epi-fluorescence microscopy with digital image acquisition.	I, II, III
Image analysis	MATLAB®	I, II
Gas composition	Gas chromatography (H_2, CO_2)	IV, V
Liquid compounds	High-performance liquid chromatography (VFA, ethanol, glucose and galactose)	III, IV, V
Biomass	Spectrophotometer (600 nm)	III, IV, V
Metabolic model	iAF1260 (Feist et al. 2007)	IV, V
Flux balance analysis	MATLAB®	IV
	MATLAB® , COBRA toolbox 2.0.5	V

5 Summary and conclusions

Microorganisms are optimal objects for systems biology studies, since they are easy to manipulate and have wide-ranging influence on many aspects of life. Rational engineering of microorganisms requires the ability to predict how the behavior of the studied organism originates from individual molecules and their interactions. Additionally, methods to implement the predicted mutations by bioengineering to achieve desired phenotypes are needed. Thus experts, who have both, computational and biological background, are needed.

Undefined mixed microbial cultures are commonly used for fermentative H_2 production. Microscopy can be used to visually examine the bacterial cultures and to reveal diversity and dominant bacterial species in the system. Manual analysis of the microscopy images is user dependent and laborious. Therefore, in Publications I and II, automatic image analysis methods were developed for monitoring changes in the culture compositions during fermentative H_2 production. In Publication I, an easy to used image analysis application, CellC, was introduced and its analysis results were validated. CellC can be applied to enumerate bacteria and measure cell length, width and intensity from digital images. It has already been applied by numerous users in experimental biology (Elazhairi-Ali et al. 2013, Gray et al. 2009, Harrold et al. 2011). In Publication II, a framework was presented for automated classification of microbial cells based on morphological features. These methods aids in the laborious analysis of microbial samples and creates opportunities for process control.

Roles of each bacterial species in mixed microbial cultures are not exactly known. Therefore, in Publication III, artificial bacterial coculture was composed and differences between coculture and pure culture were presented. Even though a pure culture of *C. butyricum* has higher yields of H_2 from glucose than the coculture of *E. coli* and *C. butyricum*, the total production of H_2 was higher with the coculture. Thus, when the goal is to efficiently degrade waste material and to minimize the residual substrate, cocultures should be applied. Additionally, facultative aerobic bacteria consume oxygen from the system, enabling the growth of strict anaerobic bacteria. On the other hand, if the goal is optimal yield of H_2 from a specific substrate, use of pure cultures is recommended.

Optimization of the environmental conditions for H_2 production is important. Anyhow, bacteria have intrinsic metabolic limitations of how much H_2 can be produced. To overcome these limitations, metabolic engineering has to be used to direct the flow of metabolites towards higher H_2 production. In Publication IV, the applicability of genome-scale metabolic model for simulating fermentative H_2 production of *E. coli* was verified.

Based on model prediction, various single gene deletions were found that experimentally increased fermentative H_2 production. Nevertheless, it is not possible to exceed the theoretical limits of fermentative H_2 production merely by gene deletions. Limit is for facultative anaerobic bacteria 2 and strictly anaerobic bacteria 4 mol of H_2 / mol of glucose (Nath and Das, 2004). Therefore new metabolic pathways need to be added to exceed these values. Those can be designed with the help of metabolic models and databases. The model used in Publication IV gave good description of the mixed acid fermentation of *E. coli*. Anyhow, the gene essentiality estimations had some inaccuracy. Therefore, in Publication V, the underlying metabolic model was modified to improve the simulation accuracy. Modifications were made based on data collected in Publications III and IV. The modified metabolic model was demonstrated to simulate the behavior of cells in the anaerobic conditions better than the original model.

Co-operation between experts of microbiology and computational sciences is crucial. When the co-operation is done from the point of planning of experiments, it can save time and money, and ease the combination of the biological data to the metabolic models. A microbiologist is able to describe what kind of possibilities there exist to produce the data, and computational experts can advise, how many samplings and which kind of samples should be taken to enable the comparison to existing models. Compromises from the both sides should be made. For example, the basic idea behind FBA is the assumption that there is no accumulation of intra cellular metabolites, the flow of nutrients is constant, pH is stable and the reactor is in steady state. Anyhow, this kind of experimentation demands the use of continuous reactors and can take days. For comparison, several parallel batch experiments can be done in a day. Therefore, the use of batch experiment and observing the rate data during the exponential growth phase is often applied to describe the steady state conditions and can be justified. If something interesting is found based on the compromised experimentation, more detailed experiments can be done later on.

A question that was often presented during the work was about the generalization of the results. An optimization of cultivation media is an example: if one have optimized growth conditions for wild-type bacteria, should the optimization be done again after one mutation? Another point of view is related to metabolic models: if one has made predictions over the metabolic fluxes within the cell in certain conditions and the pH of culture media is changed, should all the fluxes be recalculated. The correct answer to these questions is yes. But, these questions are always related to the level of accuracy of which one wants to study the given subject, and as already mentioned, compromises need to be made.

Metabolic models are generalizations over the metabolic system of studied bacteria. For example, genome-scale models include all the possible reactions that can take place within the studied microorganism based on knowledge of genome. In reality, only small portion of reactions are active at any conditions (Schuetz et al. 2012) A good example is the

difference between aerobic and anaerobic metabolism, where transition from one metabolism to other involves up and down regulation of large set of genes (Unrean et al. 2010). Additionally, the results of analysis that are not based on measured fluxes are always the same. Thus, the model has been created to reflect certain environmental conditions (e.g. pH and temperature), and those should be applied when the model is used without measured values as constrains.

