Biological hydrogen formation by thermophilic bacteria

Abraham A.M. Bielen

Thesis committee

Promotors
Prof. Dr J. van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University

Prof. Dr W.M. de Vos
Professor of Microbiology
Wageningen University

Co-promotor
Dr S.W.M. Kengen
Assistant professor, Laboratory of Microbiology
Wageningen University

Other members
Prof. Dr G. Eggink, Wageningen University
Prof. Dr W.R. Hagen, Delft University of Technology
Dr T.J.G. Ettema, Uppsala University, Sweden
Dr R. van Kranenburg, Corbion, Gorinchem

This research was conducted under the auspices of the Graduate School SENSE (Netherlands Research School for the Socio-Economic and Natural Sciences of the Environment)

Biological hydrogen formation by thermophilic bacteria

Abraham A.M. Bielen

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 10 June 2014
at 4 p.m. in the Aula.

Abraham A.M. Bielen
Biological hydrogen formation by thermophilic bacteria
234 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)
With references, with summaries in Dutch and English

ISBN 978-94-6173-936-0

Table of Contents

Chapter 1

General introduction and thesis outline

Abstract

A growing awareness of the negative aspects associated with traditional production processes for fuels and chemicals has resulted in an increase in research on the development of alternative more sustainable processes. This includes the production processes for hydrogen gas (H_2) a very important industrial commodity with indispensable roles in many processes. Moreover, H_2 could be used as an energy storage/transportation medium and is directly applicable as a fuel. In a biological H_2 (bioH_2) production process, known as dark fermentation, fermentative microorganisms are able to generate H_2 from renewable resources like carbohydrate-rich plant material or industrial waste streams. Especially thermophilic microorganisms possess several desirable traits making them attractive candidates to implement in a bioH_2 production process. The thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima* have the hydrolytic capacity to decompose complex biomass in readily fermentable di- or mono saccharides, moreover, their catabolic pathways lead only to a limited number of possible fermentation end-products. Under ideal conditions these thermophiles are able to ferment sugar substrates to H_2 with yields close to the theoretical maximum of 4 H_2/hexose. The investigation of the underlying mechanisms of the H_2 generating processes including i) the routes of carbohydrate metabolism, ii) reductant recycling and iii) the effect of cultivation conditions on performance, allows us to achieve a better understanding of the limiting factors of current bioH_2 production processes, and thereby bringing the large-scale industrial application of dark fermentation a step closer.

1. Hydrogen gas : Why do we need it and how can it be made?

Since its discovery in the 18th century [1], molecular hydrogen (H_2) has become an important chemical commodity. With an estimated annual worldwide production exceeding 50 million tons, H_2 finds its application in numerous industrial processes. Particularly, H_2 is used as a bulk chemical in the glass, food, electronics and metal-processing industries, in petroleum refining processes and during the synthesis of basic chemicals like methanol and ammonia [2, 3]. H_2 can also be used as a fuel e.g. for commercial transport or space-flight. The energy associated with the oxidative conversion of H_2 to water (H_2O) (**Eq. 1.1**) generates heat in a combustion engine or electrical current in H_2-fuelled fuel cells, where the latter process is considered to be more efficient. Moreover, H_2 can be implemented as an energy storage and transportation medium, and for this reason it has been proposed as the central chemical energy carrier in a future hydrogen energy economy [3].

(**Eq. 1.1**) $2\ H_2(g) + O_2(g) \longrightarrow 2\ H_2O(l)$
-474 kJ/reaction ($\Delta_r G^\circ$)[4]

In nature, molecular hydrogen is scarce, atmospheric H_2 concentrations are low (0.000055%) and, contrary to natural gas (methane, CH_4), there are no natural accumulation sites of H_2 gas. Hydrogen is commonly bound in other compounds like water, hydrocarbons (C_nH_m) or carbohydrates ($C_n(H_2O)_m$). To free hydrogen from these compounds several methods are available [5], including i) electrolytic [6], ii) photoelectrochemical [7], iii) thermal [8], and (iv) biological processes [9-11] (**Figure 1.1**). H_2 formation via electrolytic processes (electrolysis) requires an energy conversion from a primary energy carrier to electricity, which provides the current that drives the decomposition of water into O_2 and H_2, i.e. the reverse reaction of the reaction displayed in **Eq. 1.1** [6]. During photoelectrochemical processes solar energy (photons) directly assists in generating a current that drives the electrolysis of water [7]. Thermochemical methods which employ nuclear or solar energy use generated heat to drive the thermolysis of water [8], whereas thermal processes like steam reforming (**Eq. 1.2**), hydrocarbon partial oxidation (**Eq. 1.3**), hydrocarbon decomposition (**Eq. 1.4**) and coal gasification (**Eq. 1.5**) require fossil fuels and water as starting material alongside an input of heat to drive these reactions [8]. Hydrogen formation via reaction 2, 3, and 5 generates a gas mixture containing CO and H_2, also known as syngas. By coupling syngas production with a water-gas shift reaction (**Eq. 1.6**), the overall H_2 yield of these processes is further enhanced. The use of fossil (hydro)carbons as a source for H_2 production is inevitably coupled with the emission the greenhouse gas carbon dioxide (CO_2). Gasification of renewable biomass-derived carbohydrates (**Eq. 1.7**) is also coupled to CO_2 emission, however, the carbon

retention time in renewable biomass is much smaller than in fossil carbon compounds. Because the CO_2 fixed from the atmosphere during biomass formation is released on a relative short timescale (generally less than several years) compared to geological time scales for fossil fuels, the use of renewable biomass will not lead to a net increase in atmospheric CO_2 concentrations. Biological processes implement biological systems, i.e. microorganisms like algae or bacteria, for H_2 production. These microorganisms generate H_2 from water or biomass-derived carbon compounds, using either solar energy, electricity and/or biomass as an energy source [9-12].

(**Eq. 1.2**) $C_nH_m + n\,H_2O \leftrightarrows n\,CO + (m + 2n)/_2\ H_2$

(**Eq. 1.3**) $C_nH_m + {}^{n}/_{2}\ O_2 \longrightarrow n\,CO + {}^{m}/_{2}\ H_2$

(**Eq. 1.4**) $C_nH_m \longrightarrow n\,C + {}^{1}/_{2}\,m\,H_2$

(**Eq. 1.5**) $C + H_2O \leftrightarrows CO + H_2$

(**Eq. 1.6**) $CO + H_2O \leftrightarrows CO_2 + H_2$

(**Eq. 1.7**) $C_n(H_2O)_m + p\,O_2 + q\,H_2O \dashrightarrow x\,CO + y\,CO_2 + z\,H_2$

In all described H_2 production processes an input of energy is required to release hydrogen from the basic starting materials. This energy is either provided directly by a primary energy carrier or indirectly via electricity. In this perspective H_2 can also be considered a secondary energy carrier (**Figure 1.1**). Based on the negative effects associated with the use of nuclear and fossil resources, like their impact on the environment (e.g. radioactive waste, greenhouse gas emissions), related geopolitical issues and the non-renewable nature of these resources, the production of H_2 via their related processes are presumed to be non-sustainable. An entire H_2 production process is only considered sustainable when both the energy sources and the basic starting materials for H_2 formation are renewable. Hence H_2 is only considered a sustainable commodity when it is generated using renewable resources like solar-, wind-, water- and geothermal-energy or biomass. Currently 94% of the global H_2 production is based on fossil fuel [8]. To be able to increase the share of H_2 production processes based on renewable resources, many challenges have to be overcome and technological advances have to be made.

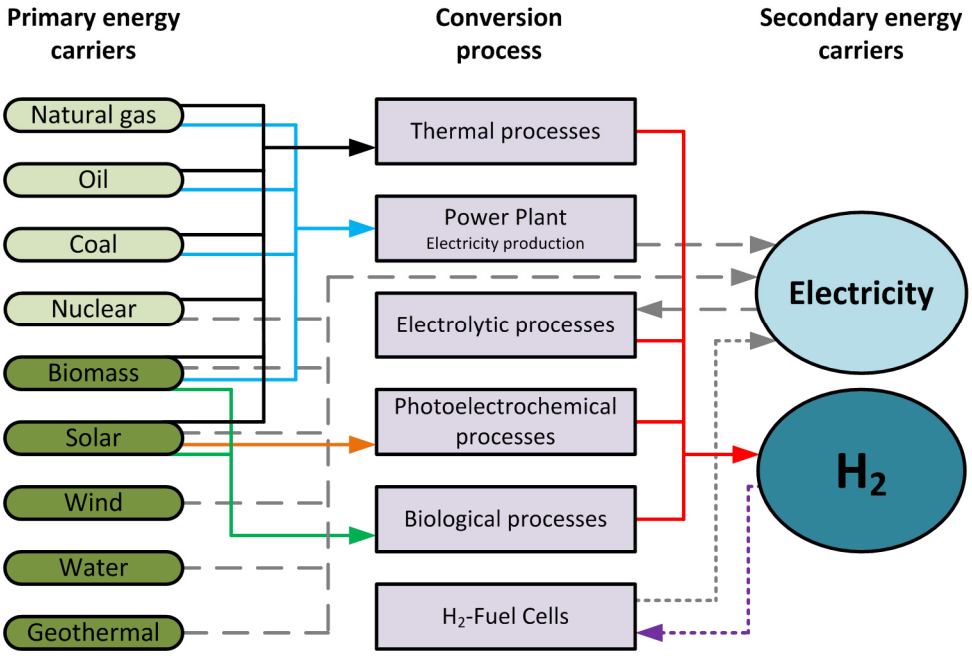


Figure 1.1 Overview of the conversion processes involved in the formation of molecular hydrogen (H_2) form renewable (dark green boxes) and non-renewable (light green boxes) primary energy carriers. The interconversion between the secondary energy carriers electricity and H_2 is achievable via electrolytic processes (electrolysis) and H_2 fuelled fuel cells.

2. Biological hydrogen production: What are the alternatives?

Biological H_2 formation processes can be classified into four major categories: (i) bio-photolytic, (ii) photo-fermentative, (iii) bio-electrochemical, and (iv) fermentative processes [9-12]. The formation of H_2 by such biological systems revolves around a seemingly simple reaction: two protons (H^+) and two electrons (e^-) brought together form H_2 (**Eq. 1.8**). Except for bio-electrochemical H_2 formation processes this reaction is catalysed by specific enzymes. The electrons involved in the catalysed reactions are generally associated to electron carriers, like NAD^+ or ferredoxin, which are able to accept electrons from a donor and transfer them to the hydrogen generating enzymes (e.g. hydrogenases or nitrogenases).

Bio-photolytic microorganisms like algae or cyanobacteria capture incoming solar energy, leading to the decomposition of water into O_2, protons (H^+) and electrons (e^-) (**Eq. 1.9**). These electrons, which are shuttled via the electron transport chain producing energy required for growth, are finally used by hydrogenases (**Eq. 1.8**) or by nitrogenases (**Eq. 1.10**). In nitrogenase catalysed reactions H_2 is actually formed as a side-product of the energy consuming N_2 fixation reaction. In photo-fermentative

organisms, like purple non-sulfur bacteria, the electrons required for H_2 formation during N_2 fixation (**Eq. 1.10**) are provided via the metabolism of biomass-derived organic acids (e.g. acetate), while the additional energy for growth is obtained via their light-driven proton pumps. In electrochemically assisted microbial H_2 production, the free electrons generated during the metabolism of organic matter are transferred to the anode electrode. Under non-augmented conditions the generated potential is too low to drive the H_2 formation reaction (**Eq. 1.8**); however, by supplementing the generated potential with an external power source, H_2 formation becomes possible. Fermentative hydrogen production, often referred to as dark fermentation, to distinguish it from the light driven photo-fermentation, involves the metabolism of biomass-derived carbohydrates (e.g. sugars like glucose) to organic acids. Electrons generated throughout metabolism are transferred to hydrogenases, subsequently leading to H_2 formation.

(**Eq. 1.8**) $2\ H^+ + 2\ e^- \leftrightarrows H_2$

(**Eq. 1.9**) $H_2O \leftrightarrows {}^1/_2\ O_2 + 2\ H^+ + 2\ e^-$

(**Eq. 1.10**) $N_2 + 8\ H^+ + 8\ e^- + 16\ ATP + 16\ H_2O \leftrightarrows 2\ NH_3 + H_2 + 16\ ADP + 16\ P_i$

(**Eq. 1.11**) $\text{Glucose} + 12\ H_2O \rightarrow 6\ HCO_3^- + 6\ H^+ + 12\ H_2$
+3.2 kJ/reaction ($\Delta_r G^{\circ\prime}$)

(**Eq. 1.12**) $\text{Glucose} + 4\ H_2O \rightarrow 2\ \text{acetate}^- + 2\ HCO_3^- + 4\ H^+ + 4\ H_2$
-206.3 kJ/reaction ($\Delta_r G^{\circ\prime}$)

(**Eq. 1.13**) $2\ \text{acetate}^- + 2\ H^+ + 8\ H_2O \rightarrow 4\ HCO_3^- + 4\ H^+ + 8\ H_2$
+209.2 kJ/reaction ($\Delta_r G^{\circ\prime}$)

Under standard conditions, the complete oxidation of the common biomass-derived carbohydrate glucose ($C_6H_{12}O_6$) to H_2 and CO_2 (**Eq. 1.11**) requires an additional input of energy, which is reflected by the positive reaction Gibbs energy ($\Delta_r G^{\circ\prime}$) associated to this reaction. In nature no single organism exists that is capable of catalysing this conversion. Due to metabolic and thermodynamic constraints, organisms performing dark fermentation can only achieve a theoretical maximal H_2 yield of 4 moles of H_2 produced per mole of glucose consumed (**Eq. 1.12**) [4]. In this situation two moles of acetate ($C_2H_3O_2^-$) and bicarbonate (HCO_3^-) are concomitantly formed as fermentation end-products. The surplus of energy generated during this conversion is used by the fermentative organisms for cell maintenance and growth. A subsequent conversion of

the fermentation end-product acetate to H_2 and CO_2 requires the input of additional energy (**Eq. 1.13**). As discussed, photo-fermentative organisms can overcome this energy barrier by utilizing solar light, while in bio-electrochemical processes the additional energy input can be provided by an externally applied potential. So by coupling dark fermentation with either photo-fermentative or bio-electrochemical processes, the theoretical maximal H_2 yield of 12 moles of H_2 produced per mole of glucose might be accomplished.

With the exception of bio-photolysis, biological H_2 formation processes require an input of biodegradable organic matter, e.g. carbohydrates like sugars or organic acids. Preferably, the utilized renewable biomass is not in competition with food production. This has been a major point of criticism with respect to the first generation of bio-commodities that are produced primarily from food crops such as grains, sugar cane or vegetable oils. Alternatively, second-generation bio-commodities are produced from sustainable feedstocks, including non-food crops and inedible waste products like argo/food/biofuel waste residues. These substrates can also be implemented in the described bio-electrochemical, photofermentative and fermentative processes. Usually some form of biomass pre-treatment is required prior to their application as substrates for biohydrogen production, like the disposal of inhibiting/toxic component or the breakdown of complex plant derived polymeric carbohydrates into readily fermentable sugars. Although the latter can be achieved by different (thermo)chemical or enzymatic pre-treatments, a more desirable process combines both the breakdown and fermentation of complex plant biomass. Such a bioprocess circumvents the negative environmental impact inherent to (thermo)chemical pre-treatment and might limit overall process costs.

3. Dark fermentation: Mechanisms of reductant disposal and why thermophiles are preferred for bio-H_2 production

Dark fermentation involves the breakdown of sugar compounds like glucose to organic acids or alcohols (**Figure 1.2**). During intermediate catabolic steps reducing equivalents are generated, e.g. two electrons are released during the formation of pyruvate from glyceraldehyde-3P. Released electrons are typically transferred to electron carriers like NAD^+ (generating NADH) or ferredoxin (Fd_{ox}) (generating reduced ferredoxin (Fd_{red})). These electron-carriers have to be recycled to maintain the glycolytic flux which provides the energy required for growth and maintenance; otherwise cell growth will hamper or completely stop. In the absence of external electron acceptors, which is intrinsic to fermentative processes, catabolic intermediates can be used as electron acceptors.

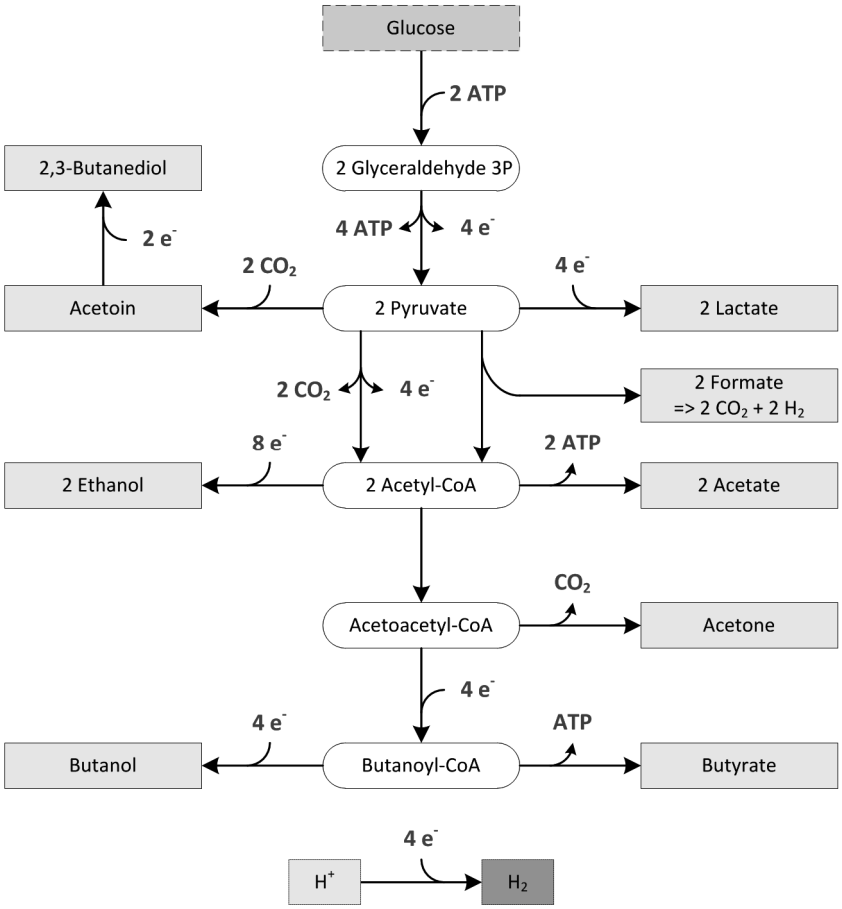


Figure 1.2 Possible routes of glucose fermentation leading to common fermentation end-products. During fermentation electrons are released (e^-, which are typically transferred to electron carriers) that can be consumed during the formation of reduced end-products like lactate or alcohols. Excess of produced reducing equivalents can be consumed in a H_2 forming reaction (bottom of the scheme).