Since the measurement methods are improving and the amount of data gained by single experimentation can be overwhelming, the significance of the co-operation will increase. The field of systems biology aims at integrating basic microbiological information with genomics, transcriptomics, metabolomics, proteomics and other data types, to create an integrated model of how a microbial cell or community functions. The aim is ambitious and current state of such models is still far from it. The currently used metabolic models contain information over the genome and the effect of the genes to the reactions within the metabolic network. Some advances in integration of, e.g., transcriptional regulation and signal transduction have been achieved (Thiele et al. 2012). Even though great progress has been made in the biochemical characterization of bacteria due to the development of high-throughput measurement technologies, much work still remains to be done. Computational methods, combining the biological data and mathematics, are needed to understand the complexity of biological systems.

This thesis presented both, biological and computational approach for the study of fermentative H_2 production. The work can be seen as a small step towards higher understanding of the complicated biological phenomena and as attempt to integrate the fields of biotechnology and computational systems biology towards the goal of enhanced hydrogen production.

Bibliography

Agapakis, C.M., Ducat, D.C., Boyle, P.M., Wintermute, E.H., Way, J.C., and Silver, P.A. (2010) Insulation of a synthetic hydrogen metabolism circuit in bacteria. *Journal of Biological Engineering*, 4:3.

Akhtar, M.K. and Jones, P.R. (2008a) Deletion of *iscR* stimulates recombinant clostridial Fe–Fe hydrogenase activity and H_2-accumulation in *Escherichia coli* BL21(DE3), *Applied Microbiology and Biotechnology*, vol. 78, no. 5, pp. 853–862.

Akhtar, M.K. and Jones, P.R. (2008b) Engineering of a synthetic *hydF-hydE-hydG-hydA* operon for biohydrogen production. *Analytical Biochemistry*, vol. 373, no. 1, pp. 170–172.

Akhtar, M.K. and Jones, P.R. (2009) Construction of a synthetic YdbK-dependent pyruvate: H_2 pathway in *Escherichia coli* BL21(DE3) *Metabolic Engineering*, vol. 11, no. 3, pp. 139-147.

Analytical Methods Committee (2013) Experimental design and optimisation (4): Plackett–Burman designs. *Analytical Methods*, vol. 5, no. 8, pp. 1901-1903.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2:2006.0008.

Bagramyan, K. and Trchounian, A. (2003) Structural and functional features of formate hydrogen lyase, an enzyme of mixed-acid fermentation from *Escherichia coli*. *Biochemistry (Mosc)*, vol, 68, no. 11, pp. 1159-1170.

Bagramyan, K., Mnatsakanyan, N., Poladian, A., Vassilian, A., and Trchounian, A. (2002) The roles of hydrogenases 3 and 4, and the F_0F_1-ATPase, in H_2 production by *Escherichia coli* at alkaline and acidic pH. *FEBS letters*, vol. 516, no. 1, pp. 172-178.

Barabási, A. L. (2009) Scale-free networks: a decade and beyond. *Science*, vol. 325, pp. 412-413.

Berks, B.C., Palmer, T. and Sargent, F. (2003) The Tat protein translocation pathway and its role in microbial physiology, Advances in Microbial Physiology, vol. 47, pp. 187-254.

Box, G. E. and Behnken, D. W. (1960) Some new three level designs for the study of quantitative variables. *Technometrics*, vol. 2, no. 4, pp. 455-475.

Box, G. E. and Wilson, K. B. (1951) On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol 13, no 1, pp. 1-45.

Plackett, R. L. and Burman, J. P. (1946) The design of optimum multifactorial experiments. *Biometrika*, vol. 33, no. 4, pp. 305-325.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., et al. (1997) The complete genome sequence of *Escherichia coli K-12*. *Science*, vol. 277, no. 5331, pp.1453–1474.

Bonarius, H. P., Schmid, G., and Tramper, J. (1997) Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends in Biotechnology*, vol. 15, no. 8, pp. 308-314.

Cai, G., Jin, B., Monis, P. and Saint, C., (2013) A genetic and metabolic approach to redirection of biochemical pathways of *Clostridium butyricum* for enhancing hydrogen production, *Biotechnology and Bioengineering*, vol. 110, no. 1, pp. 338-342.

Cai, G., Jin, B., Saint, C., and Monis, P. (2010) Metabolic flux analysis of hydrogen production network by *Clostridium butyricum* W5: Effect of pH and glucose concentrations. *International Journal of Hydrogen Energy*, vol. 35, no. 13, pp. 6681-6690.

Cai, G., Jin, B., Saint, C., and Monis, P. (2011) Genetic manipulation of butyrate formation pathways in *Clostridium butyricum*, Journal of biotechnology, vol. 155, no. 3, pp. 269-274.

Carlson, R., and Srienc, F. (2004) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. *Biotechnology and Bioengineering*, vol. *85*, no. 1, pp. 1-19.

Causey, T.B., Shanmugam, K.T., Yomano, L.P. and Ingram, L. O. (2004) Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. *Proceedings of the National Academy of Sciences USA*, vol. 101, no. 8, pp. 2235–2240.

Chaganti, S. R., Kim, D. H., and Lalman, J. A. (2011) Flux balance analysis of mixed anaerobic microbial communities: Effects of linoleic acid (LA) and pH on biohydrogen production. *International Journal of Hydrogen Energy*, vol. 36, no. 21, pp. 14141-14152.