Alternatively, the electron-carriers can be re-oxidized (recycled) in H_2 generating reactions, where protons are used as final electron acceptors (**Eq. 1.8, Figure 1.2**). Additionally, during fermentative growth oxidative phosphorylation, which acts as an electron sink and ATP source, is absent and cells mainly depend on substrate level phosphorylation for their ATP (energy) generation. Overall, carbon flow, reductant recycling and ATP generation are tightly coupled. For example, complete fermentation of glucose to two lactate does not yield any H_2 since all electrons generated during the conversion of glyceraldehyde-3P to pyruvate are consumed in the production of two

lactate from two pyruvate; in addition, one ATP is produced per lactate (**Figure 1.2**). Likewise, no H_2 is produced when ethanol is the sole end-product. Alternatively, complete oxidation of glucose to acetate yields four H_2 per glucose consumed and generates two ATP per acetate produced.

In general, the relative proportions of fermentation end-products are balanced to maximize ATP production while at the same time securing the re-oxidation of generated electron carriers. The metabolic capabilities of organisms are determined by their genomic content (whether a specific enzyme-encoding gene is present or absent), and the environmental (cultivation) conditions.

To produce the maximum amount of H_2 fermentation should not result in branched pathways leading to other end-products like lactate, ethanol or butyrate, that are not coupled to H_2 formation **Figure 1.2**. Biomass-derived carbohydrates need to be oxidized entirely to acetate (**Eq. 1.12**) while, concomitantly, all the reduced electron carriers (e.g. NADH, Fd_{red}) are recycled via H_2 formation. **Table 1.1** shows an overview of some anaerobic/facultative-anaerobic H_2 producing organisms with different optimal growth temperatures and their experimentally observed H_2-yields. Compared to mesophiles (optimal growth temperature between 20-45°C), thermophilic organisms (optimal growth temperature above 50°C) display a lower diversity in fermentation end-products and a higher H_2 yield per hexose. Especially for some of the extremely thermophilic and hyperthermophilic organisms, yields reach the theoretical maximum of 4 H_2 per hexose.

The midpoint redox potential of the electron carrier couples NAD^+/NADH and Fd_{ox}/Fd_{red} under standard conditions are -320 mV and -398 mV, respectively [4]. Since the midpoint redox potential of the H^+/H_2 couple is -414 mV it follows that under standard condition H_2 formation from either NADH or Fd_{red} is thermodynamically unfavourable (**Table 1.2**), although, under physiological conditions the midpoint redox potential of especially Fd_{red} is somewhat lower indicating that under those conditions H_2 formation from Fd_{red} might be thermodynamically feasible. However, several fermentation studies have revealed that some organisms can even overcome the thermodynamic barrier associated with H_2 formation from NADH. From a thermodynamic point of view there are several reasons to explain these observations. Firstly, the Gibbs energy ($\Delta_r G'$) for a specific reaction (**Eq. 1.14**) is calculated from the standard Gibbs energy ($\Delta_r G^{\circ\prime}$ (**Eq. 1.15**)) and the reactant concentrations by the following relation:

(**Eq. 1.14**) $$aA + bB \rightarrow cC + dD$$

(**Eq. 1.15**) $$\Delta_r G' = \Delta_r G^{\circ\prime} + RT \ln([C]^c[D]^d / [A]^a[B]^b)$$

Table 1.1. Overview of mesophilic, thermophilic, extremely-thermophilc and hyperthermophilic H_2-producing Bacteria (B) and Archaea (A). ^Abbreviations used for cultivation methods, B: batch; CB: controlled batch; CM: chemostat.

Organism	Domain	Temp. grown (°C)	Cultivation method^	Substrate
Mesophiles				
Clostridium acetobutylicum ATCC 4259	B	34	CB	glucose
Enterobacter aerogenes HU-101	B	37	CM	glucose
Klebsiella oxytoca HP1	B	38	CM	glucose
Thermophiles				
Clostridium thermosaccharolyticum LMG 6564	B	55	B	glucose
Thermoanaerobacterium saccharolyticum YS485	B	55	B	cellobiose
Clostridium thermocellum ATCC 27405	B	60	CM	α-cellulose
Thermoanaerobacterium thermosaccharolyticum PSU-2	B	60	B	starch
Extreme thermophiles				
Thermotoga elfii DSM 9442	B	65	B	glucose
Caldicellulosiruptor saccharolyticus DSM 8903	B	70	CB	sucrose
Thermoanaerobacter tengcongensis JCM 11007	B	75	B	glucose
Hyperthermophiles				
Thermotoga maritima DSM 3109	B	80	B	glucose
Thermotoga neapolitana DSM 4359	B	85	B	glucose
Thermococcus kodakaransis TSF100	A	85	CM	starch
Pyrococcus furiosus DSM 3638	A	90	B	cellobiose
		90	B	maltose

Organism	Fermentation end-products*	H_2/hexose	Reference
Mesophiles			
Clostridium acetobutylicum ATCC 4259	acetate, ethanol, acetone, butanol, butyrate	1.0	[13]
Enterobacter aerogenes HU-101	acetate, lactate, ethanol, acetoin, 2,3-butanediol	0.7	[14]
Klebsiella oxytoca HP1	nd	1.0	[15]
Thermophiles			
Clostridium thermosaccharolyticum LMG 6564	acetate, lactate, ethanol, butanol, butyrate	1.6	[16]
Thermoanaerobacterium saccharolyticum YS485	acetate, lactate, ethanol	0.9	[17]
Clostridium thermocellum ATCC 27405	acetate, lactate, ethanol, formate	1.7	[18]
Thermoanaerobacterium thermosaccharolyticum PSU-2	acetate, ethanol, butyrate	2.8	[19]
Extreme thermophiles			
Thermotoga elfii DSM 9442	acetate	3.3	[20]
Caldicellulosiruptor saccharolyticus DSM 8903	acetate, lactate	3.3	[21]
Thermoanaerobacter tengcongensis JCM 11007	acetate	4.0	[22]
Hyperthermophiles			
Thermotoga maritima DSM 3109	acetate	4.0	[23]
Thermotoga neapolitana DSM 4359	acetate, lactate	3.8	[24]
Thermococcus kodakaransis TSF100	acetate, alanine	3.3	[25]
Pyrococcus furiosus DSM 3638	acetate, alanine	2.8	[26]
	acetate, alanine	3.5	[26]

*Fermentation end-products other than H_2 and CO_2.

Where $\Delta_r G^{\circ\prime}$ is the standard Gibbs energy (J/mol, 1 molar concentration of all reactants, at a neutral pH and at a specific temperature, R is the gas constant (J/mol*K); T is the temperature (K), A and B are the substrate concentrations with respective stoichiometric reaction coefficients a, b; and C and D are the reaction products with respective stoichiometric reaction coefficients c, d. So, if H_2 concentrations are kept low, H_2 formation from NADH could even become exergonic, i.e. at a H_2 partial pressure of 10^{-4} atm proton reduction by NADH is associated with a negative $\Delta_r G'$ (-4.7 kJ/mol). **Figure 1.3** shows how the $\Delta_r G'$ value for proton reduction by NADH and Fd_{red} changes as a function of H_2 partial pressure. Secondly, also the temperature affects the thermodynamics of a reaction (**Eq. 1.16**)

(**Eq. 1.16**) $$\Delta G = \Delta H + T \cdot \Delta S$$

ΔG is the change in Gibbs energy; ΔH is the change in enthalpy; ΔS is the change in entropy; T is the temperature (K). **Eq. 1.16** indicates that H_2 formation becomes energetically more favourable at higher temperatures. The temperature dependency of $\Delta_r G'$ for proton reduction by NADH and Fd_{red} is illustrated in **Figure 1.3**. Finally, a third reason to overcome the unfavourable thermodynamics is to couple an energetically unfavourable reaction to an energetically more favourable reaction. For example, the coupling of NADH and Fd_{red} in a single H_2-forming reaction, catalysed by a bifurcating hydrogenase [30], decreases the overall Gibbs free energy of that reaction compared to a reaction with NADH alone (**Table 1.2, Figure 1.3**).

These data indicate that H_2 formation at elevated temperatures is thermodynamically more feasible, and that thermophilic H_2 producing organisms have a more favourable metabolism since they generally display a lower diversity in their fermentation end-products. Moreover, several thermophiles display an extensive enzyme inventory aiding in the decomposition and metabolism of complex biomass [33-35]. Overall this indicates that thermophiles are good candidates for biohydrogen production by dark fermentation.

Table 1.2. Gibbs free energy values for different fermentative reactions [27]. Data were calculated using [4, 28, 29].

Fermentative reaction	$\Delta G^{0\prime}$ kJ/reaction
$NADH + H^+ + pyruvate^- \rightarrow NAD^+ + lactate^-$	-25.0
$2NADH + 2H^+ + acetyl\text{-}CoA \rightarrow 2NAD^+ + ethanol + CoA$	-27.5
$NADH + H^+ + pyruvate^- + NH_4^+ \rightarrow NAD^+ + alanine + H_2O$	-36.7
$NADH + H^+ \rightarrow NAD^+ + H_2$	+18.1
2 reduced Ferredoxin(Fd_{red}) + $2H^+ \rightarrow$ 2 oxidized Ferredoxin(Fd_{ox}) + H_2	+2.7
$1/2\ NADH + Fd_{red} + 3/2\ H^+ + \text{-> } 1/2\ NAD^+ + Fd_{ox} + H_2$	+10.4

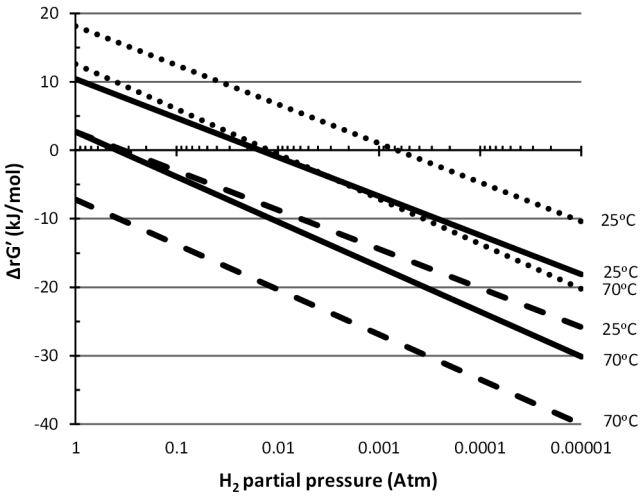


Figure 1.3 Effect of the H_2 concentration and temperature on the Gibbs energy change of reactions involved in H_2 formation. $\Delta G'$ of H_2 formation from NADH (dotted line), reduced ferredoxin (dashed line) and via the bifurcating system (50% NADH and 50% Fd_{red}) (solid line) at 25°C and 70°C. Values were calculated from data presented in [4, 28, 29, 31, 32].

4. BioH$_2$ formation by *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima*

In search for novel thermophilic cellulolytic micro-organisms, *Caldicellulosiruptor saccharolyticus* (**Figure 1.4**) was isolated from a natural enrichment site from at the Rotorua-Taupo thermal area in New-Zealand [36]. *C. saccharolyticus* is a gram-positive, thermophilic and strictly anaerobic bacterium that is capable of sustaining growth at a temperature range of 45-80°C (T_{opt} = 70 °C) and pH range of 5.5-8.0 (pH_{opt} = 7) [37]. *C. saccharolyticus*, belonging to the order Thermoanaerobacteriales (phylum Firmicutes, class Clostridia), has the capacity to grow on a broad substrate range, including different pentoses and hexoses, di-saccharides and polysaccharides like cellulose and xylan [36-39]. Especially its capacity to use cellulose at high temperatures is exceptional. The growth of *C. saccharolyticus* on sugar mixtures revealed the co-utilization of hexoses and pentoses, without any signs of carbon catabolite repression [40, 41]. *C. saccharolyticus* possesses a variety of endo- and exo-glycoside hydrolases [33, 42, 43], which allow it to degrade and grow on crop-based feedstock or on industrial waste stream derived biomass, resulting in high yields of H_2 [44-47].

In *C. saccharolyticus* the Embden-Meyerhof-Parnas (EMP) is the main route for glycolysis [48]. As fermentation end-products acetate, lactate and ethanol have been reported alongside CO_2 and H_2 (**Figure 1.5**). During glycolysis, NADH and reduced Fd

are produced as intermediate electron carriers. For reoxidation of NADH, formation of H_2 is believed to be catalysed by a cytoplasmic heterotetrameric Fe-only hydrogenase (HydA-D) as described for *Clostridium pasteurianum* [49] and *T. tengcongensis* [22]. The recycling of Fd_{red} to Fd_{ox} can be catalysed by a H_2 producing heteromultimeric membrane-bound NiFe hydrogenase (EchA-F), as identified in *T. tengcongensis* [22]. However, increased H_2 levels have been shown to result in a shift in the fermentation profile [50]. When H_2 formation becomes less favourable, produced NADH will partially be used by lactate dehydrogenase to produce lactate (**Figure 1.5**), while the observed formation of ethanol as fermentation end-product suggest the conversion of NADH and/or reduced ferredoxin to NADPH, which is the cofactor for its alcohol dehydrogenase [51].

Thermotoga maritima (**Figure 1.4**), is a gram-negative, hyperthermophilic, anaerobic H_2-producing bacterium belonging to the order Thermotogales (phylum Thermotogae) and was isolated from geothermal heated marine sediment at Vulcano, Italy. *T. maritima* is able to growth between 55°C and 90°C (T_{opt} = 80°C) and within a pH range of 5.5-9.0 (pH_{opt} = 6.5) [52]. It ferments a variety of simple and complex carbohydrates, including glucose, sucrose, starch, cellulose and xylan, to mainly acetate, lactate, CO_2 and H_2 [23, 35, 52]. *T. maritima* can use elemental sulphur S^0 as an external electron acceptor, however, the addition of S^0 reduces H_2-formation [23, 52]. Twenty-four percent of the genes from *T. maritima* are most similar to archaeal genes indicating a high frequency of lateral gene transfer between this bacterium and archaeal species during its evolutionary history [53]. The organisms genome contains a wide diversity of glycoside hydrolases and sugar transporters [35, 54]. Interestingly, *T. maritima* appears to have a preference for complex carbohydrates compared to mono-saccharides, as growth on the latter substrates is slower than growth in the presence of oligo/polysaccharides [55].

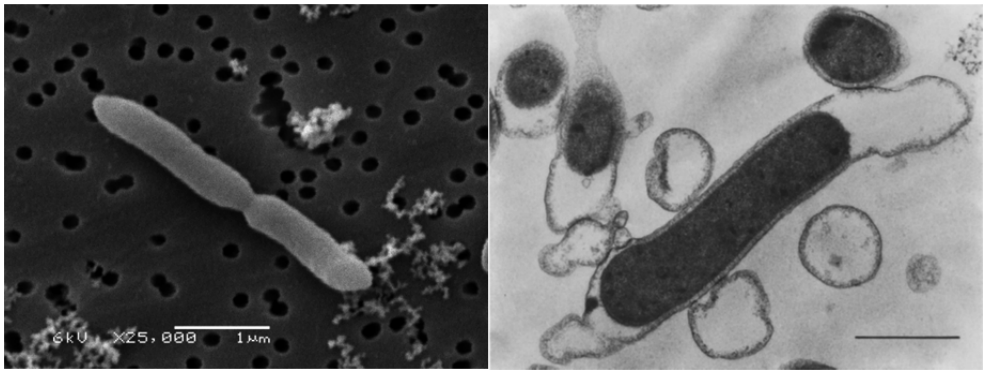


Figure 1.4 Left panel, EM-micrograph of a dividing *Caldicellulosiruptor saccharolyticus* cell (A. Pereira and M. Verhaart). Right panel, thin section of *Thermotoga maritima*, visualising its "toga", bar indicates 1µm [52].

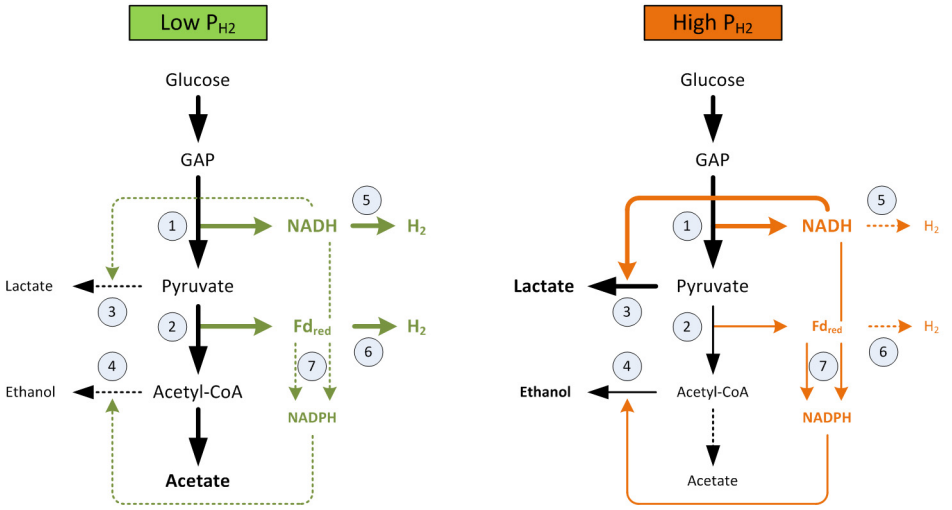


Figure 1.5 Scheme of typical carbon (black) and electron (green/orange) flow during glucose fermentation by *C. saccharolyticus* at low and high P(H_2). **1**, Glyceraldehyde-3-P dehydrogenase (GAPDH); **2**, Pyruvate:ferredoxin oxidoreductase (POR); **3**, Lactate dehydrogenase (LDH); **4**, Alcohol dehydrogenase; **5**, NADH-dependent hydrogenase; **6**, Ferredoxin-dependent hydrogenase; **7**, NADPH:ferredoxin / NADPH:NADH oxidoreductase.

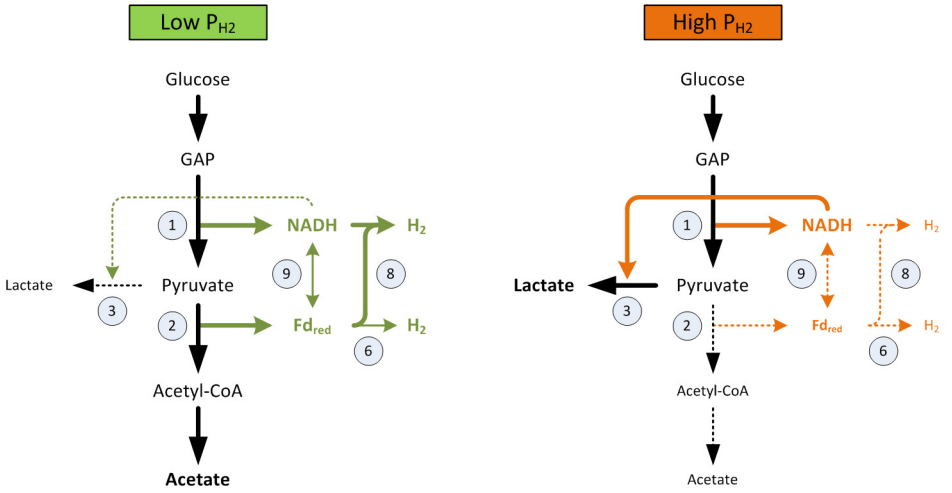


Figure 1.6 Scheme of typical carbon (black)- and electron (green/orange) flow during glucose fermentation by *T. maritima* at low and high P(H_2). **1**, Glyceraldehyde-3-P dehydrogenase (GAPDH); **2**, Pyruvate:ferredoxin oxidoreductase (POR); **3**, Lactate dehydrogenase (LDH); **6**, Ferredoxin-dependent hydrogenase; **8**, Bifurcating NADH/Ferredoxin dependent hydrogenase; **9**, NADH:ferredoxin oxidoreductase (Nfo).

T. maritima ferments its sugars mainly via the (EMP) pathway [23], generating the electron carriers NADH and Fd_{red} (**Figure 1.6**). Both reducing equivalents are consumed in a 1:1 ratio by a bifurcating trimeric Fe-only hydrogenase [30], indicating that the exergonic oxidation of Fd_{red} is used to drive the unfavourable oxidation of NADH, during H_2 formation. However, since part of the carbon intermediates are used for biosynthesis (e.g. biomass formation), the Fd_{red}/NADH ratio is always lower than 1. To maintain a 1:1 ratio of Fd_{red} and NADH, required in the bifurcating hydrogenase reaction, *T. maritima* seems to employ a Fd:NADH oxidoreductase. Non-ideal cultivation conditions lead to a shift in the fermentation profile, from mainly acetate to a mixed acid profile of acetate and lactate.

Because of their ability to use a broad range of biomass components and their outstanding H_2 evolving capacities, the thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima* have become model organisms for the study on biomass utilization and H_2 formation. Currently, several cellulolytic and weakly cellulolytic *Caldicellulosiruptor* species have been isolated and characterised [37, 56-62]. Likewise several *Thermotoga* species have been characterized [52, 63-66]. The availability of fully sequenced genomes of these species has allowed the investigation of the possible differences in their cellulolytic traits and the analysis of other remarkable features of these two genera [42, 67]. Overall a deeper understanding of the mechanisms involved in H_2 formation, the factors limiting H_2 production and the investigation of usable substrates will bring biohydrogen formation via dark fermentation a step closer.

5. Variants of the central glycolytic pathways: A role for pyrophosphate in energy metabolism?

The EMP pathway is one of the most widely distributed metabolic routes involved in the degradation of glucose to pyruvate (glycolysis). Especially at the level of fructose 6P conversion to fructose 1,6-bisphosphate, and of phosphoenolpyruvate (PEP) conversion to pyruvate, sets of fundamentally different reactions exist, involving different phosphoryl donors, like ATP, ADP and Pyrophosphate (PP_i) (**Figure 1.7**). Some organisms possess only enzymes which can use one of these phosphoryl donors, while other organisms have multiple genes coding for different enzymes requiring different phosphoryl donors. For some organisms, PP_i-dependent phosphofructokinase and pyruvate, phosphate dikinase are assumed to have a function in gluconeogenesis. On the other hand, several reports demonstrate a glycolytic function of these two enzymes. The involvement of PP_i-dependent conversions in catabolism and a role for PP_i as phosphoryl donor is intriguing. Participation of PP_i in catabolism would save an investment of ATP at the indicated catalytic steps. Especially for fermentative organisms, which only gain ATP via

substrate phosphorylation, this would lead to an overall increase in ATP yield. Pyrophosphate originates from biosynthetic reactions, like DNA and protein synthesis [68]. The effective removal of PP_i is required to drive these reactions forward [69], which is generally believed to occur via hydrolysis by soluble pyrophosphatases (PPase), generating inorganic phosphate. Alternatively, PPi may also be used in glycolysis as a phosphoryl donor. Furthermore, some organisms possess a membrane-bound pyrophosphatase which can either use the energy released upon PP_i hydrolysis to generate an electrochemical gradient or use this gradient to generate PP_i. Overall it seems that PP_i metabolism might be an important feature of the cell's energy metabolism. However, little is known about the genomic distribution of these PP_i-dependent glycolytic enzymes or their genomic co-occurrence with soluble or membrane-bound PPases.

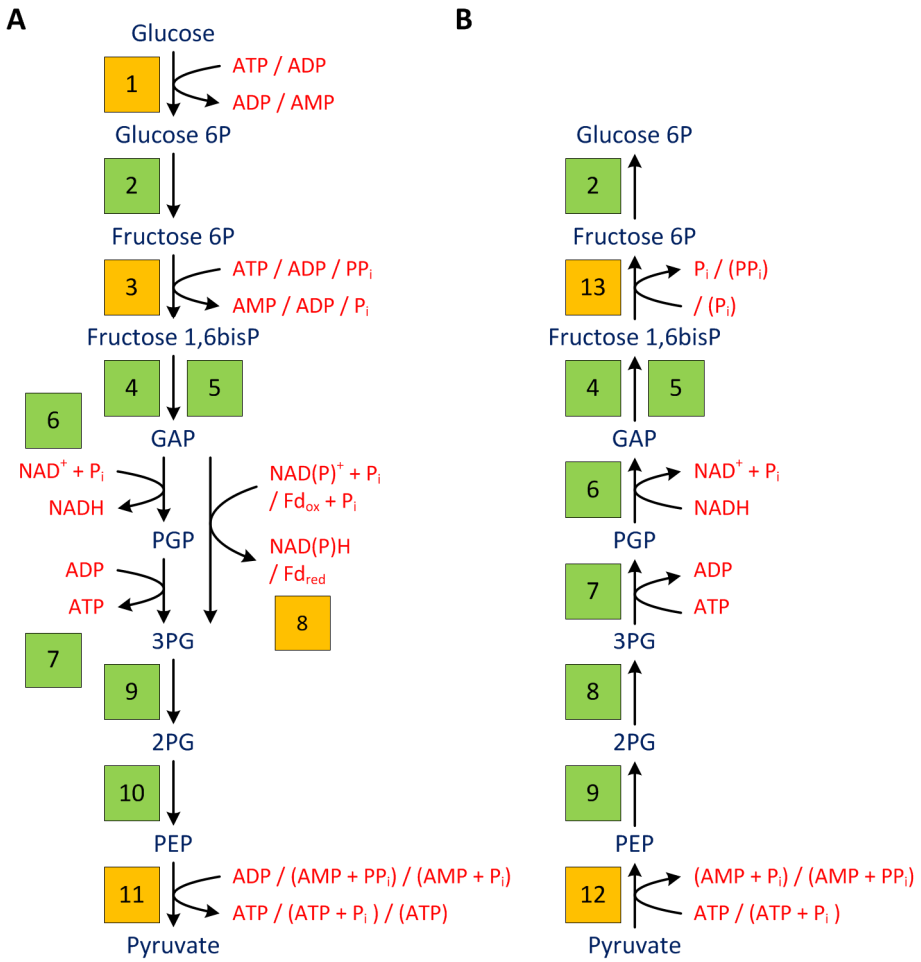


Figure 1.7 Variations of the cofactors involved in the enzyme catalysed conversion steps of the **A)** EMP pathway and **B)** gluconeogenic pathway. P_i, inorganic phosphate; PP_i inorganic pyrophosphate. Enzymes **1)** glucokinase; **2)** phosphoglucose isomerase; **3)** phosphofructokinase: **4)** fructose-bisphosphate aldolase; **5)** triosephosphate isomerase; **6)**; glyceraldehyde 3-phosphate dehydrogenase; **7)** phosphoglycerate kinase; **8)** non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase / glyceraldehyde-3-phosphate ferredoxin oxidoreductase; **9)** phosphoglycerate mutase; **10)** enolase; **11)** pyruvate kinase / pyruvate, phosphate dikinase / pyruvate, water dikinase; **12)** pyruvate, water dikinase /pyruvate, phosphate dikinase; **13)** fructose-bisphosphatase.

Thesis outline

Because of their favourable biomass degrading capabilities and H_2-forming features both *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima* have become model organisms in the study of thermophilic H_2 production. Novel insights in substrate usability, associated fermentation pathways and the mechanism involved in H_2 formation will provide steps forward in the application of these organisms for H_2 production and sustainable biological H_2 formation via dark fermentation in general.
Glycerol is formed as a by-product during biodiesel formation. Given the highly reduced state of carbon in glycerol this low cost substrate is of special interest for sustainable biofuel production. In **chapter 2** the use of glycerol for H_2-formation by *T. maritima* is investigated. Growth and high yield H_2 formation on glycerol is demonstrated in both batch and chemostat cultivation setups. In addition, the route of glycerol fermentation and the exceptional bioenergetics associated with H_2 formation from glycerol in *T. maritima* are discussed.

Elevated H_2 levels are known to inhibit H_2-formation during dark fermentations. In **chapter 3** the response of *C. saccharolyticus* to the exposure of elevated H_2 levels is investigated in different chemostat cultivation setups. The analysis of the fermentation profiles and transcriptome data associated with low and high H_2 levels provides insight into this organism's strategy to deal with elevated H_2 levels.

Chapter 4 describes several chemostat studies for elucidating the effect of increased H_2 levels on the fermentation profile of *C. saccharolyticus* with respect to i) growth on ammonium deficient media and ii) low/high substrate loads. It discusses the thermodynamics of H_2 formation with respect to the dissolved H_2 concentration. Moreover, the fermentation pathway of rhamnose, a sugar moiety of pectin, and its associated mechanisms of reductant disposal are discussed. This chapter also provides and extensive literature overview of the hydrolytic capability, sugar metabolism and H_2-producing capacity of *C. saccharolyticus*.

The role of inorganic pyrophosphate (PP_i) in the energy metabolism of *C. saccharolyticus* is investigated in **chapter 5**. In agreement with the annotated genome sequence PP_i-dependent phosphofructokinase, pyruvate phosphate dikinase and membrane bound pyrophosphatase activity can be detected in glucose-grown cultures. Pyrophosphate is demonstrated to inhibit pyruvate kinase activity. Furthermore, the dynamics in ATP and PP_i levels throughout batch growth is discussed.

Chapter 6 describes the genomic distribution of PP_i-dependent glycolytic enzymes and their genomic co-occurrence with soluble or membrane-bound pyrophosphatases in 495 fully sequenced genomes. An *ab initio* classification of enzyme-subtypes, which elaborates on known classifications systems and incorporates characterized protein features e.g. catalytic site residues and allosteric regulatory site residues, is presented.

The potential functional role of the PPi-dependent enzymes and membrane-bound pyrophosphatases is discussed. Overall the presented data indicates that the involvement of pyrophosphate in glycolysis/gluconeogenesis is a widespread phenomenon throughout the three domains of life.

Finally, **chapter 7** summarizes the research presented in this thesis and discusses the data in a broader context. In addition, several novel research strategies are proposed to increase our understanding of the metabolism involved in H_2 formation by the fermentative thermophiles *C. saccharolyticus* and *T. maritima*.

Chapter 2

Glycerol fermentation to hydrogen by *Thermotoga maritima*: proposed pathway and bioenergetic considerations

This chapter has been published as:

B.T. Maru[#], **A.A.M. Bielen**[#], M. Constantí, F. Medina, S.W.M. Kengen. *Glycerol fermentation to hydrogen by Thermotoga maritima: proposed pathway and bioenergetic considerations.* International Journal of Hydrogen Energy, 2013 38(14): 5563-5572

[#] These authors contributed equally to this paper

Abstract

The production of biohydrogen from glycerol, by the hyperthermophilic bacterium *Thermotoga maritima* DSM 3109, was investigated in batch and chemostat systems. *T. maritima* converted glycerol to mainly acetate, CO_2 and H_2. Maximal hydrogen yields of 2.84 and 2.41 hydrogen per glycerol were observed for batch and chemostat cultivations, respectively. For batch cultivations: i) hydrogen production rates decreased with increasing initial glycerol concentration, ii) growth and hydrogen production was optimal in the pH range of 7-7.5, and iii) a yeast extract concentration of 2 g/l led to optimal hydrogen production. Stable growth could be maintained in a chemostat, however, when dilution rates exceeded 0.025 h^{-1} glycerol conversion was incomplete. A detailed overview of the catabolic pathway involved in glycerol fermentation to hydrogen by *T. maritima* is given. Based on comparative genomics the ability to grow on glycerol can be considered as a general trait of *Thermotoga* species. The exceptional bioenergetics of hydrogen formation from glycerol is discussed.

1. Introduction

Hydrogen gas (H_2) is considered an attractive alternative to fossil fuels, as it burns cleanly, without emitting carbon dioxide (CO_2) or any other environmental pollutants [70]. H_2 possesses the highest energy content per unit of weight compared to other fuels, and it can be used in energy-efficient hydrogen fuel cells [71]. However, nearly 96% of the total current H_2 production, by catalytic steam reforming of natural gas, coal gasification or the partial oxidation of refinery oil, is still derived from fossil fuels. Since these processes are not based on renewable resources and still cause a net increase of atmospheric CO_2, they are not considered sustainable [72, 73]. To overcome the use of fossil hydrocarbons as sources for H_2 production, alternative methods, like electrolysis, thermal decomposition of water and biological methods, are preferred. The electro- and thermo-chemical means are very energy inefficient. Moreover, they still depend on fossil fuels for the generation of electricity and heat [9]. Biological hydrogen (bioH_2) production by bacteria, on the other hand, is far more promising due to its potential as an inexhaustible, low-cost and environmentally friendly process, especially when it is generated from a variety of renewable resources [5, 74]. BioH_2 is produced either by biophotolysis, microbial electrolysis, photofermentation, using light-dependent organisms, or by dark fermentation [10]. BioH_2 production by dark fermentation is an anaerobic process, involving heterotrophic fermentative bacteria or archaea that convert biomass or biomass-derived hydrocarbons mainly to H_2 and acetate. To enhance the economy of H_2 production by dark fermentation it is important to explore suitable biomass substrates which can be utilized by a broad range of H_2 producing microorganisms.

Recently many research efforts have been devoted to microbial conversion of low-priced industrial and agricultural waste into bioenergy [75-78]. One of these industrial wastes concerns crude glycerol [76, 79, 80]. Crude glycerol is an inevitable by-product of the production of biodiesel; about 10 kg crude glycerol, containing 50-60% pure glycerol, is produced for every 100 kg of biodiesel [81]. In recent years, the accelerated growth of the biodiesel industry has generated a surplus of glycerol, that resulted in a 10-fold decrease in crude glycerol prices. Furthermore, this has generated environmental concern associated with glycerol disposals [80]. As a result, glycerol has gone from being a chemical commodity to a chemical waste in less than a decade. Its availability, low price and its potential to mitigate possible environmental hazards make glycerol an attractive carbon source for industrial microbiology including the dark fermentation processes.

Yet another advantage is that fuels and reduced chemicals can be produced from glycerol at yields higher than those obtained from common sugars like glucose and xylose [80]. This is due to its highly reduced redox state of carbon, the degree of reduction per carbon for glucose and xylose is 4 compared to 4.67 for glycerol [82].

Until recently, the fermentative metabolism of glycerol had been reported in species of the genera *Klebsiella, Citrobacter*, *Enterobacter*, *Clostridium, Lactobacillus, Bacillus*, *and Anaerobiospirillum* [79, 80]. However, the potential for using these mesophilic organisms for H_2 production in dark fermentation is limited due to the low yield. In previous studies converting pure glycerol or glycerol-containing wastes [79, 83, 84] the maximal H_2 yield obtained was ~1 mol H_2 per mol of glycerol , concomitant with the production of ~1 mol of ethanol per mol of glycerol. Moreover, mesophilic microorganisms often produce reduced end-products such as diols and lactic acid, at the expense of H_2 [79, 85]. Therefore, for maximal H_2 production, oxidation of glycerol to acetic acid is required.

In light of yield optimization of H_2 from biomass, extreme thermophilic anaerobic bacteria are preferred. Their yields are reported to be approximately 83-100% of the maximum theoretical value of 4 mol hydrogen/mol glucose, in contrast to the mesophilic facultative anaerobes which show a H_2 yield of less than 2 [86]. Furthermore, thermophilic H_2 production benefits from some general advantages of performing processes at elevated temperatures, like a lower viscosity, better mixing, less risk of contamination, higher reaction rates and no need for reactor cooling [87]. Among the thermophiles, the order of the *Thermotogales* is characterized by the ability of its members to utilize a wide variety of carbohydrates [52] and to ferment sugars predominantly to acetate, CO_2, and H_2 [21, 23].