Chang, J. J., Chou, C. H., Ho, C. Y., Chen, W. E., Lay, J. J., and Huang, C. C. (2008). Syntrophic co-culture of aerobic *Bacillus* and anaerobic *Clostridium* for bio-fuels and bio-hydrogen production. *International Journal of Hydrogen Energy*, vol. *33*, no. 19, pp. 5137-5146.

Chen, W.-M., Tseng, Z.-J., Lee, K.-S. and Chang, J.-S. (2005) Fermentative hydrogen production with *Clostridium butyricum* CGS5 isolated from anaerobic sewage sludge. *International Journal of Hydrogen Energy*, vol. 30, no. 10, pp. 1063–1070.

Chen, X., Alonso, A. P., Allen, D. K., Reed, J. L., and Shachar-Hill, Y. (2011) Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*. *Metabolic Engineering*, vol. 13, no. 1, pp. 38-48.

Chen, X., Sun, Y., Xiu, Z., Li, X., and Zhang, D. (2006) Stoichiometric analysis of biological hydrogen production by fermentative bacteria. *International Journal of Hydrogen Energy*, vol. 31, no. 4, pp. 539-549.

Chittibabu, G., Nath, K., and Das, D. (2006) Feasibility studies on the fermentative hydrogen production by recombinant *Escherichia coli* BL-21. *Process Biochemistry*, vol. 41, no. 3, pp. 682-688.

Chong, M. L., Rahman, A., Rahim, R. A., Aziz, S. A., Shirai, Y., and Hassan, M. A. (2009a) Optimization of biohydrogen production by *Clostridium butyricum* EB6 from palm oil mill effluent using response surface methodology. *International Journal of Hydrogen Energy*, vol. *34*, no. 17, pp. 7475-7482.

Chong, M. L., Sabaratnam, V., Shirai, Y., and Hassan, M. A. (2009b) Biohydrogen production from biomass and industrial wastes by dark fermentation. *International Journal of Hydrogen Energy*, vol. 34, no. 8, pp. 3277-3287.

Chou, C. H., Han, C. L., Chang, J. J., and Lay, J. J. (2011) Co-culture of *Clostridium beijerinckii* L9, *Clostridium butyricum* M1 and *Bacillus thermoamylovorans* B5 for converting yeast waste into hydrogen. *International Journal of Hydrogen Energy*, vol. 36, no. 21, pp. 13972-13983.

Crown, S. B., Indurthi, D. C., Ahn, W. S., Choi, J., Papoutsakis, E. T., and Antoniewicz, M. R. (2011) Resolving the TCA cycle and pentose-phosphate pathway of *Clostridium acetobutylicum* ATCC 824: Isotopomer analysis, in vitro activities and expression analysis. *Biotechnology journal*, vol. 6, no. 3, pp. 300-305.

Das D. and Veziroğlu T.N. (2001) Hydrogen production by biological processes: a survey of literature. *International Journal of Hydrogen Energy*, vol. 26, no. 1, pp. 13–28.

Das, D., Khanna, N., and Veziroğlu, N. T. (2008) Recent developments in biological hydrogen production processes. *Chemical Industry and Chemical Engineering Quarterly/CICEQ*, vol. 14, no. 2, pp.57-67.

Demirel, B., and Yenigün, O. (2006) Changes in microbial ecology in an anaerobic reactor. *Bioresource technology*, vol. 97, no. 10, pp. 1201-1208.

Dong, L., Zhenhong, Y., Yongming, S., Xiaoying, K. and Yu, Z. (2009). Hydrogen production characteristics of the organic fraction of municipal solid wastes by anaerobic mixed culture fermentation, *International Journal of Hydrogen Energy*, vol. 34, no. 2, pp. 812-820.

Dräger, A., Kronfeld, M., Ziller, M. J., Supper, J., Planatscher, H., Magnus, J. B., et al. (2009) Modeling metabolic networks in *C. glutamicum*: a comparison of rate laws in combination with various parameter optimization strategies. *BMC Systems Biology, 3*:5.

Durot, M., Bourguignon, P. Y., and Schachter, V. (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, vol. 33, no. 1, pp.164-190.

Edwards, J. S., Ibarra, R. U., and Palsson, B. O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, vol. *19*, no. 2, pp. 125-130.

Elazhari-Ali, A., Singh, A. K., Davenport, R. J., Head, I. M., and Werner, D. (2013) Biofuel components change the ecology of bacterial volatile petroleum hydrocarbon degradation in aerobic sandy soil. *Environmental Pollution*, vol. 173, pp. 125-132.

Fan, Z., Yuan, L., and Chatterjee, R. (2009). Increased hydrogen production by genetic engineering of *Escherichia coli*. *PloS one*, vol. 4, no. 2, e4432.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D. et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, *3*:121.

Feist, A. M., Herrgård, M. J, Thiele, I., Reed, J. L., and Palsson B. Ø. (2008) Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology,* vol. 7, no. 2, pp. 129-143.

Fischer, E., and Sauer, U. (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *European Journal of Biochemistry,* vol. 270, no. 5, pp. 880-891.

Fischer, E., Zamboni, N., and Sauer, U. (2004) High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived ^{13}C constraints. *Analytical Biochemistry,* vol. 325, no. 2, pp. 308-316.

Ghosh, D. and Hallenbeck, P.C. (2009) Fermentative hydrogen yields from different sugars by batch cultures of metabolically engineered *Escherichia coli* DJT135. *International Journal of Hydrogen Energy*, vol. 34, no. 19, pp. 7979-7982.

Ghosh, D., and Hallenbeck, P. C. (2010) Response surface methodology for process parameter optimization of hydrogen yield by the metabolically engineered strain *Escherichia coli* DJT135. *Bioresource Technology*, vol. 101, no. 6, pp. 1820-1825.