However, in literature some contradiction exists concerning the ability of *Thermotoga* species to convert glycerol. Previous studies reported that *Thermotoga maritima* contains the coding sequences for a complete pathway for both glycerol uptake and conversion [53]. A positive signal indicating oxidation of glycerol by *Thermotoga neapolitana* was found in a microplate assay [88]. Ngo *et al.* describes hydrogen production by *T. neapolitana* on biodiesel waste with a maximal yield of 2.73 mol H_2/mol glycerol consumed [89]. However, Eriksen *et al.* could not observe glycerol conversion by *T. maritima, T. neapolitana,* or *Thermotoga elfii* [90]. These, opposing results prompted us to reinvestigate the ability of *Thermotoga* species to use glycerol. Our preliminary data showed that *T. neapolitana*, but also *T. maritima* were able to form hydrogen from glycerol [91].

Here, bioH_2 production from glycerol by *T. maritima* was investigated in detail. Optimum growth parameters and cultivation conditions were determined. A putative glycerol catabolic pathway leading to hydrogen is presented, and the unusual thermodynamics and biochemistry of high yield hydrogen formation from glycerol are discussed.

2. Material and Methods

2.1. Organisms and growth conditions

Thermotoga maritima strain DSM 3109 [52] and *Thermotoga neapolitana* strain DSM 4359 [64] were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen and cultivated in M3 medium. M3 medium, which was based on M2 medium [92], consisted of (amounts are in grams per liter of deionized water): 1.5 g KH_2PO_4; 2.4 g $Na_2HPO_4 \cdot 2H_2O$; 0.5 g NH_4Cl; 0.2 g $MgCl_2 \cdot 6H_2O$; 2.0 mg $NiCl_2 \cdot 6H_2O$; NaCl, 2.7% (w/v) for *T. maritima* and 2.0% (w/v) for *T. neapolitana*; 11.9 g HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2 ethanesulphonic acid); 2 g yeast extract (YE); 15 mL trace element solution (DSM-TES, see DSMZ medium 141, complemented with Na_2WO_4 3.00 mg/L); 1.0 mL of vitamin solution (Biotin 2 mg, Nicotinamide 20 mg, *p*-Aminobenzoic acid 10 mg, Thiamine (Vit.B_1) 20 mg, Pantothenic acid 10 mg, Pyridoxamine 50 mg, Cyanocobalamin and Riboflavin 10 mg); 1.0 g/L of cysteine hydrochloride as reducing agent and 1 mg resazurin, which was used as a redox indicator. Anaerobic conditions were achieved by flushing the headspace of the serum bottles with N_2 gas. The starting pH of the medium was adjusted to pH 6.9 for *T. maritima* and pH 7.3 for *T. neapolitana* with 10 mM NaOH.

Batch cultivations were performed in 120- and 240-mL serum bottles with a working volume of 25 ml or 50 mL, at a constant temperature of 80°C and shaking at 200 rpm. Cultures were inoculated with a 10% (v/v) pre-culture. The effect of the glycerol concentration (2.5 - 40 g/L) on the fermentation performance was tested for both *T. maritima* and *T. neapolitana*. Optimal growth parameters (pH, YE concentration) for glycerol (2.5 g/L) conversion by *T. maritima* were investigated for the pH range of 4.9-9.2 and YE concentration range of 0-4 g/L.

Chemostat cultivations of *T. maritima* were performed in a 2-l jacketed bioreactor (Applikon) with a working volume of 1 L. Fermentations were run at 80°C, using a stirring speed of 300 rpm and pH was controlled at 7.0 by automatic addition of 2 N NaOH. The broth was continuously sparged with N_2 gas (4 NL/h). To prevent the loss of volatile end-products via the gas phase, off-gas was led through a water cooled condenser (4°C). Cultivations were performed in the M3 medium without HEPES, using a glycerol concentration of 2.5 g/L and a YE concentration of 2 g/L. The medium was inoculated with an exponentially growing pre-culture (5% (v/v)). During the batch start-up phase the broth was not sparged and the gas outlet was closed, mimicking the closed bottle setup. Fermentation performance was investigated during growth at different dilution rates (0.017, 0.025, 0.036 and 0.050 h^{-1}). The system was assumed to be in steady state when H_2 and CO_2 concentrations in the off gas and fermentation profiles were constant, which in all cases occurred after ~1.5 volume

change. For each dilution rate three samples at different time points were taken for further analysis.

2.2. Analytical methods

Substrate and fermentation end-product concentrations were determined by HPLC, using a Shodex RSpak KC-811 ion exclusion chromatography column operating at 80°C with a eluent of 3 mM H_2SO_4 (0.8 mL/min). Crotonic acid (10 mM) was added to the culture supernatant (16,000 × g, 10 min at 20°C) as an internal standard in a 1:1 ratio to correct for differences in HPLC injection volumes. Concentrations were quantified using standard curves of the respective compounds.

During batch experiments the serum bottles headspaces were analysed for H_2 and CO_2 levels by GC, equipped with a Molsieve 13X column and a CP Poraplot Q column, respectively. For the chemostat cultivations, off-gas composition was continuously monitored using a Compact GC equipped with a Carboxen 1010 PLOT column and a Micro thermal conductivity detector, using He as carrier gas.

Optical cell densities were determined at 600 nm (OD_{600}). Additionally, cell dry weight (CDW) was used to quantify the amount of biomass in the bioreactor during the continuous cultivations. CDWs were determined in technical duplicates. 2 x 15 mL culture was sampled and centrifuged (4800 g, 15 min at 4°C). Each pellet was re-suspended in 2 mL ultrapure water. CDWs were determined after drying the samples for 2 days in an oven at 120°C.

2.3. Data analysis

A modified Gompertz equation **Eq. 2.1** [93] was used to estimate the maximum production rates and the production potentials of the fermentation end-products acetate, lactate, H_2 and CO_2:

(**Eq. 2.1**) $$P_i(t) = P_{max,i} \exp\{-\exp[((R_{max,Pi} * e / P_{max,i})*(\lambda_i - t))+1]\}$$

where: P_i – concentration of product i (mmol/L), t – fermentation time (h), $P_{max,i}$ – maximum concentration of product i formation (mmol/L), $R_{max,Pi}$ – maximum production rate of product i formation (mmol/L*h), λ_i – lag phase time. Accordingly, for the consumption of glycerol a modified Gompertz equation **Eq. 2.2** [93] was used:

(**Eq. 2.2**) $$S_0 - S(t) = S_{max} \exp\{-\exp[((R_{max,s} * e / S_{max})*(\lambda_s - t))+1]\}$$

where: S_0 – initial substrate concentration (mmol/L), S – substrate concentration (mmol/L), S_{max} – maximum concentration of consumed substrate (mmol/l), $R_{max,S}$ –

maximum rate of substrate consumption (mmol/L*h). The fitting of the fermentation data was performed using Sigma plot application software version 12.3, where accuracy of the fit was given by correlation coefficients (R^2).
For batch cultivation yields of the fermentation end-products, expressed in mole product produced per mole glycerol consumed, were calculated using the values obtained from the data fits (**Eq. 2.1** and **Eq. 2.2**), according to **Eq. 2.3**:

(**Eq. 2.3**) $$Y_{Pmax,i} = P_{max,i} / (S_0 - S_{max})$$

Where: $Y_{Pmax,i}$ – substrate yield for fermentation product *i*, S_0 – initial glycerol concentration (mol/l), S_{max} – maximum glycerol concentration (mol/L). For the chemostat cultivations molar yields were calculated using the biomass specific production and consumption rates (mmol/g*h).

Carbon balances (C-balance) and a balances of degree of reduction (γ-balance) were calculated according to Heijnen *et al.* [94] using the standard elemental biomass composition $CH_{1.8}O_{0.5}N_{0.2}$, which corresponds to a biomass carbon content of 48.8% and a degree of reduction of biomass of 4.2 electrons per C atom. For the batch cultivations biomass levels were estimated from OD_{max} using the relation (CDW (g/L) = 0.84 * OD_{max}, R^2 = 0.658), which was derived from the chemostat experiments.

When calculating the biomass yield in grams of CDW per mole of ATP produced (Y_{ATP}) four assumptions were made: (I) During the anaerobic oxidation of 1 mole of glycerol to 1 mole of acetate, 3 moles of ATP are produced, (II) glycerol enters the cell via passive transport, (III) 1 ATP is required for the phosphorylation of glycerol to glycerol-3-phosphate, and (IV) 1 ATP is required for the uphill formation of H_2 by proton reduction coupled to the oxidation of glycerol 3-phosphate to dihydroxyacetone phosphate. Overall this results in the formation of 1 mole ATP per mole of acetate.

2.4. Genomic neighborhood analysis of genes involved in glycerol conversion

The genomic neighborhoods of the *T. maritima* genes involved in glycerol metabolism were investigated and compared with the fully sequenced species of the order Thermotogales: *Thermotoga lettingae* TMO, *Thermotoga naphthophila* RKU-10, *Thermotoga neapolitana* DSM 4359, *Thermotoga petrophila* RKU-1, *Thermotoga* sp. RQ2, *Kosmotoga olearia* TBF 19.5.1, *Petrotoga mobilis* SJ95, *Thermosipho africanus* TCF52B, *Thermosipho melanesiensis* BI429 and *Fervidobacterium nodosum* Rt17-B1 using the Integrated Microbial Genomes (IMG) system (img.jgi.doe.gov). The genomic context of the *T. maritima* orthologs of the glycerol-3-phosphate dehydrogenase coding gene cluster (TM1432-1434) was determined for *Pyrococcus furiosus* DSM 3638, *Thermococcus kodakarensis* KOD1, *Thermoanaerobacter tengcongensis* MB4T,

Caldicellulosiruptor saccharolyticus DSM 8903, *Clostridium acetobutylicum* ATCC 824 and *Clostridium butyricum* E4, BoNT E BL5262 (Draft genome).

Table 2.1 Effect of different glycerol concentration on substrate consumption, end-product production, H_2 productivities and yields for *T. maritima* and *T. neapolitana*.

Initial glycerol conc.	Maximal consumption (S_{max}) and production ($P_{max,i}$)*				
(mmol/L)	(mmol/L)				
S_0	S_{max}	$P_{max,Act}$	$P_{max,Lact}$**	$P_{max,CO2}$	$P_{max,H2}$
T. maritima					
29.9	24.8 (0.998)	21.0 (0.996)	0.1	23.9 (0.984)	70.5 (0.993)
71.4	22.4 (0.964)	17.1 (0.983)	0.0	19.2 (0.969)	62.6 (0.995)
164.2	17.0 (0.946)	16.0 (0.976)	0.1	16.9 (0.954)	48.2 (0.977)
210.8	22.0 (0.913)	18.9 (0.968)	0.1	17.4 (0.981)	46.7 (0.994)
T. neapolitana					
29.9	27.4 (0.990)	26.6 (0.996)	0.0	25.5 (0.982)	78.3 (0.997)
59.3	27.9 (0.999)	24.9 (0.995)	0.3	20.1 (0.994)	76.3 (0.999)
140.2	28.9 (0.985)	26.2 (0.997)	0.1	25.0 (0.997)	70.0 (0.998)
198.0	28.1 (0.970)	23.6 (0.995)	0.1	24.6 (0.996)	69.4 (0.999)

Initial glycerol conc.	Maximal H_2 production rate	Molar yields			OD_{max}	C-balance
(mmol/L)	(mmol/L/h)	(mol/mol)				%
S_0	$R_{max,H2}$	Y_{Act}	Y_{CO2}	Y_{H2}		
T. maritima						
29.9	1.01	0.84	0.96	2.84	0.64	105
71.4	0.63	0.76	0.85	2.79	0.43	97
164.2	0.38	0.94	0.99	2.84	0.37	101
210.8	0.38	0.86	0.79	2.12	0.45	100
T. neapolitana						
29.9	1.58	0.97	0.93	2.86	0.60	109
59.3	1.52	0.89	0.72	2.74	0.52	98
140.2	0.78	0.91	0.87	2.42	0.52	95
198.0	0.48	0.84	0.87	2.45	0.50	93

*Correlation coefficients (R^2) of the curve fits with the Gompertz equation (**Eq. 2.1** or **Eq. 2.2**) are given between brackets. **Low lactate concentrations prevented proper curve fitting.

3. Results and Discussions

3.1. Growth on glycerol

In contrast to earlier reports by Eriksen *et al.* [90] *T. maritima* was found to grow on glycerol as source of carbon and energy. Proper growth of *T. maritima* on glycerol required some adaptation time when the inoculum was pre-cultured on another substrate, like glucose. After several transfers on glycerol, the lag phase decreased and growth initiated directly after inoculation in standard medium. Glycerol was fermented mainly to acetate, CO_2, H_2 and minor amounts of lactate (**Figure 2.1A**; **Table 2.1**). The closely related *T. neapolitana,* that has been shown to grow on glycerol as well [89], was investigated here for comparison (**Figure 2.1B**; **Table 2.1**). To be able to quantify and compare the different growth experiments, time courses of substrate consumption and products formation were modelled using modified Gompertz equations (**Eq. 2.1** and **Eq. 2.2**). **Figure 2.1** shows a typical growth experiment for *T. maritima* and *T. neapolitana* with fitted curves for the main metabolites. The various growth parameters are summarized in **Table 2.1**. These data suggest that glycerol is fermented to acetate, CO_2 and H_2 in a ratio of 1:1:3. End-products commonly found in mesophilic glycerol fermentation by enterobacteria [95] or clostridia [96], like ethanol, butanol, 1,3-propanediol, 1,2-propanediol, succinate, or formate, were never observed. The relatively constant C-balance near 100% also indicates that no major end-product has been overlooked. In contrast to earlier data for *T. neapolitana* ([89]; ~5 mM lactate), almost no lactate was found. In accordance, the hydrogen yields of around 2.8 mol H_2/mol glycerol found here, were higher compared to the data of Ngo *et al*., who reported a value of 1.23 mol H_2/mol glycerol under non-N_2-sparged conditions [89]. This discrepancy could probably be a result of different culturing conditions, leading to variations in the dissolved H_2 concentration. For instance, Ngo *et al.* showed that N_2-sparging of the cultures led to H_2 yields (2.73 mol H_2/mol glycerol) [89], which are similar to the values found here. *T. neapolitana* is apparently able to adapt its metabolism from producing mainly H_2 to producing a mixture of H_2 and lactate, as reduced end-products.

Both *Thermotoga* species showed a substantial decrease in optical density when the culture was approaching the stationary phase (**Figure 2.1A & 2.1B**). A similar decrease in cell density has been reported by Ngo *et al.* for *T. neapolitana* [89]. The reason for the cell lysis is not known. However, this phenomenon did not affect the C-balance calculations, for which we used the maximum OD to estimate the carbon content of the biomass.

Our results clearly show that not only *T. neapolitana*, but also *T. maritima*, is perfectly able to grow on glycerol as source of carbon and energy. The inability of *T. neapolitana* and *T. maritima* to grow on glycerol, as reported by Eriksen *et al.* [90] is

likely a result of differences in the growth medium. For instance, the medium used by Eriksen *et al.* [90], which was based on a medium by Van Ooteghem *et al.* [88] had an initial pH of 8.5, which is at the boundary of the pH range supporting growth on glycerol, found in our study (see section 3.3). Moreover, Eriksen *et al.* used a lower NaCl concentration (10 versus 27 g/L) and added no additional Ni^{2+}, which is an essential metal in many hydrogenases.

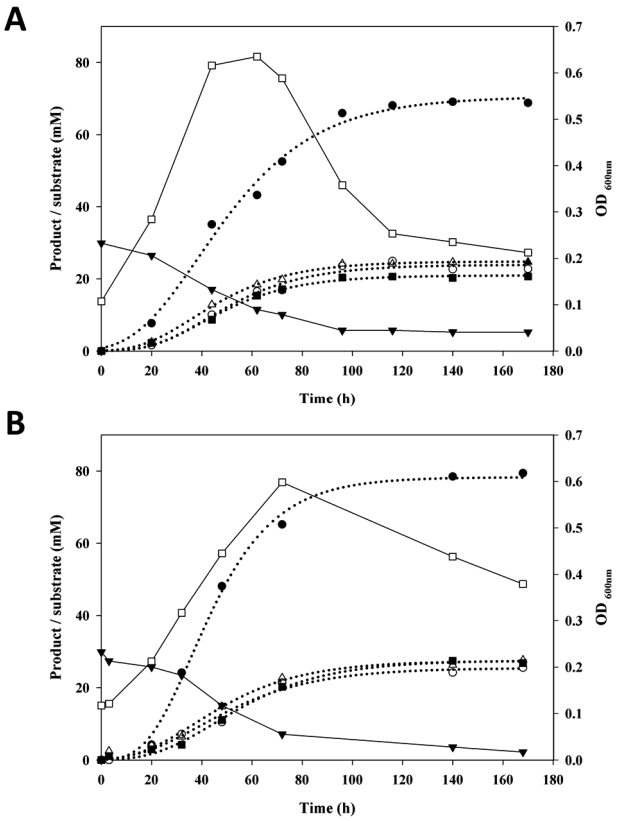


Figure 2.1 Fermentation profiles of batch cultivations on glycerol (2.5 g/L) of **A)** *T. maritima* and **B)** *T. neapolitana*. Residual glycerol (▼), glycerol consumed (Δ), acetate (■), H_2 (●), CO_2 (○) and OD (□). For glycerol consumed, acetate, H_2 and CO_2 data was fitted using the modified Gompertz equation (**Eq. 2.1** and **Eq. 2.2**) (dotted lines).

3.2. Effect of glycerol concentration

The type of carbon source and the initial substrate concentration usually play an important role on the bacterial growth and product yield [97]. Therefore, the effect of the initial glycerol concentration (29 mM – 210 mM) on hydrogen production by *T. maritima* and *T. neapolitana* was investigated (**Table 2.1**). No growth or H_2 formation

was observed when glycerol was omitted from the medium. Maximal H_2 production rates ($R_{max,H2}$) decreased with increasing initial glycerol concentration. And it can be observed that irrespective of the initial glycerol concentration, total glycerol consumption is rather constant and amounts to a maximum value of approximately 25 mM and 29 mM for *T. maritima* and *T. neapolitana*, respectively. This observation suggests that glycerol conversion is not limited by the amount of glycerol present, but by some other environmental parameter. A possible reason could be the accumulation of fermentation end-products, especially acetic acid, which has been shown before to cause end-product inhibition [50]. The accumulation of acetate, being a weak acid, may impair the growth of bacteria by dissipation of the membrane potential [50]. Alternatively, growth inhibition may be a result of the lowering of the pH (*vide infra*).

3.3. Effect of pH on fermentative H_2 production

During glycerol fermentation, the pH value dropped from ~7 at the start to ~6 in the stationary phase. Therefore, we were interested in the effect of the initial pH on the growth, which was assessed here by measuring the H_2 production. **Figure 2.2A** depicts the maximal H_2 concentration and the H_2 production rate for *T. maritima* as a function of the initial pH. Below pH 6 and above pH 8, H_2 production decreased considerably (**Supplementary Table S2.A**). The results are in agreement with the cessation of growth around pH 6 as shown in **Figure 2.1**. This might also explain the observation that higher initial glycerol concentrations do not lead to higher glycerol conversion. However, the investigation of the pH-dependence of *T. neapolitana* on glycerol, by Ngo *et al*. [98], showed a broader pH range with growth even possible at pH 5 and pH 9.

3.4. Effect of YE concentration on fermentative H_2 production

Yeast extract (YE) is an important environmental determinant for the growth of many microorganisms. Here, different YE concentrations (0, 0.5, 1, 2 and 4 g/L) were tested on the glycerol conversion efficiency and the H_2 producing capacity. As seen from **Figure 2.2B** glycerol conversion and H_2 production are low in the absence of YE. Addition of 0.5 g/L and 1 g/L YE results in a significantly better performance (**Supplementary Table S2.B**), but growth stimulation at YE concentrations exceeding 2 g/L is marginal. Overall, it is demonstrated that addition of YE enhances growth and hydrogen production on glycerol, which is in line with the report of Schröder *et al*. [23], who found that the addition of YE enhances growth and hydrogen production by *T. maritima* during growth on glucose.

Table 2.2 Fermentation details of *T. maritima* grown on glycerol (2.5 g/L) in a chemostat cultivation setup at different dilution rates.

Dilution rate	**Substrate and product concentrations**									
	Medium	Effluent				Gas-phase*				
h^{-1}	mM									
	Glycerol	Glycerol	Lactate	Acetate		H_2		CO_2		
0.017	27.01 ± 0.02	0.45 ± 0.02	0.44 ± 0.02	21.58 ± 0.06		64.07 ± 1.40		20.12 ± 0.39		
0.025	27.01 ± 0.02	0.27 ± 0.01	0.20 ± 0.02	22.37 ± 0.22		60.66 ± 0.45		19.26 ± 0.44		
0.035	27.01 ± 0.02	1.05 ± 0.04	0.17 ± 0.04	21.89 ± 0.19		57.93 ± 0.35		18.71 ± 0.05		
0.050	27.01 ± 0.02	6.72 ± 0.12	0	16.71 ± 0.51		46.73 ± 0.71		14.89 ± 0.56		

	Volumetric consumption/production rate				
	mmol/L*h				
	q(Glycerol)	q(Lactate)	q(Acetate)	q(H_2)	q(CO_2)
0.017	-0.45 ± 0.01	0.01 ± 0.01	0.37 ± 0.01	1.09 ± 0.02	0.34 ± 0.01
0.025	-0.67 ± 0.01	0.01 ± 0.01	0.56 ± 0.01	1.52 ± 0.01	0.48 ± 0.01
0.035	-0.93 ± 0.01	0.01 ± 0.01	0.79 ± 0.01	2.09 ± 0.01	0.67 ± 0.01
0.050	-1.01 ± 0.01	0	0.84 ± 0.03	2.34 ± 0.04	0.73 ± 0.01

	Molar yields					
	Per Glycerol consumed				Per Acetate produced	
	mol/mol					
	Y(Lactate)	Y(Acetate)	Y(H_2)	Y(CO_2)	Y(H_2)	Y(CO_2)
0.017	0.02 ± 0.01	0.81 ± 0.01	2.41 ± 0.05	0.76 ± 0.01	2.97 ± 0.06	0.93 ± 0.02
0.025	0.01 ± 0.01	0.84 ± 0.01	2.27 ± 0.02	0.72 ± 0.02	2.71 ± 0.01	0.86 ± 0.01
0.035	0.01 ± 0.01	0.84 ± 0.01	2.23 ± 0.01	0.72 ± 0.01	2.65 ± 0.04	0.85 ± 0.01
0.050	0	0.82 ± 0.01	2.30 ± 0.01	0.73 ± 0.02	2.80 ± 0.08	0.89 ± 0.05

*Values are expressed normalized to the liquid phase (1 L).

Dilution rate	**Biomass**			
				Biomass/ATP
h^{-1}		g/L	g/mol	g/mol
	OD_{600}	CDW	Yxs	Y_{ATP}
0.017	0.47 ± 0.01	0.42 ± 0.01	15.73 ± 0.34	19.35 ± 0.45
0.025	0.63 ± 0.01	0.54 ± 0.03	20.22 ± 1.27	24.16 ± 1.45
0.035	0.70 ± 0.01	0.53 ± 0.03	20.98 ± 0.98	24.89 ± 1.27
0.050	0.66 ± 0.02	0.50 ± 0.01	24.68 ± 0.48	30.20 ± 1.41

	Specific consumption/production rate				
	mmol/g*h				
	q(Glycerol)	q(Lactate)	q(Acetate)	$q(H_2)$	$q(CO_2)$
0.017	-1.08 ± 0.02	0.02 ± 0.01	0.88 ± 0.02	2.61 ± 0.06	0.82 ± 0.02
0.025	-1.24 ± 0.08	0.01 ± 0.01	1.04 ± 0.06	2.82 ± 0.17	0.89 ± 0.07
0.035	-1.72 ± 0.08	0.01 ± 0.01	1.45 ± 0.08	3.84 ± 0.17	1.24 ± 0.06
0.050	-2.03 ± 0.04	0	1.67 ± 0.09	4.67 ± 0.13	1.49 ± 0.05

	C-balance	**γ-balance**
	Recovery	
	%	
0.017	102.4 ± 0.8	101.5 ± 1.0
0.025	107.9 ± 2.1	105.5 ± 2.0
0.035	109.0 ± 1.4	106.0 ± 1.2
0.050	109.6 ± 0.9	107.6 ± 0.9

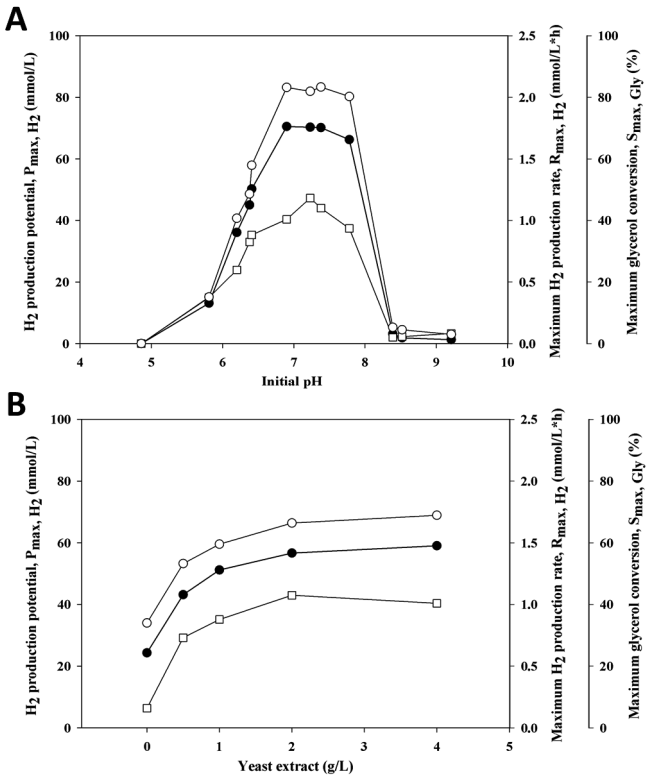


Figure 2.2 The effect of (a) initial pH and (b) yeast extract concentrations on batch fermentation of *T. maritima* grown on glycerol (2.5 g/L). Maximum glycerol conversion (○), maximal H_2 production rate (□) and H_2 production potential (●).

3.5. Glycerol fermentation by T. maritima in a continuous system

The possibility to grow *T. maritima* on glycerol in a chemostat setup was investigated. Our results (**Table 2.2**) show that *T. maritima*, grown on glycerol (2.5g/L), can be maintained in a continuous cultivation setup at different dilution rates (d = 0.015 h^{-1} – 0.050 h^{-1}). Similar to the batch cultivations, glycerol was mainly converted to acetate, H_2 and CO_2. Trace amounts of lactate were observed but no ethanol formation. Hydrogen yields ranged between 2.23-2.41 mol/mol glycerol consumed, which is somewhat lower than calculated for the batch cultivations as presented above. H_2 yields per acetate reach an average of 2.8 H_2/acetate, showing that almost all electrons released during acetate formation end up in H_2. Overall carbon recovery exceeded 100%, which probably reflects the consumption of YE for biomass formation. At higher dilution rates (d = 0.035 and 0.050), glycerol conversion was not

complete. However, since no washout occurred, it is assumed that a factor other than glycerol was limiting growth. However, volumetric and specific H_2 production rates increased with increasing dilution rates. Interestingly, biomass yields (Y_{xs}) increased with increasing dilution rate, which may reflect changes in ATP usage for maintenance (**Table 2.2**).

3.6. Comparative genomics of Thermotogales species

T. maritima was proposed to have the complete glycerol degradation pathway [53]. In general, glycerol is assumed to enter the glycolysis, either via dihydroxyacetone (DHA) or glycerol-3-phosphate (G3P), at the dihydroxyacetone phosphate (DHAP) node. A closer inspection of the *T. maritima* genome revealed that the pathway of glycerol assimilation via DHA is incomplete. Although a gene coding for a glycerol dehydrogenase (*dhaD*, EC 1.1.1.6, TM0432) is present, genes coding for the ATP-dependent (*dhaK,* EC 2.7.1.29) or PEP-dependent (*dhaKLM* EC 2.7.1.121) dihydroxyacetone kinase could not be identified.

On the other hand, the glycerol 3-phosphate pathway, consisting of a glycerol kinase (GK) and a glycerol-3-phosphate dehydrogenase (G3PDH), is present. Two genes coding for a glycerol kinase (*glpK-1* & *-2,* EC 2.7.1.30, TM0952 & TM1430) were identified*,* while TM0378 and gene cluster (TM1432-1434) both code for a G3PDH. TM0378 catalyses the NAD(P)H-dependent reduction of DHAP to glycerol-3-phosphate and is usually associated with the biosynthesis of phospholipids (*gpsA,* EC:1.1.1.94). In *Escherichia coli*, *gpsA* deficient strains were identified as glycerol-3-phosphate auxotrophs [99]. No gene coding for an aerobic sn-glycerol-3-phosphate dehydrogenase (*glpD*, EC: 1.1.5.3) was identified in *T. maritima*. In *E. coli* GlpD is an essential membrane enzyme, shuttling electrons via FAD and ubiquinone into the respiratory pathway, finally ending up at oxygen [100, 101]. Under anaerobic conditions the anaerobic glycerol-3-phosphate dehydrogenase (*E. coli*, glpABC, EC: 1.1.5.3) [102] fulfils a similar function, where reducing equivalents are passed on via an electron transport chain to electron acceptors like fumarate or nitrate [103]. The first gene of the *T. maritima* G3PDH gene cluster (TM1432-TM1434) is annotated as a *glpA* homologue. However, the genes in the *T. maritima* cluster belong to different clusters of orthologous groups (COGs 0579, 1249 and 3862, respectively) with respect to the *E. coli glpABC* genes (COGs 0578, 3075 and 0247, respectively), and probably code for a different uncharacterized type of G3PDH.

Like *T. maritima,* all sequenced Thermotogales (order) species contain such a G3PDH gene cluster. Except for *Kosmotoga olearia* and *Fervidobacterium nodosum*, this cluster co-localized with a glycerol kinase and a glycerol uptake operon anti-terminator (**Supplementary Table S2.C**). The co-localization of this G3PDH gene cluster with a glycerol kinase and a glycerol uptake operon anti-terminator can also be

observed in other organisms (**Figure 2.3**). Based on its genomic context, these findings suggest that this G3PDH gene cluster is likely to perform the oxidation of G3P to DHAP during the glycerol degradation in *T. maritima* and possibly in other organisms as well.

Interestingly, the B subunits of the G3PDH gene clusters in *T kodakaraensis, T. tengcongensis* and the investigated *Clostridium* species are represented by a different COG (COG0492), with respect to the B subunits from both *T. maritima* and *E. coli*. But all the G3PDH gene cluster B subunits contain a pyridine nucleotide-disulphide oxidoreductase, FAD/NAD(P)-binding domain (InterProScan, IPR023753, ebi.ac.uk/interpro/), which encompasses a small NADH binding domain within a larger FAD binding domain (Pfam, PF07992, pfam.sanger.ac.uk). The function of this oxidoreductase is not known, but it may be involved in the transfer of electrons from G3P to the hydrogenase complex. However, further characterisation is required to determine its exact role and whether NAD^+ is the final electron acceptor in this reaction.

With respect to the trans-membrane transport of glycerol, of all investigated *Thermotogales* species only *T. maritima* contains a gene coding for a glycerol transport facilitator (**Supplementary Table S2.C**). Since *T. neapolitana* is also capable to grow on glycerol (This study and Ngo *et al.* [89]), this probably indicates that the glycerol transport facilitator is not mandatory for growth on glycerol and glycerol enters the cell by diffusion.

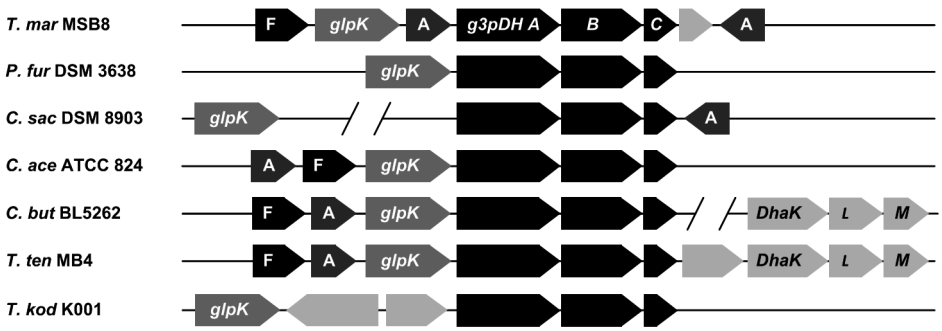


Figure 2.3 Gene ortholog neighborhood of the *Thermotoga maritima* glycerol-3-phosphate dehydrogenase gene cluster (*g3pDH ABC*, TM1432-34) in *Pyrococcus furiosus* DSM 3638, *Thermococcus kodakarensis* KOD1, *Thermoanaerobacter tengcongensis* MB4T, *Caldicellulosiruptor saccharolyticus* DSM 8903, *Clostridium acetobutylicum* ATCC 824 and *Clostridium butyricum* E4, BoNT E BL5262. Arrowhead direction indicates relative gene orientation. F, probable glycerol uptake facilitator protein (*glpF*, TM1429); *glpK,* Glycerol kinase (TM1430); A, Glycerol uptake operon antiterminator (TM1431). *DhaKLM*, PEP-dependent dihydroxyacetone kinase subunits K,L &M.

3.7. Glycerol degradation pathway in Thermotoga maritima

The proposed pathway for glycerol degradation in *T. maritima* is presented in **Figure 2.4**. *T. maritima* utilizes both the EMP and Entner-Doudoroff (ED) pathways [104] when grown on glucose. The presence of all the conventional EM-pathway enzymes have been confirmed in cell extracts [23], and the corresponding genes have been identified in the *T. maritima* genome [53]. It was shown that the oxidation of glyceraldehyde-3-phosphate is catalysed by an NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase [23] and the conversion of pyruvate to acetyl-CoA is catalysed by a pyruvate: ferredoxin oxidoreductase [105]. This indicates that, when glucose is converted to acetate, the reducing equivalents NADH and reduced ferredoxin are generated in a 1:1 ratio. The same reductant ratio is required in the H_2 formation step, which is catalysed by a bifurcating hydrogenase (TM1424-1426) [30]. However, the catabolism of glycerol via G3P requires an additional oxidation step. As discussed above, in *T. maritima* this is catalysed by a G3PDH, generating an additional reduced electron carrier. Based on the analogy with respiring glycerol utilizing microorganisms, and considering the redox potential of the DHAP/G3P couple of E_0 = -190 mV [4], FAD is a likely electron acceptor (FAD/$FADH_2$ couple of E_0 = -220 mV). The clustering on the genome with the pyridine nucleotide-disulphide oxidoreductase (FAD/NAD(P)-binding domain) suggests that $FADH_2$ might be used for the uphill formation of NADH. The formation of either $FADH_2$ or NADH disturbs the preferred 1:1 NADH/ferredoxin ratio, which is needed for H_2 formation via the bifurcating hydrogenase. Nevertheless, the observed H_2 yields of 2.5-3.0 mol/mol acetate, suggests that reductant derived from the oxidation of G3P, is also channelled to H_2. From a thermodynamic viewpoint the oxidation of G3P to DHAP and H_2 is not feasible ($\Delta G^{0\prime}$ +61 kJ/mol). Reverse electron flow, coupled to a Na^+ gradient, may enable the uphill electron transfer from NADH to the level of ferredoxin [106]. *T. maritima* contains a Rnf-cluster (TM0244-0249) that could accommodate this energy-dependent formation of reduced ferredoxin.