Gray, N. D., Sherry, A., Larter, S. R., Erdmann, M., Leyris, J., Liengen, T., et al. (2009) Biogenic methane production in formation waters from a large gas field in the North Sea. *Extremophiles*, *13*(3), 511-519.

Gírio, F. M., Fonseca, C., Carvalheiro, F., Duarte, L. C., Marques, S., & Bogel-Łukasik, R. (2010). Hemicelluloses for fuel ethanol: a review. *Bioresource Technology*, Vol. *101*, no. 13, pp. 4775-4800.

Guo, X. M., Trably, E., Latrille, E., Carrère, H. and Steyer, J.-P. (2010) Hydrogen production from agricultural waste by dark fermentation: A review. *International Journal of Hydrogen Energy*, vol. 35, no. 19, pp. 10660–10673.

Hamdan, A. A., Liebgott, P. P., Fourmond, V., Gutiérrez-Sanz, O., De Lacey, A. L., Infossi, P., et al. (2012) Relation between anaerobic inactivation and oxygen tolerance in a large series of NiFe hydrogenase mutants. *Proceedings of the National Academy of Sciences*, vol. 109, no. 49, pp. 19916-19921.

Harrold, Z. R., Hertel, M. R., and Gorman-Lewis, D. (2011) Optimizing *Bacillus subtilis* spore isolation and quantifying spore harvest purity. *Journal of microbiological methods*, vol. 87, no. 3, pp. 325-329.

Haverkorn van Rijsewijk, B. R. B., Nanchen, A., Nallet, S., Kleijn, R. J., and Sauer, U. (2011) Large-scale ^{13}C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Molecular Systems Biology*, *7*:477.

He, G. Q., Kong, Q., and Ding, L. X. (2004) Response surface methodology for optimizing the fermentation medium of *Clostridium butyricum*. *Letters in Applied Microbiology*, vol. 39, no. 4, pp. 363-368.

Heinemann, M., and Sauer, U. (2010) Systems biology of microbial metabolism. *Current Opinion in Microbiology*, vol. 13, no. 3, pp. 337-343.

Hillmann, F., Döring, C., Riebe, O., Ehrenreich, A., Fischer, R. J., and Bahl, H. (2009) The role of PerR in O_2-affected gene expression of *Clostridium acetobutylicum*. *Journal of bacteriology*, vol. 191, no. 19, pp. 6082-6093.

Hu, C.C., Giannis, A., Chen C.-L., Qi, W. and Wang, J.-Y. (2013) Comparative study of biohydrogen production by four dark fermentative bacteria. *International Journal of Hydrogen Energy*, In press, corrected proof available online 21 April 2013.

Infantes, D., González del Campo, A., Villaseñor, J., and Fernández, F. J. (2011) Influence of pH, temperature and volatile fatty acids on hydrogen production by acidogenic fermentation. *International Journal of Hydrogen Energy*, vol. *36*, no. 24, pp. 15595-15601.

Jamshidi, N., and Palsson, B. Ø. (2008) Formulating genome-scale kinetic models in the post-genome era. *Molecular Systems Biology*, *4*:171.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A. L. (2000) The large-scale organization of metabolic networks. *Nature*, vol. 407, pp. 651-654.

Jo, J. H., Lee, D. S., Park, D., and Park, J. M. (2008) Statistical optimization of key process variables for enhanced hydrogen production by newly isolated *Clostridium tyrobutyricum* JM1. *International Journal of Hydrogen Energy*, vol. *33*, no. 19, pp. 5176-5183.

Jo, J. H., Lee, D. S., Park, D., Choe, W. S., and Park, J. M. (2008) Optimization of key process variables for enhanced hydrogen production by *Enterobacter aerogenes* using statistical methods. *Bioresource Technology*, vol. 99, no. 6, pp. 2061-2066.

Karadag, D., and Puhakka, J. A. (2010) Effect of changing temperature on anaerobic hydrogen production and microbial community composition in an open-mixed culture bioreactor. *International Journal of Hydrogen Energy*, vol. *35*, no. 20, pp.10954-10959.

Kaksonen, A. H., Lavonen, L., Kuusenaho, M., Kolli, A., Närhi, H., Vestola, E., Puhakka, J. A. and Tuovinen, O. H. (2011) Bioleaching and recovery of metals from final slag waste of the copper smelting industry. *Minerals Engineering*, vol 24, no. 11, pp. 1113-1121.

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B, et al. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell*, vol. 150, no. 2, pp. 389-401.

Karube, I., Matsunaga, T., Tsuru, S. and Suzuki, S. (1976) Continuous hydrogen production by immobilized whole cells of *Clostridium butyricum*. *Biochimica et Biophysica Acta*, vol. 444, pp. 338-343

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñiz-Rascado, L., et al. (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research*, vol. 39(suppl 1), D583-D590.

Kim, S., Seol, E., Oh, Y.-K., Wang, G.Y. and Park, S. (2009) Hydrogen production and metabolic flux analysis of metabolically engineered *Escherichia coli* strains. *International Journal of Hydrogen Energy*, vol. 34, no. 17, pp. 7417–7427.

Klamt, S., and Stelling, J. (2003) Two approaches for metabolic pathway analysis? *TRENDS in Biotechnology*, vol. 21, no. 2, pp. 64-69.

Koskinen, P. E., Kaksonen, A. H., and Puhakka, J. A. (2007) The relationship between instability of H_2 production and compositions of bacterial communities within a dark fermentation fluidised-bed bioreactor. *Biotechnology and bioengineering*, vol. 97, no. 4, pp. 742-758.

Koskinen, P. E., Lay, C. H., Puhakka, J. A., Lin, P. J., Wu, S. Y., Örlygsson, J., & Lin, C. Y. (2008) High-efficiency hydrogen production by an anaerobic, thermophilic enrichment culture from an Icelandic hot spring. *Biotechnology and bioengineering*, vol. 101, no. 4, pp. 665-678.

Kwon, Y.D., Kwon, O.H., Lee, H.S. and Kim, P. (2007) The effect of NADP-dependent malic enzyme expression and anaerobic C4 metabolism in *Escherichia coli* compared with other anaplerotic enzymes. *Journal of Applied Microbiology*, vol. 103, no. 6, pp. 2340-2345.

Lehmussola, A. (2009) Image processing for systems biology - validation and performance evaluation through simulation. Tampereen yliopistopaino Oy, Tampere, Finland.

Levin, D.B., Pitt, L. and Love, M. (2004) Biohydrogen production: prospects and limitations to practical application. *International Journal of Hydrogen Energy*, vol. 29, no. 2, pp. 173–185.

Liebgott, P. P., Dementin, S., Léger, C., and Rousset, M. (2011) Towards engineering O_2-tolerance in [Ni–Fe] hydrogenases. *Energy & Environmental Science*, Vol. *4*, no. 1, pp. 33-41.

Maeda, T., Sanchez-Torres, V. and Wood, T.K. (2007) *Escherichia coli* hydrogenase 3 is a reversible enzyme possessing hydrogen uptake and synthesis activities. *Applied Microbiology and Biotechnology*, vol. 76, no. 5, pp. 1035-1042.

Maeda, T., Sanchez-Torres, V. and Wood, T.K. (2008) Metabolic engineering to enhance bacterial hydrogen production. *Microbial Biotechnology*, vol. 1, no. 1, pp. 30–39.

Mandenius, C. F., and Brundin, A. (2008) Bioprocess optimization using design-of-experiments methodology. *Biotechnology Progress*, vol. *24*, no. 6, pp. 1191-1203.

Manish, S., Venkatesh, K.V. and Banerjee, R. (2007) Metabolic flux analysis of biological hydrogen production by *Escherichia coli*. *International Journal of Hydrogen Energy*, vol. 32, no. 16, pp. 3820-3830.

Mason, D. J., Shanmuganathan, S., Mortimer, F. C., and Gant, V. A. (1998) A fluorescent Gram stain for flow cytometry and epifluorescence microscopy. *Applied and environmental microbiology*, vol. 64, no. 7, pp. 2681-2685.

Masset, J., Hiligsmann, S., Hamilton, C., Beckers, L., Franck, F., and Thonart, P. (2010) Effect of pH on glucose and starch fermentation in batch and sequenced-batch mode with a recently isolated strain of hydrogen-producing *Clostridium butyricum* CWBI1009. *International Journal of Hydrogen Energy*, vol. 35, no. 8, pp. 3371-3378.

Mathews, J., Li, Q. and Wang, G. (2010) Characterization of hydrogen production by engineered *Escherichia coli* strains using rich defined media. *Biotechnology and Bioprocess Engineering*, vol. 15, no. 4, pp. 686-695.

McAnulty, M. J., Yen, J. Y., Freedman, B. G., and Senger, R. S. (2012) Genome-scale modeling using flux ratio constraints to enable metabolic engineering of *clostridial* metabolism *in silico*. *BMC Systems Biology*, *6*:42.

McCloskey, D., Palsson, B. Ø., and Feist, A. M. (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Molecular Systems Biology*, *9*:661.

Mishra, J., Kumar, N., Ghosh, A. K., and Das, D. (2002) Isolation and molecular characterization of hydrogenase gene from a high rate of hydrogen-producing bacterial strain *Enterobacter cloacae* IIT-BT 08. *International Journal of Hydrogen Energy*, vol. *27*, no. 11, pp. 1475-1479.

Mohd Yusoff, M. Z., Maeda, T., Sanchez-Torres, V., Ogawa, H. I., Shirai, Y., Hassan, M. A., and Wood, T. K. (2012) Uncharacterized *Escherichia coli* proteins YdjA and YhjY are related to biohydrogen production. *International Journal of Hydrogen Energy*, vol. 37, no. 23, pp. 17778-17787.

Müller, V (2001) Bacterial fermentation. *John Wiley and Sons, Ltd*, Available: http://dx.doi.org/10.1002/9780470015902.a0001415.pub2.

Nakashimada, Y., Rachman, M. A., Kakizono, T., and Nishio, N. (2002) Hydrogen production of *Enterobacter aerogenes* altered by extracellular and intracellular redox states. *International Journal of Hydrogen Energy*, vol. 27, no. 11, pp. 1399-1405.

Nath, K. and Das, D. (2004) Improvement of fermentative hydrogen production: various approaches. *Applied Microbiology and Biotechnology*, vol. 65, no. 5, pp. 520-529.

Nath, K., and Das, D. (2011). Modeling and optimization of fermentative hydrogen production. *Bioresource Technology*, vol. *102*, no. 18, pp. 8569-8581.

Nissilä, M, Biohydrogen, bioelectricity and bioalcohols from cellulosic materials (2013) Suomen Yliopistopaino Oy,Tampere, Finland, TUT Publication 1103.

Nissilä, M. E., Tähti, H. P., Rintala, J. A., and Puhakka, J. A. (2011a). Effects of heat treatment on hydrogen production potential and microbial community of thermophilic compost enrichment cultures. *Bioresource technology*, vol. *102*, no 6, pp. 4501-4506.