Based on these considerations it is expected that, the complete conversion of glycerol to acetate results in an ATP yield of 1 ATP per acetate. In *T. maritima,* glycerol enters the cell via passive transport either by diffusion or facilitated transport; one ATP is consumed when glycerol is phosphorylated; another ATP is required for the uphill formation of reduced ferredoxin from $FADH_2$ or NADH to restore the desired 1:1 NADH/ferredoxin ratio. Two ATP are generated during the conversion of 1,3-bisphosphoglycerate to 3-phosphoglycerate and the conversion phosphoenolpyruvate to pyruvate. The latter reaction could be catalysed either by a pyruvate kinase or a pyruvate, phosphate dikinase (PPDK). Based on the *T. maritima* genome, a catabolic role for PPDK seems very likely since the PPDK gene (TM0272) clusters with the fructose bis-phosphate aldolase (TM0273) and the acetate kinase (TM0274) genes as

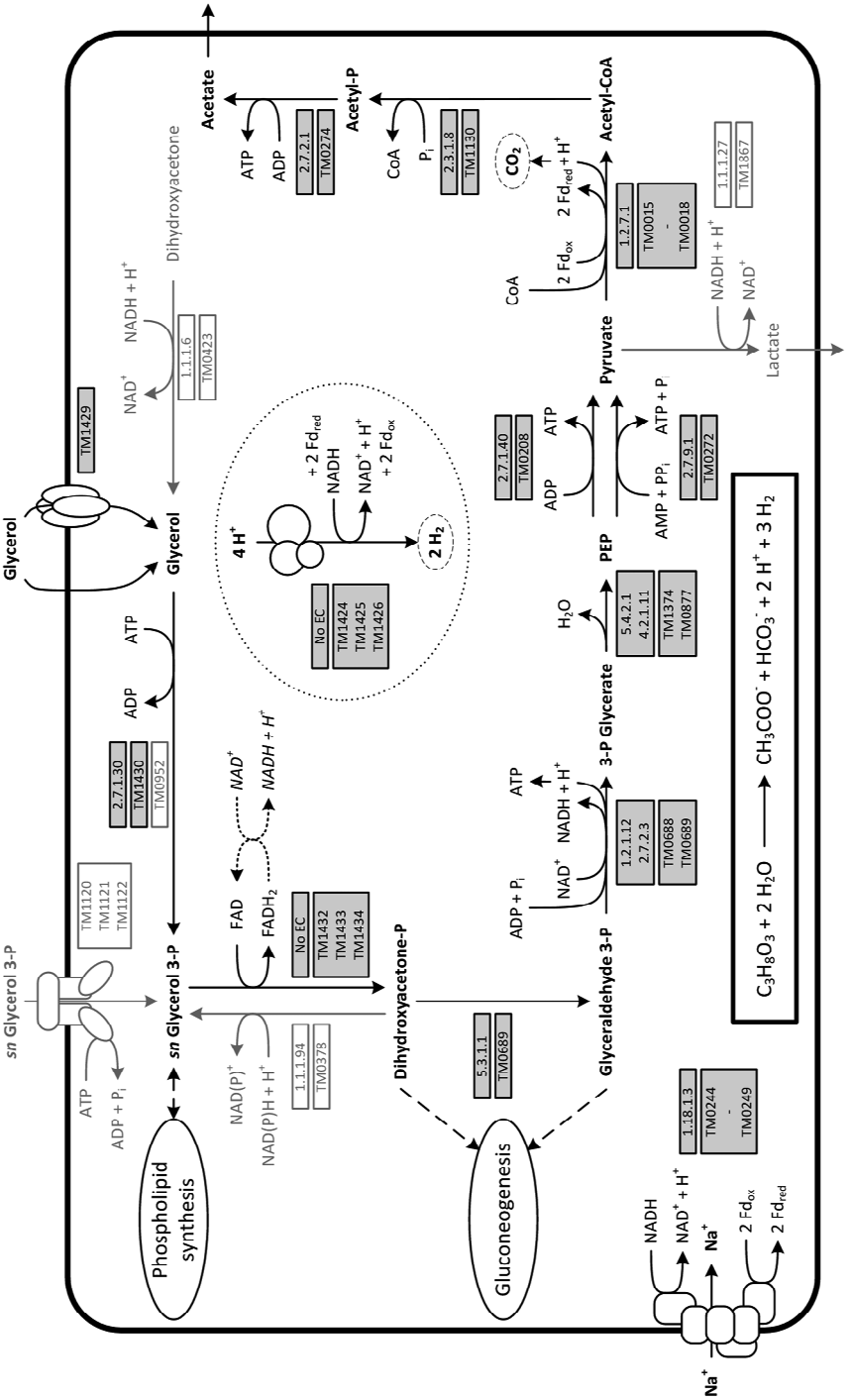

Acetate
Acetyl-P
Acetyl-CoA
Dihydroxyacetone
Glycerol
Pyruvate
Lactate
PEP
3-P Glycerate
Glyceraldehyde 3-P
Dihydroxyacetone-P
sn Glycerol 3-P
Phospholipid synthesis
Gluconeogenesis
ATP
ADP
CoA
NADH + H+
NAD+
2 Fdox
2 Fdred
4 H+
2 H2
H2O
CO2
ADP + Pi
NAD+ NADH + H+
FAD
FADH2
NAD(P)+
NAD(P)H + H+
AMP + PPi
2.7.2.1
TM0274
2.3.1.8
TM1130
1.2.7.1
TM0015 - TM0018
1.1.1.27
TM1367
1.1.1.6
TM0423
TM1429
2.7.1.30
TM1430
TM0952
No EC
TM1424
TM1425
TM1426
2.7.1.40
TM0208
2.7.9.1
TM0272
5.4.2.1
4.2.1.11
TM1374
TM0877
1.2.1.12
2.7.2.3
TM0688
TM0689
No EC
TM1432
TM1433
TM1434
1.1.1.94
TM0378
TM1120
TM1121
TM1122
5.3.1.1
TM0689
1.18.1.3
TM0244 - TM0249
C3H8O3 + 2 H2O → CH3COO- + HCO3- + 2 H+ + 3 H2

Figure 2.4 Proposed pathway for glycerol catabolism in *Thermotoga maritima*. For each reaction the locus tags of the genes coding for the respective enzymes and their EC numbers are given. Additional reactions involved in glycerol metabolism are indicated in light grey. Overall the complete conversion of glycerol to acetate yields 3 H_2 and a CO_2; lactate can be considered a minor side product of glycerol degradation. The membrane bound ion-translocating Fd:NADH oxidoreductase (EC 1.18.1.3) (Rnf) ensures the required NADH/Ferredoxin (Fd_{red}) ratio for the bifurcating hydrogenase reaction (encircled, no EC) at the expense of ion influx.

observed previously [107]. Finally, a third ATP is generated during the formation of acetate from acetyl phosphate. The ATP yield corresponds with the Gibbs energy of (-3.2 kJ/mol) that can be calculated for the oxidation of glycerol to acetate and H_2 and the average amount of Gibbs energy required for the synthesis of 1 ATP (-70 kJ/mol, [108]).

4. Conclusions

In this study, experimental evidence clearly shows that the hyperthermophile *Thermotoga maritima* DSM 3109 is able to grow on glycerol in both batch and chemostat cultivation setups. *T. maritima* converted glycerol to mainly acetate, CO_2 and H_2, with a maximal observed H_2 yield of 2.84 mol H_2 per mol glycerol consumed. The fermentation data suggest a stoichiometry of 1:1:3 ratio for acetate, CO_2, and H_2, respectively. The observed low diversity in fermentation end-products corresponds with the high H_2 yields, which are superior compared to those generally observed for mesophilic organism (~1 mol/mol).

For batch cultivations optimal H_2 production was realised using an initial pH of 7-7.5 and a yeast extract concentration of 2 g/L. Fermentation performances of *T. maritima* on the different initial glycerol concentrations were comparable to those observed for *T. neapolitana*, with maximal observed H_2 production rates of 1.0 and 1.6 mmol/ L*h, respectively. The H_2 production rates decreased with increasing initial glycerol concentration, and substrate consumption was incomplete. Growth in controlled batch systems with fixed pH, might allow complete substrate conversion at higher glycerol loads, thus improving the cumulative H_2 production.

Stable growth on glycerol could be achieved for *T. maritima* in a chemostat system. H_2 was produced with yields ranging between 2.23 and 2.41 mol/mol. For the chemostat cultivations, the H_2 production rate increased with increasing dilution rate (from 1.1 till 2.34 mmol/L*h), however, at dilution rates exceeding 0.025 h-1, glycerol (2.5 g/L) conversion was incomplete.

A pathway for glycerol fermentation by *T. maritima* is proposed. Based on comparative genomics, the ability to grow on glycerol can be considered as a general trait of *Thermotoga* species. In all probability, glycerol enters glycolysis via glycerol-3-phosphate. The co-localization of the genes coding for a glycerol kinase and an

uncharacterized multimeric glycerol-3-phosphate dehydrogenase suggest their involvement in glycerol catabolism. The observed H_2 yields of 2.5-3.0 mol H_2 per mol acetate, indicated that reductant derived from the oxidation of glycerol-3-phosphate, is also channelled to H_2. However, the exact mechanism of how to overcome the endergonic electron transfer from glycerol-3-phosphate to H_2 requires further investigation.

Supplementary materials

Supplementary Table S2.A
Supplementary Table S2.B
Supplementary Table S2.C

Supplementary materials can be found online.
(http://dx.doi.org/10.1016/j.ijhydene.2013.02.130)

Chapter 3

A thermophile under pressure: Transcriptional analysis of the response of *Caldicellulosiruptor saccharolyticus* to different H_2 partial pressures

This chapter has been published as:

A.A.M. Bielen, M.R.A. Verhaart, A.L. VanFossen, S.E. Blumer-Schuette, A.J.M. Stams, J. van der Oost, R.M. Kelly and S.M.W. Kengen. *A thermophile under pressure: Transcriptional analysis of the response of Caldicellulosiruptor saccharolyticus to different H_2 partial pressures.* International Journal of Hydrogen Energy, 2013 38 (4): 1837-1849

Abstract

Increased hydrogen (H_2) levels are known to inhibit H_2 formation in *Caldicellulosiruptor saccharolyticus*. To investigate this organism's strategy for dealing with elevated H_2 levels the effect of the hydrogen partial pressure (P_{H_2}) on fermentation performance was studied by growing cultures under high and low P_{H_2} in a glucose limited chemostat setup. Transcriptome analysis revealed the upregulation of genes involved in the disposal of reducing equivalents under high P_{H_2}, like lactate dehydrogenase and alcohol dehydrogenase as well as the NADH-dependent and ferredoxin-dependent hydrogenases. These findings are line with the observed shift in fermentation profiles from acetate production to the production of acetate, lactate and ethanol under high P_{H_2}. In addition, differential transcription was observed for genes involved in carbon metabolism, fatty acid biosynthesis and several transport systems. In addition, presented transcription data provide evidence for the involvement of the redox sensing Rex protein in gene regulation under high P_{H_2} cultivation conditions.

1. Introduction

Hydrogen (H_2) has its applications in a diversity of industrial processes, and can also be used as a fuel. However, only H_2 produced from renewable resources is a sustainable production process. Several sustainable technologies for the biological production of H_2 are currently under investigation. One of the promising alternatives is the anaerobic H_2 fermentation in the absence of light, often referred to as dark fermentation [109]. A model organisms in the field of thermophilic H_2 production is *Caldicellulosiruptor saccharolyticus,* a thermophilic, gram-positive anaerobe that ferments a broad range of sugar-substrates to mainly acetate, CO_2, and H_2 [37]. Its high hydrolytic and H_2-producing capacity makes this bacterium an attractive candidate for microbial biohydrogen production [40, 43, 46, 48].

For *C. saccharolyticus* the main route of glycolysis is the EMP pathway [48]. In the oxidation of glucose to pyruvate NADH is generated via the glyceraldehyde-3-phosphate dehydrogenase reaction and the subsequent oxidation of pyruvate to acetyl-CoA, performed by a pyruvate:ferredoxin oxidoreductase, produces reduced ferredoxin (Fd_{red}). During the successive conversion of acetyl-CoA to acetate, via acetyl phosphate, an additional ATP is generated [40]. Electron carriers (NAD^+, ferredoxin) that become reduced during the oxidation of glucose, have to be recycled to maintain the glycolytic flux. ATP formation is maximized if all the reducing equivalents are used for the reduction of protons instead of the reduction of metabolic intermediates, resulting in the maximum acetate formation. Hydrogenases couple the oxidation of NADH and/or Fd_{red} to proton reduction, thus generating H_2. The genome of *C. saccharolyticus* contains a gene cluster coding for a NADH-dependent cytosolic hetero-tetrameric Fe-only hydrogenase (*hyd*) and a cluster encoding a membrane bound multimeric [NiFe] hydrogenase (*ech*), which couples the oxidation of Fd_{red} to H_2 production [40]. These hydrogenase gene clusters have similar gene arrangements as the clusters identified in *Caldanaerobacter subterraneus* subsp. *tengcongensis* (formerly '*Thermoanaerobacter tengcongensis*') [22].

When the carbon flux toward biomass is neglected, the complete oxidation of 1 mol of glucose yields 2 mol of acetate, 2 mol of CO_2 and 4 mol of hydrogen. H_2 yields close to this theoretical limit have been reported for *C. saccharolyticus* in both batch [110] and continuous cultivations systems [48], 3.8 and 3.6 respectively. However, cultivation parameters like increased growth rate [48], increased substrate load [48] and type of substrate [45, 47] negatively affect H_2 yields. Furthermore, H_2 productivities decrease when *C. saccharolyticus* is grown under increased osmotic pressure [50, 111] or when H_2 is allowed to accumulate in the system [50, 112, 113]. H_2 accumulation in the liquid phase occurs when H_2 production rates exceed the H_2 liquid-to-gas mass transfer rate. This increased level of dissolved H_2 inhibits hydrogenase activity, partially preventing reducing equivalents to be used for H_2

formation. Instead, these reductants are consumed during the generation of other reduced end-products like lactate and/or ethanol. This shift in the fermentation profile from mainly acetate to a mixed end-product profile of acetate and lactate and ethanol, results in a decrease of both the H_2 and ATP yields.

Although studies on the effect of the hydrogen partial pressure (P_{H_2}) on fermentation profiles have been performed [50, 112, 113], nothing is known about the effects at the level of gene expression. To investigate the organism's molecular mechanisms for dealing with elevated H_2 levels and to identify alternative pathways involved in the recycling of the reducing equivalents, the effect of the P_{H_2} on fermentation performance and gene expression was studied. For these purposes *C. saccharolyticus* cultures were grown under different P_{H_2} in a glucose-limited chemostat setup.

2. Materials and methods

2.1. Organism, medium and cultivation conditions

C. saccharolyticus DSM 8903 (also ATCC 43494) was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen. *C. saccharolyticus* was grown in sterilized modified DSMZ 640 medium containing NH_4Cl (0.9 g/L), KH_2PO_4 (0.75 g/L), K_2HPO_4 (1.5 g/L), yeast extract (1 g/L), modified trace element solution SL-10 (1 ml/L; van Niel *et al.* [21], where $MnCl_2{\cdot}4H_2O$ and $Na_2SeO_3{\cdot}5H_2O$ were replaced by $MnSO_4{\cdot}H_2O$ and SeO_2), glucose (4 g/L), $MgCl_2$ (0.4 g/L) and cysteine-HCl (0.5 g/L). Resazurin (0.5 mg/L) was added as a redox indicator. Glucose, $MgCl_2$ and cysteine-HCl solutions were sterilized separately and added to the medium prior to inoculation with a batch pre-culture (5% v/v). The continuous cultures were glucose-limited unless stated otherwise. Cultivations were performed under anaerobic conditions in a 2-L jacketed bioreactor (Applikon) with a working volume of 1 L and an internal reactor pressure equal to the atmospheric pressure. Fermentations were run at 70°C with a dilution rate of 0.1 h^{-1}. The pH was controlled at 7.0 by automatic addition of 2 N NaOH. To prevent the loss of volatile end-products via the gas phase, off-gas was led through a water cooled condenser (4°C). The system was assumed to be in steady state after 4 volume changes (40 h), giving stable fermentation profiles and growth parameters (pH OD_{620}). Eight different experimental cultivation conditions were examined. The experimental variables used were: type of gas flushing (sparging or headspace flushing; 4 NL/h), type of flushing gas (N_2 or H_2), and stirring speed (300 or 50 rpm). For each condition two samples at different time points were taken for further analysis.

Samples used for transcriptome analysis were taken under low H_2 partial pressure (low P_{H_2}) and high P_{H_2} conditions. The low P_{H_2} condition was achieved by sparging the broth with N_2 using a stirring speed of 300 rpm, to minimize H_2 build-up in the broth. The high P_{H_2} condition was established by flushing the headspace with H_2 gas at a lower stirring speed of 50 rpm, to saturate the system with H_2. A low stirring speed was maintained to secure mixing and headspace flushing prevented CO_2 build-up in the system.

2.2. Analytical methods

Substrate and fermentation end-product concentrations were determined by HPLC, using a Shodex RSpak KC-811 ion exclusion chromatography column operating at 80 ^oC with a eluent of 3 mM H_2SO_4 (0.8 ml/min). Changes in the refractive index and UV absorption were detected by a Spectrasystem RI-150 and UV6000LP, respectively. Crotonic acid (10 mM) was added to the culture supernatant (16000xg, 10 min at 20^oC) as an internal standard in a 1:1 ratio to correct for differences in HPLC injection volumes. Concentrations were quantified using standard curves of the respective compounds. Cell dry weight (CDW) was used to quantify the amount of biomass in the bioreactor. CDW values were determined in technical duplicates. 2 x 10 ml culture was sampled and centrifuged (4800xg, 15 min at 4^oC). Each pellet was re-suspended in 2 ml ultrapure water. CDW values were determined after the samples had been incubated for 2 days in an oven at 120^oC. Additionally, optical cell densities were determined at 660 nm (OD_{660}).

2.3. Fermentation data analysis

Molar yields (mol/mol) were calculated using the biomass specific production and consumption rates (mmol produced (consumed) per g biomass per hour). CO_2 concentrations were calculated assuming that per mole of acetate or ethanol 1 mol of CO_2 is produced. H_2 concentrations were calculated assuming that per mol of acetate 2 mol of H_2 is produced. Biomass yields (g/mol) is expressed in grams of CDW per mol of consumed glucose (Y_{xs}). When calculating the biomass yield in grams of CDW per mol of ATP produced (Y_{ATP}) three assumptions were made: I) During the anaerobic oxidation of 1 mol of glucose to 2 mol of acetate, 6 mol of ATP are produced, II) 1 ATP is needed for glucose transport across the membrane via an ABC type transport system, and III) 2 ATP are invested in the first part of the glycolysis. Overall this results in the formation of 1.5 mol ATP per mol of acetate; since 1 mole ATP is formed during the acetate kinase step it follows that only 0.5 mol ATP is produced per mole of lactate or ethanol.

For each sample point a carbon balance (C-balance) and a balance of degree of reduction (γ-balance) was calculated [94]. To incorporate the biomass in these balances the elemental biomass composition of *C. saccharolyticus*, $CH_{1.62}O_{0.46}N_{0.23}S_{0.0052}P_{0.0071}$, was used [48], which corresponds to a carbon content of 40.63 mmol C per g CDW and a degree of reduction (γ) of 4.08 mol electrons per C-mole.

2.4. RNA isolation and microarray protocols

At each sample point (50 ml) culture broth from the bioreactor was rapidly cooled to 4°C and cells were harvested by centrifugation (4800xg, 20 min at 4 °C). Cell pellets were stored at -80°C. Total RNA was isolated using a modified Trizol (Invitrogen) protocol in combination with an RNAeasy kit (Qiagen). RNA quality was determined with an Experion bioanalyzer (Bio-Rad). The biological replicates were pooled prior to cDNA construction. cDNA was constructed using Superscript III reverse transcriptase (Invitrogen), random primers (Invitrogen), and the incorporation of 5-(3-aminoallyl)-2'-deoxyuridine-5'-triphosphate (Ambion, Austin, TX). The spotted whole-genome *C. saccharolyticus* microarray was developed [40] from unique 60-mer oligonucleotides based on at least 2679 ORFs. cDNA was labelled with either Cy3 or Cy5 dye (GE Healthcare, Little Chalfont, United Kingdom) using a dye-flip experimental design and hybridized to Corning Ultra-Gap II slides (Corning, Acton, MA). Microarray slides were scanned by using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA). GenePix Pro (6.0.1.25; Molecular Devices) was used to quantitate signal intensities. Output data was imported into JMP Genomics (v5.0, SAS, Cary, NC) for background correction, normalization across both arrays (Loess), and log transformation during ANOVA normalization which estimated global variation based on fixed (dye and treatment) and random (array and block) effects. Normalized data was used for row-by-row modelling (ANOVA) using the same fixed and random effects and the Holm multiple testing method to ascertain gene-specific variation. ORFs that met the Bonferroni statistical criterion (-log10(p-value) of 4.92) and were differentially transcribed 1.74-fold (log2 value 0.8) or more were considered to be regulated.

Additionally, genomic regions were taken into account during analysis, so a predicted operon was considered to be regulated if it contained at least one gene differentially transcribed 1.74-fold and all other genes in the cluster showed the same trend of regulation.

The microarray platform used in the present study is available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GPL6681. The raw data along with the final log2-fold changes have been deposited in the same database under accession number GSE40430.

2.5. Transcriptional regulation analysis

The involvement of the redox responsive repressor Rex (Csac_1220) in gene transcription regulation in *C. saccharolyticus* was investigated. Promoter/operator regions of *C. saccharolyticus* were scanned for putative Rex binding motifs with the Firmicutes/Clostridiaceae Rex Position Weight Matrix (PWM) as presented in the RegPredict database, using the RegPredict webtool (regpredict.lbl.gov, [114]). The gene set composed of the PWM scan hits was compared with a set containing genes which were upregulated under high P_{H_2}. This latter set also included the first genes of putative operons of which members showed the same trend of regulation and contained at least one gene significantly upregulated. Genes that were present in both sets were considered as putative targets for Rex mediated transcriptional regulation. In addition, the genomes of the fully sequenced *Caldicellulosiruptor* species: *C. hydrothermalis* 108, *C. kristjanssonii* 177R1B, *C. bescii Z-1320, C. kronotskyensis* 2002, *C. obsidiansis* OB47 and *C. owensensis* OL [57, 115, 116] were analyzed for the presence of orthologs of the genes identified as potential Rex targets. Their operator regions were manually inspected for the presence of possible Rex binding motifs.

Also the involvement of the ferric uptake regulator Fur (Csac_1741) and the fatty acid/phospholipid biosynthesis regulator FapR (Csac_1601) in gene transcription regulation was investigated. For this purpose promoter/operator regions of *C. saccharolyticus* were scanned for putative Fur and FapR binding motifs using the Thermotogales Fur and Staphylococcus FapR Position Weight Matrixes (PWMs). To determine if the *C. saccharolyticus* candidate binding sites and their corresponding operons are putative members of a regulon, PWM scans were performed using the regulon inference by known PWM webtool: RegPredict (regpredict.lbl.gov, [114]). The analysis was performed on *C. saccharolyticus* and several closely related organisms: *Clostridium phytofermentans* ISDg, *Thermoanaerobacter tengcongensis* MB4, *Thermoanaerobacter ethanolicus* X514, *Clostridium thermocellum* ATCC 27405 and *Clostridium cellulolyticum* H10. Transcription data was analyzed for signs of regulation of the identified *C. saccharolyticus* operons. The genomes of the fully sequenced *Caldicellulosiruptor* species were analyzed for the presence of orthologs of the regulated genes identified as potential Fur and FapR targets in *C. saccharolyticus,* and their operator regions were manually inspected for the presence of possible binding motifs.

Thresholds for the PWMs scans were adapted from RegPredict tool: Clostridiaceae Rex, 4.51; Thermotogales Fur, 5.42; Staphylococcus FapR 4.72. Upstream region sequences (position -400 to -1) were retrieved using the Regulatory Sequence Analysis Tools (RSAT) (rsat.ulb.ac.be).

Table 3.1 Fermentation details of *C. saccharolyticus* grown under eight different continuous cultivation conditions (dilution rate, 0.1 h^{-1}). Y_{ATP}, biomass yield in grams of CDW per mole of ATP produced. Y_{xs}, biomass yield per mol glucose.

Cultivation condition	Substrate and product concentrations						
	Measured					Calculated	
	mM						
	Glucose in	Glucose out	Lactate	Acetate	Ethanol	CO_2	H_2
N_2, sparging, 300rpm (low P_{H_2})	20.8 ± 0.5	0	0	30.9 ± 0.4	2.1 ± 0.1	32.9 ± 0.4	61.7 ± 0.8
N_2, sparging, 50rpm	19.7 ± 0.1	0	0.2 ± 0.1	28.1 ± 0.3	2.6 ± 0.1	30.6 ± 0.2	56.2 ± 0.6
N_2, headspace, 300rpm	20.4 ± 0.5	0	0	28.0 ± 0.6	2.8 ± 0.1	30.8 ± 0.6	55.9 ± 1.3
N_2, headspace, 50rpm	21.0 ± 0.3	2.1 ± 0.1	2.6 ± 0.1	17.8 ± 0.1	4.9 ± 0.0	22.7 ± 0.1	35.6 ± 0.9
H_2, sparging, 300rpm	20.8 ± 0.5	0	0	27.8 ± 0.2	4.0 ± 0.3	31.7 ± 0.1	55.6 ± 0.5
H_2, sparging, 50rpm	19.5 ± 0.1	0	1.3 ± 0.1	24.2 ± 0.1	3.6 ± 0.1	27.8 ± 0.2	48.4 ± 0.2
H_2, headspace, 300rpm	20.8 ± 0.5	0	0	24.4 ± 0.7	3.6 ± 0.1	28.0 ± 0.8	48.8 ± 1.4
H_2, headspace, 50rpm (high P_{H_2})	20.8 ± 0.5	0	3.9 ± 0.1	18.0 ± 0.3	5.3 ± 0.1	23.3 ± 0.3	36.0 ± 0.6

Cultivation condition	Specific consumption/production rates						Molar yields	
							Per glucose	
	mmol/g*h						mol/mol	
	q(Glu)	q(Lac)	q(Ace)	q(Eth)	q(CO_2)	q(H_2)	Y(Lac)	Y(Ace)
N_2, sparging, 300rpm (low P_{H_2})	4.02 ± 0.01	0	5.97 ± 0.10	0.40 ± 0.01	6.37 ± 0.10	11.94 ± 0.20	0	1.49
N_2, sparging, 50rpm	5.43 ± 0.06	0.05 ± 0.01	7.74 ± 0.16	0.71 ± 0.02	8.45 ± 0.15	15.48 ± 0.33	0.01	1.43
N_2, headspace, 300rpm	4.29 ± 0.01	0	5.88 ± 0.12	0.60 ± 0.01	6.47 ± 0.11	11.75 ± 0.23	0	1.37
N_2, headspace, 50rpm	5.47 ± 0.14	0.76 ± 0.05	5.15 ± 0.07	1.42 ± 0.01	6.57 ± 0.08	10.30 ± 0.14	0.14	0.94
H_2, sparging, 300rpm	4.82 ± 0.07	0	6.46 ± 0.14	0.92 ± 0.06	7.38 ± 0.09	12.92 ± 0.29	0	1.34
H_2, sparging, 50rpm	6.38 ± 0.35	0.43 ± 0.02	7.93 ± 0.46	1.19 ± 0.11	9.12 ± 0.57	15.86 ± 0.92	0.07	1.24
H_2, headspace, 300rpm	5.15 ± 0.09	0	6.05 ± 0.06	0.90 ± 0.01	6.95 ± 0.07	12.11 ± 0.13	0	1.18
H_2, headspace, 50rpm (high P_{H_2})	5.96 ± 0.10	1.13 ± 0.01	5.18 ± 0.17	1.53 ± 0.02	6.70 ± 0.12	10.35 ± 0.34	0.19	0.87

[a]SUM Yield, the sum of the lactate, acetate and ethanol yields.

Cultivation condition	ATP		Biomass		
	Total	Y_{ATP}			
	mmol	g/mol		g/L	g/mol
			OD_{660}	CDW	Yxs
N_2, sparging, 300rpm (low P_{H_2})	47.3 ± 0.6	10.4 ± 0.2	1.14 ± 0.01	0.49 ± 0.01	23.7 ± 0.1
N_2, sparging, 50rpm	43.5 ± 0.3	8.3 ± 0.2	0.95 ± 0.05	0.36 ± 0.01	18.4 ± 0.2
N_2, headspace, 300rpm	43.4 ± 0.9	10.4 ± 0.2	0.99 ± 0.01	0.45 ± 0.01	22.1 ± 0.1
N_2, headspace, 50rpm	30.5 ± 0.1	10.8 ± 0.2	0.88 ± 0.01	0.33 ± 0.01	17.4 ± 0.4
H_2, sparging, 300rpm	43.7 ± 0.2	9.4 ± 0.2	0.99 ± 0.01	0.41 ± 0.01	19.7 ± 0.3
H_2, sparging, 50rpm	38.8 ± 0.2	7.9 ± 0.5	0.85 ± 0.01	0.31 ± 0.02	15.7 ± 0.9
H_2, headspace, 300rpm	38.4 ± 1.1	10.0 ± 0.1	0.93 ± 0.01	0.38 ± 0.01	18.5 ± 0.3
H_2, headspace, 50rpm (high P_{H_2})	31.7 ± 0.4	10.5 ± 0.3	0.88 ± 0.01	0.33 ± 0.01	16.0 ± 0.3

Cultivation condition					C-balance	γ-balance
					Recovery	
					%	
	Y(Eth)	SUM Yields[a]	$Y(CO_2)$	$Y(H_2)$		
N_2, sparging, 300rpm (low P_{H_2})	0.10	1.59	1.59	2.97	95.3 ± 0.9	95.6 ± 0.9
N_2, sparging, 50rpm	0.13	1.56	1.56	2.85	90.6 ± 0.2	91.9 ± 0.2
N_2, headspace, 300rpm	0.14	1.51	1.51	2.74	90.4 ± 1.5	90.7 ± 1.5
N_2, headspace, 50rpm	0.26	1.34	1.20	1.88	81.0 ± 0.6	81.2 ± 0.6
H_2, sparging, 300rpm	0.19	1.53	1.53	2.68	89.9 ± 0.3	90.1 ± 0.3
H_2, sparging, 50rpm	0.19	1.50	1.43	2.48	85.4 ± 0.1	85.6 ± 0.1
H_2, headspace, 300rpm	0.17	1.35	1.35	2.35	80.1 ± 2.0	80.3 ± 2.0
H_2, headspace, 50rpm (high P_{H_2})	0.26	1.31	1.13	1.74	76.5 ± 0.2	76.7 ± 0.2

3. Results

3.1. Fermentation profiles

C. saccharolyticus was grown in a continuous cultivation setup (D=0.1 h^{-1}) under eight different conditions. The cumulative effects of stirring speed (300 or 50 rpm), type of reactor flushing (sparging or headspace flushing) and applied flushing gas (N_2 or H_2) on the fermentation performance were investigated (**Table 3.1**). Stable fermentation profiles and growth parameters indicated that the systems were in steady state when the samples were taken (data not shown).

Compared to the low P_{H_2} condition [N_2, sparging, 300 rpm] all other investigated cultivation conditions negatively affected fermentation performance, resulting in lower acetate and H_2 yields. During all cultivations glucose (4 g/L) was mainly converted to acetate, H_2 and CO_2, with acetate concentrations ranging between 17.8 and 30.9 mM. H_2 yields ranged from 1.7 to 3.0 mol/mol glucose for the high P_{H_2} and low P_{H_2}, respectively. Ethanol was formed during all cultivations (2.1 - 5.3 mM), with the lowest ethanol concentration detected in the low P_{H_2} condition [N_2, sparging, 300 rpm]. Lactate (0.2-3.9 mM) was only formed under low stirring speed (50 rpm) conditions, with only marginal amounts observed for condition [N_2, sparging, 50 rpm]. Cultivations were glucose limited except for condition [N_2, headspace, 50 rpm], where residual glucose was detected in the culture supernatant. Residual glucose was also occasionally detected in high P_{H_2} cultivation conditions [H_2, headspace, 50 rpm] as, for example, shown in **Figure 3.2A**. This indicates that headspace flushing in combination with a low stirring speed results in a different limitation than the imposed substrate limitation and is independent of the type of flushing gas used, though, the effect is more pronounced when using N_2 gas.

Using a pairwise comparison between conditions, where only one parameter changed and the others were kept the same, several observations could be made. Carbon recovery (**Figure 3.1**) was always lower under relatively high P_{H_2} conditions, i.e. under headspace flushing conditions compared to sparging. Likewise, carbon recovery was always lower when H_2 was used as a flushing gas compared to N_2 gas, although fermentation patterns were relatively similar. The carbon recovery and biomass levels were always lower under low stirring speed conditions (50 rpm) with respect to the high stirring speed (300 rpm) conditions. Also CDW/OD ratios were lower at low stirring speed (~0.37±0.01) compared to high stirring speed (~0.43±0.02), which could reflect a change in cell morphology.

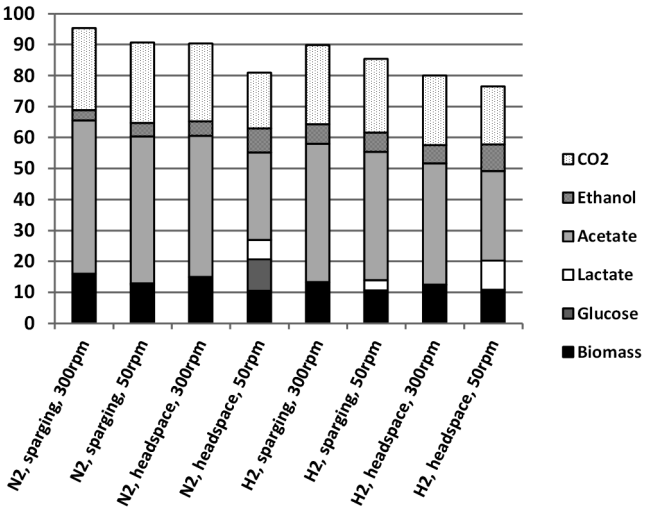


Figure 3.1 Carbon balance for eight chemostat cultivations. The differences in cultivation conditions are indicated on the x-axis, and encompass: gas used for flushing (N_2/H_2), type of flushing (headspace/sparging) and stirring speed (300/50 rpm). The carbon recovery for a product is expressed as the percentage (%) of the total carbon content of the consumed glucose (= 100 %) in each condition.

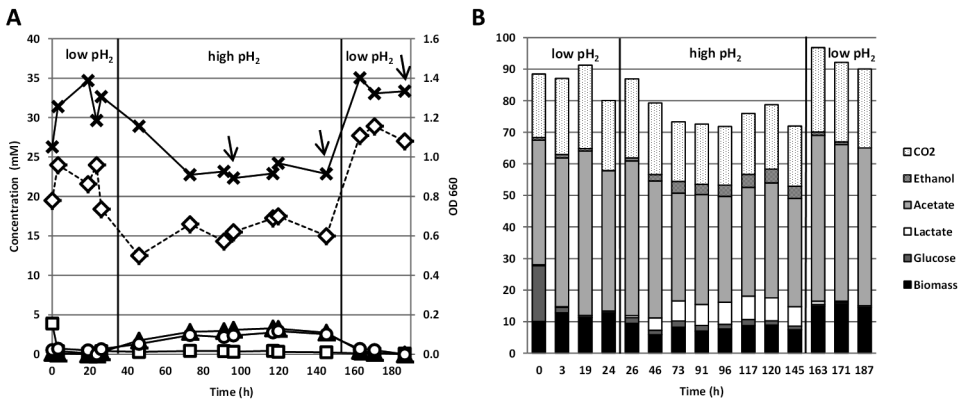


Figure 3.2 Fermentation profile and carbon balance of *C. saccharolyticus* growing in a continuous culture under low and high P_{H_2}. Low P_{H_2}: broth flushed with N_2 (4L/h), stirring speed 300rpm. High P_{H_2}: headspace flushed with H_2 (4L/h), stirring speed 50rpm. (a) Fermentation profile. Arrows indicate sampling points for transcriptome analysis. Residual glucose (□), ethanol (○), lactate (▲), acetate (**X**) and optical density (◊, secondary axis); (b) The carbon balances at different time points (x-axis). The carbon recovery for each compound is expressed as the percentage (%) of the total carbon content of the consumed glucose (= 100 %) at each sampling point.