Nissilä, M. E., Tähti, H. P., Rintala, J. A., and Puhakka, J. A. (2011b) Thermophilic hydrogen production from cellulose with rumen fluid enrichment cultures: Effects of different heat treatments. *International Journal of Hydrogen Energy*, vol. *36*, no. 2, pp. 1482-1490.

NIST/SEMATECH (2013) e-Handbook of Statistical Methods, Chapter 5: Prosess improvement. Available: http://www.itl.nist.gov/div898/handbook/ 5.9.2013.

Nnyepi, M.R., Peng, Y. and Broderick J.B. (2007) Inactivation of *E. coli* pyruvate formate-lyase: Role of AdhE and small molecules. *Archives of Biochemistry and Biophysics*, vol. 459, no. 1, pp. 1-9.

Ordal, E.J. and Halvorson, H.O. (1939) A comparison of hydrogen production from sugars and formic acid by normal and variant strains of *Escherichia coli*. *Journal of Bacteriology*, vol. 38, no. 2, pp. 199-220.

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., and Palsson, B. Ø. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*, *7*:535.

Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010) What is flux balance analysis? *Nature biotechnology*, vol. 28, no. 3, pp. 245-248.

Pan, C. M., Fan, Y. T., Xing, Y., Hou, H. W., and Zhang, M. L. (2008) Statistical optimization of process parameters on biohydrogen production from glucose by *Clostridium* sp. Fanp2. *Bioresource Technology*, vol. 99, no. 8, pp. 3146-3154.

Penfold, D.W. and Macaskie, L.E, (2004) Production of H_2 from sucrose by *Escherichia coli* strains carrying the pUR400 plasmid, which encodes invertase activity. Biotechnology Letters, vol 26, no. 24, pp. 1879-1883.

Penfold, D.W., Forster, C.F. and Macaskie, L.E. (2003) Increased hydrogen production by *Escherichia coli* strain HD701 in comparison with the wild-type parent strain MC4100. *Enzyme and Microbial Technology*, vol., 33, no. 2-3, pp. 185–189.

Penfold, D.W., Sargent, F. and Macaskie L.E. (2006) Inactivation of the *Escherichia coli* K-12 twin-arginine translocation system promotes increased hydrogen production, *FEMS Microbiology Letters*, vol. 262, no. 2, pp. 135–137.

Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012) Microbial engineering for the production of advanced biofuels. *Nature*, vol. *488*, pp. 320-328.

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M., and Mattanovich, D. (2011) Production of recombinant proteins and metabolites in yeasts. *Applied Microbiology and Biotechnology*, vol. 89, no. 4, pp. 939-948.

Quintana, N., Van der Kooy, F., Van de Rhee, M. D., Voshol, G. P., and Verpoorte, R. (2011) Renewable energy from *Cyanobacteria*: energy production optimization by metabolic pathway engineering. *Applied Microbiology and Biotechnology*, vol. 91, no. 3, pp. 471-490.

Raman, K., and Chandra, N. (2009) Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, vol. 10, no. 4, pp. 435-449.

Redwood, M.D., Mikheenko, I.P., Sargent, F. and Macaskie, L.E. (2008) Dissecting the roles of *Escherichia coli* hydrogenases in biohydrogen production, *FEMS Microbiology Letters*, vol. 278 no.1, pp. 48–55.

Rezola, A., de Figueiredo, L. F., Brock, M., Pey, J., Podhorski, A., Wittmann, C., et al. (2011) Exploring metabolic pathways in genome-scale networks via generating flux modes. *Bioinformatics*, vol. 27, no. 4, pp. 534-540.

Rosales-Colunga, L. M., Razo-Flores, E., Ordoñez, L. G., Alatriste-Mondragón, F. and De León-Rodríguez, A. (2010) Hydrogen production by *Escherichia coli* Δ*hycA* Δ*lacI* using cheese whey as substrate. *International Journal of Hydrogen Energy*, vol. 35, no. 2, pp. 491-499.

Rosales-Colunga, L.M., Razo-Flores, E. and De León Rodríguez, A. (2012) Fermentation of lactose and its constituent sugars by *Escherichia coli* WDHL: Impact on hydrogen production. *Bioresource Technology*, vol. 111, pp. 180–184.

Saint-Amans, S., Girbal, L., Andrade, J., Ahrens, K. and Soucaille, P. (2001) Regulation of carbon and electron flow in *Clostridium butyricum* VPI 3266 grown on glucose-glycerol mixtures. *Journal of Bacteriology*, vol. 183, no. 5, pp. 1748–1754.

Salmon, K., Hung S.P., Mekjian, K., Baldi, P., Hatfield, G.W. and Gunsalus, R.P. (2003) Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *The Journal of Biological Chemistry*, vol. 278, no. 32, pp.29837-29855.

Sanchez-Torres, V., Maeda, T., and Wood, T.K. (2009) Protein engineering of the transcriptional activator FhlA to enhance hydrogen production in *Escherichia coli*. *Applied and Environmental Microbiology*, vol. 75, no. 17, pp. 5639–5646.

Sauer, U. W. E., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., et al. (1999) Metabolic flux ratio analysis of genetic and environmental modulations of

Escherichia coli central carbon metabolism. *Journal of Bacteriology*, vol. 181, no. 21, pp. 6679-6688.

Sawers, R.G. (2005) Formate and its role in hydrogen production in *Escherichia coli*. *Biochemical Society Transactions*, vol. 33, pp. 42-46.

Schilling, C. H., Edwards, J. S., Letscher, D., and Palsson, B. Ø. (2000) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering*, vol. 71, no. 4, pp. 286-306.

Schleifer, K. H. (2009) Classification of *Bacteria* and *Archaea*: Past, present and future. *Systematic and Applied Microbiology*, vol. 32, no. 8, pp. 533-542.

Schrader, P. S., Burrows, E. H., and Ely, R. L. (2008) High-throughput screening assay for biological hydrogen production. *Analytical chemistry*, vol. 80, no. 11, pp. 4014-4019.

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012) Multidimensional optimality of microbial metabolism. *Science*, vol. 336, pp. 601-604.

Schuster, S., and Hilgetag, C. (1994) On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, vol. 2, no. 2, pp. 165-182.

Seibert, M., Flynn, T., Benson, D., Tracy, E., and Ghirardi, M. (1999) Development of selection and screening procedures for rapid identification of H_2 -producing algal mutants with increased O_2 tolerance. In *BioHydrogen*. Zaborsky, O.R., Benemann, J.R., Matsunaga, T., Miyake, J., and San Pietro, A. (eds) New York, NY, USA: Plenum Press, pp. 227–234.

Selembo, P. A., Perez, J. M., Lloyd, W. A. and Logan, B. E. (2009) Enhanced hydrogen and 1,3-propanediol production from glycerol by fermentation using mixed cultures. *Biotechnology and bioengineering*, vol. *104*, no. 6, pp. 1098-1106.

Self, W.T., Hasona, A. and Shanmugam, K. T. (2004) Expression and regulation of a silent operon, *hyf*, Coding for Hydrogenase 4 Isoenzyme in *Escherichia coli*. *Journal of Bacteriology*, vol. 186, no. 2, pp. 580–587.

Serres, M. H., Gopal, S., Nahum, L. A., Liang, P., Gaasterland, T., and Riley, M. (2001). A functional update of the *Escherichia coli* K-12 genome. *Genome Biology*, vol. 2, no. 9:research0035.1–0035.7

Shima, S., Pilak, O., Vogt, S., Schick, M., Stagni, M. S., Meyer-Klaucke, W, Warkentin, E., Thauer R. K., and Ermler, U. (2008) The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. *Science*, vol. 321, no. 5888, pp. 572-575.

Singh, S., Kang, S. H., Mulchandani, A., and Chen, W. (2008) Bioremediation: environmental clean-up through pathway engineering. *Current opinion in biotechnology*, vol. 19, no. 5, pp. 437-444.

Sode, K., Watanabe, M., Makimoto, H. and Tomiyama, M. (1999) Construction and characterization of fermentative lactate dehydrogenase *Escherichia coli* mutant and its potential for bacterial hydrogen production. *Applied Biochemistry and Biotechnology*, vol. 77, no. 1-3, pp. 317-323

Song, Z.-X., Dai, Y., Fan, Q.-L., Li, X.-H., Fan, Y.-T. and Hou, H.-W. (2012) Effects of pretreatment method of natural bacteria source on microbial community and bio-hydrogen production by dark fermentation. *International Journal of Hydrogen Energy*, vol. 37, no. 7, pp. 5631–5636.

Stelling, J. (2004) Mathematical models in microbial systems biology. *Current Opinion in Microbiology*, vol. 7, no. 5, pp. 513-518.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature*, vol. 420, pp. 190-193.

Steuer, R. (2011) Exploring the dynamics of large-scale biochemical networks: a computational perspective. *Open Bioinformatics Journal*, vol. 5, pp. 4-15.

Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., and Stahl, D. A. (2007) Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, *3*:92.

Subudhi, S., and Lal, B. (2011) Fermentative hydrogen production in recombinant *Escherichia coli* harboring a [FeFe]-hydrogenase gene isolated from *Clostridium butyricum*. *International Journal of Hydrogen Energy*, vol. 36, no. 21, pp. 14024-14030.

Suzuki, H., Kamatani, S., Kim, E.-S., and Kumagai, H. (2001) Aminopeptidases A, B, and N and dipeptidase D are the four cysteinylglycinases of *Escherichia coli* K-12. *Journal of Bacteriology*, vol. 183, no. 4, pp. 1489-1490.

Szyperski, T., Bailey, J. E., and Wüthrich, K. (1996) Detecting and dissecting metabolic fluxes using biosynthetic fractional ^{13}C labeling and two-dimensional NMR spectroscopy. *Trends in Biotechnology*, vol. 14, no. 12, pp. 453-459.

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, vol. 329, no. 5991, pp. 533-538.

Thiele, I., and Palsson, B. Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, vol. 5, no. 1, pp. 93-121.

Thiele, I., Fleming, R. M., Que, R., Bordbar, A., Diep, D., and Palsson, B. O. (2012) Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PloS One*, vol. *7*, no. 9, e45635.

Tosatto, S. C., Toppo, S., Carbonera, D., Giacometti, G. M., and Costantini, P. (2008) Comparative analysis of [FeFe] hydrogenase from *Thermotogales* indicates the molecular basis of resistance to oxygen inactivation. *International Journal of Hydrogen Energy*, vol. 33, no. 2, pp. 570-578.

Trinh, C. T., Li, J., Blanch, H. W., and Clark, D. S. (2011) Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. *Applied and Environmental Microbiology*, vol. 77, no. 14, pp. 4894-4904.

Trinh, C. T., Wlaschin, A., and Srienc, F. (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, vol. 81, no. 5, pp. 813-826.

Turcot, J., Bisaillon, A., Hallenbeck, P.C. (2008) Hydrogen production by continuous cultures of *Escherichia coli* under different nutrient regimes. *International Journal of Hydrogen Energy*, vol. 33, no. 5, pp. 1465-1470.

Unrean, P., Trinh, C. T., and Srienc, F. (2010) Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. *Metabolic Engineering*, vol. 12, no. 2, pp. 112-122.

Varma, A., and Palsson, B. O. (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/technology*, vol. 12, pp. 994-998.

Veit, A., Akhtar, M.K., Mizutani, T. and Jones, P.R. (2008) Constructing and testing the thermodynamic limits of synthetic NAD(P)H:H_2-pathways. *Microbial Biotechnology*, vol. 1, no. 5. pp. 382–394.

Vilar, J. M., Guet, C. C., and Leibler, S. (2003) Modeling network dynamics: the *lac* operon, a case study. *The Journal of Cell Biology*, vol. 161, no. 3, pp. 471-476.

Wagner, A., and Fell, D. A. (2001) The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, pp. 1803-1810.

Wang H. and Gunsalus R.P. (2003) Regulation of the *Escherichia coli* formate dehydrogenase *fdnGHI* and *fdhF* genes in response to nitrate, nitrite, and formate: roles for NarL and Nar. *Journal of Bacteriology*, vol. 185, no. 17, pp. 5076–5085.

Wang, J., and Wan, W. (2009) Factors influencing fermentative hydrogen production: a review. *International Journal of Hydrogen Energy*, vol. *34*, no. 2, pp. 799-811.

Wang, W., Li, G. W., Chen, C., Xie, X. S., and Zhuang, X. (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, vol. *333*, no. 6048, pp. 1445-1449.

Wells, M. A., Mercer, J., Mott, R. A., Pereira-Medrano, A. G., Burja, A. M., Radianingtyas, H., and Wright, P. C. (2011) Engineering a non-native hydrogen production pathway into *Escherichia coli* via a cyanobacterial [NiFe] hydrogenase. *Metabolic Engineering*, vol. 13, no. 4, pp. 445-453.

Wiechert, W. (2001) ^{13}C Metabolic Flux Analysis. *Metabolic engineering*, vol. 3, no. 3, pp. 195-206.

Wu, X., Lin, H., & Zhu, J. (2013). Optimization of continuous hydrogen production from co-fermenting molasses with liquid swine manure in an anaerobic sequencing batch reactor. *Bioresource technology*, vol, 136, pp. 351-359.

Yokoi, H., Maeda, Y., Hirose, J., Hayashi, S. and Takasaki, Y. (1997) H_2 production by immobilized cells of *Clostridium butyricum* on porous glass beads, *Biotechnology Techniques*, vol 11, no. 6, pp. 431-433.

Yoshida, A., Nishimura, T., Kawaguchi, H., Inui, M. and Yukawa, H. (2006) Enhanced hydrogen production from glucose using *ldh*- and *frd*-inactivated *Escherichia coli* strains. *Applied Microbiology and Biotechnology*, vol. 73, no. 1, pp. 67-72.

Yoshida, A., Nishimura, T., Kawaguchi, H., Inui, M. and Yukawa, H. (2005) Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing *Escherichia coli* strains. *Applied and Environmental Microbiology*, vol. 71, no. 11, pp. 6762–6768.

Zamboni, N. (2011) ^{13}C metabolic flux analysis in complex systems. *Current Opinion in Biotechnology*, vol. *22*, pp. 103-108.

Zamboni, N., and Sauer, U. (2009) Novel biological insights through metabolomics and ^{13}C-flux analysis. *Current opinion in microbiology*, vol. 12, no. 5, pp. 553-558.

Zamboni, N., Fischer, E., and Sauer, U. (2005) FiatFlux–a software for metabolic flux analysis from ^{13}C-glucose experiments. *BMC Bioinformatics*, *6*:209.

Zomorrodi, A. R., and Maranas, C. D. (2012) OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Computational Biology*, vol. 8, no. 2, e1002363.

Publications

Publication I

Selinummi, J., Seppälä, J., Yli-Harja, O and Puhakka, J.A. (2005) Software for quantification of labeled bacteria from digital microscope images by automated image analysis. *BioTechniques*, 39(6), 859-862.

Publication II

Ruusuvuori, P., Seppälä, J., Erkkilä, T., Lehmussola, A., Puhakka, J. A. and Yli-Harja, O. (2008) Efficient automated method for image-based classification of microbial cells. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008)*, Tampa, Florida, USA, December 7-11, 2008.

Publication III

Seppälä, J.J., Puhakka, J.A., Yli-Harja, O., Karp, M.T. and Santala, V. (2011) Fermentative hydrogen production by *Clostridium butyricum* and *Escherichia coli* in pure and cocultures. *International Journal of Hydrogen Energy*, 36, 10701-10708.

Publication IV

Seppälä, J.J.*, Larjo, A.*, Aho, T., Yli-Harja, O., Karp, M.T. and Santala, V. (2013) Prospecting hydrogen production of *Escherichia coli* by metabolic network modeling. *International Journal of Hydrogen Energy*, 38, 11780-11789. *Equal contribution

Publication V

Seppälä, J.J., Larjo, A., Aho, T., Kivistö, A., Karp, M.T. and Santala, V. (2013) Modification of the *Escherichia coli* metabolic model iAF1260 based on anaerobic experiments. In *Proceedings of the 10th TICSP Workshop on Computational Systems Biology (WCSB 2013)*, Tampere, Finland, June 10-12, 2013, pp. 79-85.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3227-6
ISSN 1459-2045