The balances of degree of reduction showed a similar pattern as the carbon balances (**Table 3.1**). Although differences in biomass levels were observed between the high and low P_{H_2} condition, these could not account for the decreases in carbon recovery. These findings suggest that a product or products with a similar reduction state as glucose (γ = 4 electrons per C atom) are still missing in the balances. The product or products must have been excreted from the cell, else one would expect higher biomass levels (CDW). Overall, the largest deficit in the C-balance and the highest lactate and ethanol yields were observed under high P_{H_2} condition [H_2, headspace, 50 rpm].

For the analysis of the effect of P_{H_2} on the transcription level, *C. saccharolyticus* was grown using the extremes of the different growth conditions, viz. the high P_{H_2} [H_2, headspace, 50 rpm] and low P_{H_2} [N_2, sparging, 300 rpm] condition. The fermentation profiles and carbon balance of this continuous cultivation are displayed in **Figure 3.2A** and **Figure 3.2B**, respectively. During the initial low P_{H_2} phase [N_2, sparging, 300 rpm] glucose is mainly converted to acetate, and an OD of ~1.0 was reached. At t=0 h there is still some residual glucose from the startup batch cultivation, but is consumed within 3 h. After the switch to the high P_{H_2} condition: [H_2, headspace, 50 rpm] a clear shift in the fermentation profile was observed. A decrease in OD (~0.65) coincided with a decreased acetate formation and an increased ethanol and lactate formation. Switching back to the low P_{H_2} condition resulted in similar fermentation profiles as during the initial low P_{H_2} phase, indicating that the bacterium completely recovered from the high P_{H_2} condition. Only the optical density was slightly increased to an OD of ~1.1. Although very small amounts of oxaloacetate and pyruvate (below 1 mM) were detected under high P_{H_2} conditions they do not significantly contribute to the carbon recovery. Additionally, no other compounds (organic acids, alcohols or amino acids) could be detected via HPLC and analysis of the supernatant did not reveal a relative increase in protein level under high P_{H_2} (data not shown). These findings suggest that the unaccounted carbon might be distributed over multiple end-products, whose concentrations are below detection limits and/or ends up in one or more compound(s) that cannot be detected by HPLC.

3.2. Transcriptome analysis

Changes in transcription (mRNA) levels of cells grown under high P_{H_2} with respect to a low P_{H_2} were investigated by microarray analysis. For this purpose two samples were taken under high P_{H_2} [H_2, headspace, 50 rpm] cultivation condition (**Figure 3.2A**) and two samples under low P_{H_2} [N_2, sparging, 300 rpm] conditions. One low P_{H_2} sample was taken after the condition changed from high P_{H_2} to low P_{H_2} and, as an additional biological duplicate for the low P_{H_2} condition, cells were harvested from

another chemostat culture under identical conditions and comparable fermentation profile (data not shown).

Overall the differences in transcript levels between the two conditions were moderate. 131 ORFs could be considered to be up- or downregulated (**Supplementary Table S3.A**), where upregulation indicates a relative increase of transcript level under the high P_{H_2} cultivation condition with respect to the low P_{H_2} condition. During the analysis the genomic neighborhood was also taken into account. An overview of the regulated genes/gene clusters can be found in **Supplementary Table S3.B & Supplementary Table S3.C**. The genes/gene clusters presented in **Supplementary Table S3.B** are discussed below with respect to their physiological function in *C. saccharolyticus*.

3.2.1. Upregulation under high P_{H_2} versus low P_{H_2} conditions

3.2.1.1. Carbon metabolism

Transcription levels of the genes involved in the upper part of the glycolysis, the conversion of glucose to glyceraldehyde-3-phosphate, did not reveal any signs of regulation. Genes involved in the lower part of the glycolysis, from glyceraldehyde-3-phosphate to pyruvate, are clustered together (locus tags: Csac_1951-1955) and, although the differences in transcript level for these genes did not meet the 1.74-fold criterion (log2 value 0.8), most of them showed the same trend of upregulation. This cluster includes pyruvate phosphate dikinase, an alternative enzyme to catalyze the conversion of phosphoenolpyruvate to pyruvate besides the pyruvate kinase (Csac_1831) which does not seem to be regulated. All subunits of the pyruvate:ferredoxin oxidoreductase (Csac_1458-1461) enzyme complex, responsible for the conversion of pyruvate to acetyl-CoA, were moderately upregulated at high P_{H_2}. Also both genes (Csac_2040-2041) involved in the acetyl-CoA conversion to acetate were upregulated.

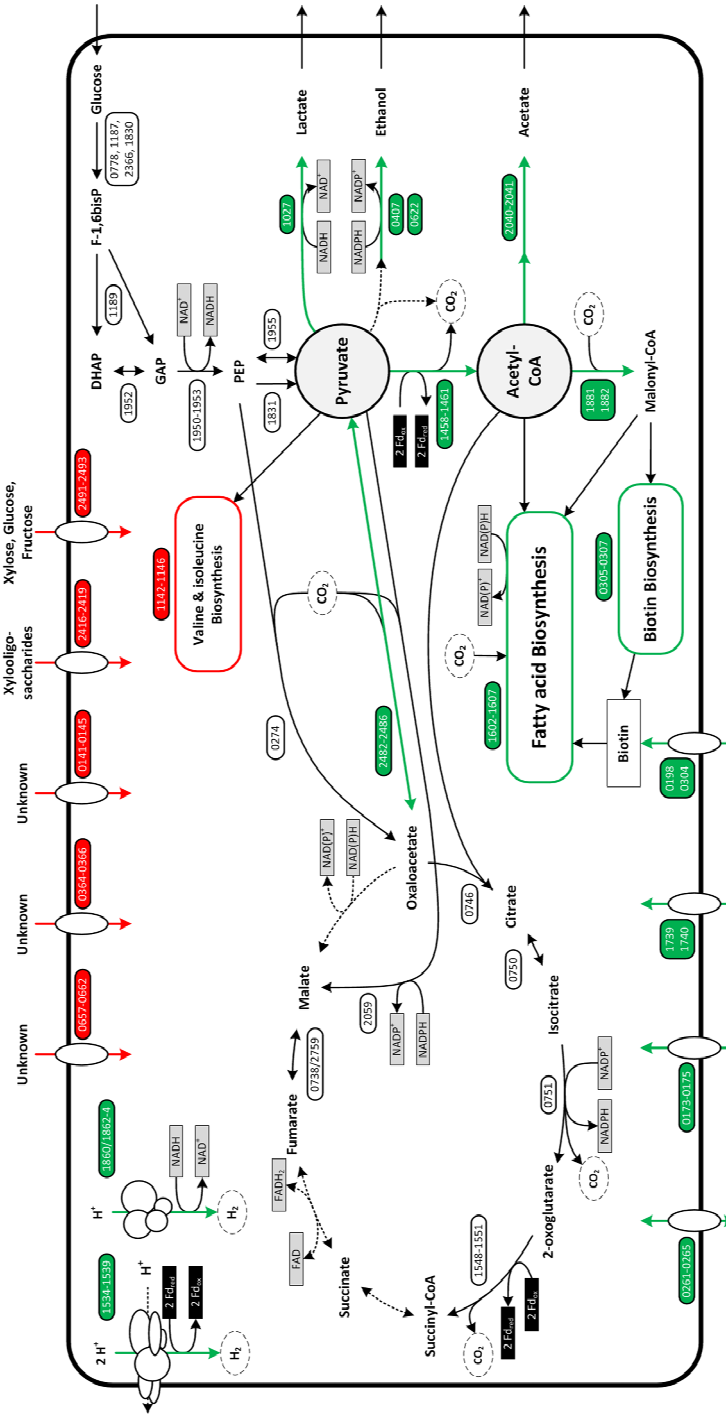

Glucose
F-1,6bisP
0778, 1187, 2366, 1836
DHAP
1189
1952
GAP
1950-1953
PEP
1955
1831
Pyruvate
Lactate
1027
NADH
NAD+
Ethanol
NADPH
NADP+
0407
0622
Acetate
2040-2041
Acetyl-CoA
CO2
1881
1882
Malonyl-CoA
Biotin Biosynthesis
0305-0307
Biotin
0198
0304
Biotin
Fatty acid Biosynthesis
1602-1607
NAD(P)+
NAD(P)H
Valine & isoleucine Biosynthesis
1142-1146
2 Fdred
2 Fdox
1455-1460
0274
2482-2486
Oxaloacetate
Malate
0738/2759
Fumarate
FADH2
FAD
Succinate
Succinyl-CoA
1548-1551
2059
NADP+
NADPH
0746
Citrate
0750
Isocitrate
0751
2-oxoglutarate
1739
1740
Fe2+
Fe3+-Siderophore
0173-0175
Fructose, Sucrose
0261-0265
Xylose, Glucose, Fructose
2491-2493
Xylooligo-saccharides
2416-2419
Unknown
0141-0145
Unknown
0364-0366
Unknown
0657-0662
H+
3860/1862-4
2 H+
1534-1539
H2

Figure 3.3 Schematic overview of the results of the transcriptome analysis performed on *C. saccharolyticus* growing in a continuous culture under low and high P_{H_2}. Low P_{H_2}: broth flushed with N_2 (4L/h), stirring speed 300rpm. High P_{H_2}: headspace flushed with H_2 (4L/h), stirring speed 50rpm. In green: metabolic pathways and transport systems upregulated under high P_{H_2} versus low P_{H_2} In red: metabolic pathways and transport systems downregulated under high P_{H_2} versus Low P_{H_2}. Numbers in boxes represent the locus tags of the reaction associated genes. Dotted arrows: reaction associated genes were not identified in *C. saccharolyticus*.

3.2.1.2. Hydrogenases and alternative electron sinks

In *C. saccharolyticus* two gene clusters have been predicted to code for an NADH-dependent- and a ferredoxin-dependent hydrogenase (Csac_1860/1862-1864 and Csac_1534-1539, respectively) [40]. Both clusters, including the ferredoxin-dependent hydrogenase maturation genes, were upregulated under the high P_{H_2} condition, although difference in transcript levels for the NADH-dependent-hydrogenase genes were slightly below the chosen log2 threshold value. In agreement with the observed fermentation profile, the annotated L-lactate dehydrogenase (Csac_1027) and two iron-dependent alcohol dehydrogenases (Csac_0407, Csac_0622) were upregulated. These enzymes catalyze reactions acting as electron sinks (**Figure 3.3**). Interestingly, an L-lactate permease (Csac_0583, log2 value -0.76) was downregulated.

The reductive part of the TCA cycle, leading to succinate, can potentially act as an alternative electron sink. However, the genes coding for a malate dehydrogenase and fumarate reductase, necessary for this metabolic pathway, were not detected in the genome [40] (**Figure 3.3**). Malate might also be formed directly from pyruvate via the NADP-dependent malic enzyme (oxaloacetate decarboxylating), but the predicted gene (Csac_2059) showed no sign of regulation nor did the genes encoding a fumarase (Csac_0738 & Csac_2759). Furthermore, succinate was not detected (HPLC) in the culture supernatant. For *C. saccharolyticus* carbon exchange between the EM-pathway and the TCA cycle could also be catalyzed by a phosphoenolpyruvate carboxykinase (GTP) (Csac_0274) or an oxaloacetate decarboxylase (EC 4.1.1.3, coupled to Na^+-translocation) (Csac_2482-2485), although both enzymes are usually associated with gluconeogenesis. The oxaloacetate decarboxylase gene cluster shows a clear upregulation under high P_{H_2} conditions.

Csac_1883 codes for an oxaloacetate decarboxylase alpha subunit as well (62% AA sequence identity with Csac_2485), but it clusters with genes coding for an acetyl-CoA carboxylase and a biotin carboxyl carrier protein (Csac_1881-1882); all three genes are upregulated. Acetyl-CoA carboxylase (EC 6.4.1.2) catalyzes the conversion of acetyl-CoA to malonyl-CoA, both compounds are the precursors for fatty acid biosynthesis (**Figure 3.3**).

Gene cluster (Csac_1601-1607) is coding for several proteins involved in fatty acid biosynthesis and its regulation. Including FapR, a fatty acid and phospholipid biosynthesis regulator, PslX, a protein involved in fatty acid and/or phospholipid synthesis, and an acyl carrier protein. The cluster has a similar gene arrangement as the *Escherichia coli* K12 operon (regulonDB, http://regulondb.ccg.unam.mx/) and was upregulated under high P_{H_2}. Fatty acid biosynthesis acts both as a carbon (Acetyl-CoA and CO_2) and electron sink (NADH and NADPH).

3.2.1.3. Biotin biosynthesis and transport

Biotin is an essential cofactor in enzymatic carboxylation, decarboxylation, and transcarboxylation reactions, which include the previously discussed oxaloacetate decarboxylase and acetyl-CoA-carboxylase. Genes involved in biotin transport and biosynthesis from the precursor 7-keto-8-aminopelargonate (Csac_0304-0307/0198) were upregulated under high P_{H_2} conditions. The *de novo* synthesis of 7-keto-8-aminopelargonate from malonyl-CoA involves several steps catalysed by enzymes also participating in the fatty acid biosynthesis. However, the final step in 8-amino-7-oxononanoate synthesis, catalyzed by 8-amino-7-oxononanoate synthase (Csac_1225), appears to be downregulated (Log2 value -1.31). A *bioY* gene (Csac_0304) is, likewise in other clostridia [117], located at a locus encoding biotin biosynthesis. Studies of the *bioY* gene from proteobacterium *Rhodobacter capsulatus* revealed BioY to be a high-capacity transporter, which converts into a high-affinity system in the presence of BioMN [117]. However, the *bioMNY* operon is absent in *C. saccharolyticus*. The second *bioY* gene (Csac_0198) is not clustered with any biotin related genes, and the bifunctional biotin-[acetyl-CoA-carboxylase] ligase/biotin operon repressor (Csac_0109) [118] showed no signs of regulation.

3.2.1.4. Fe^{2+}, Fe^{3+} and Co^{2+} -transport

Two of the four iron uptake related transport systems present in *C. saccharolyticus* were upregulated. The first is a ferrous ion transporter (Csac_1739-1741), with similar gene arrangements as the *feoABC* operon in *E. coli* K-12 genome (regulonDB). The second iron transport system is an ABC-type Fe^{3+}-siderophore transporter (Csac_0173-0175). Additionally, several other transport related gene clusters were upregulated. *CbiMNQO* (Csac_2682-2685) encodes the widespread ECF-type Co^{2+} transporter system, where CbiMN was shown to function as the core transporter [119].

3.2.2. Downregulated under high P_{H_2} versus low P_{H_2} conditions

3.2.2.1. Carbon metabolism and biosynthesis

Some enzymatic steps involved in carbon catabolism were downregulated. Including a predicted L-fucose-1-P aldolase (Csac_0425) and a gene cluster (Csac_2051-2054) coding for a phenylacetate-CoA ligase and an indolepyruvate:ferredoxin oxidoreductase (IOR). The IOR has been shown to be involved in peptide fermentation [120]. Downregulation of several key enzymes in biosynthesis pathways of valine, leucine and isoleucine could be observed, the genes coding for acetolactate synthase (EC 2.2.1.6), ketol-acid reductoisomerase (EC 1.1.1.86) and 2-isopropylmalate synthase (EC 2.3.3.13) are all located within the same gene cluster (Csac_1142-1146). Additionally, phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21, Csac_1629-1630), which is involved in the formation of inosine monophosphate (IMP) from D-ribose 5-P was downregulated. IMP is a precursor for the *de novo* biosynthesis of the guanosine and adenosine nucleotides.

3.2.2.2. Transport systems

A large variety of transport systems are downregulated. This included some sugar transport systems proposed to be specific for xylooligosaccharides (Csac_2416-2619), fructose/sucrose (Csac_0261-0265) and xylose/glucose/fructose (Csac_2491-2494) [41] and a highly regulated ABC-transporter related gene cluster (Csac_0140-0145) with unknown function. An ABC-type transporter (Csac_0657-0662), potentially involved in antimicrobial peptide transport, and a potential bacteriocin ABC-transporter (Csac_0559-0567), were downregulated. The ABC-2 type transport system (Csac_0364-0366) and a type II secretion system (Csac_1060-1063), coding for pilus retraction protein PilT, are both associated with gliding motility [121]. Finally, the gene coding for a facilitator (Csac_1517 (log 2 value -0.93)) was downregulated, however, its specificity is unknown.

3.2.3. Transcription regulators and genes with unknown function

Several transcription or stress response regulators were regulated under high P_{H_2} conditions. This included LacI family regulatory protein Csac_1228 (-0.75), a putative anti-sigma regulatory factor, serine/threonine protein kinase Csac_0346 (1.53), a HTH-type transcriptional regulator Csac_0140 (-0.90) and a translation elongation factor G Csac_2373 (-1.04). The potential targets of these regulators are unknown. Additionally, some gene clusters for which the role of their gene products in the

cellular processes of *C. saccharolyticus* is rather unclear show signs of regulation (**Supplementary Table S3.C**). This includes gene clusters coding for proteins potentially involved in the CRISPR-Cas immunity systems, DNA repair, sporulation or cell division (Csac_2394-2401, Csac_1451-1455, Csac_0484-0486, Csac_1897-1911, Csac_1961-1967, Csac_1865-1873). Sporulation was never observed for *C. saccharolyticus*, but cell morphology might be affected by the change in cultivation condition.

3.3. Transcriptional regulation

The involvement of the redox responsive repressor Rex (Csac_1220), iron uptake regulator Fur (Csac_1741) and the fatty acid biosynthesis regulator FapR (Csac_1601) in transcriptional regulation was investigated in *C. saccharolyticus* grown under high P_{H_2} conditions.

3.3.1. Redox responsive repressor, Rex

Rex is a transcriptional regulator sensing changes in the intracellular ratio of NADH/NAD^+ [122]. Putative Rex binding motifs have been observed upstream of genes whose associated operons where significantly upregulated under high P_{H_2} (**Table 3.2**). Orthologs of these given genes are present in all the investigated *Caldicellulosiruptor* species, except for Csac_1452, Csac_2485 and the operons coding for the alcohol dehydrogenases (Csac_0407, Csac_0622) (**Supplementary Table S3.D**). Orthologs of the *C. saccharolyticus* oxaloacetate decarboxylase (Csac_2485) are absent in the *C. bescii* and *C. obsidiansis*. While only *C. hydrothermalis*, *C. bescii* and *C. kronotskyensis* contain an ortholog of the hypothetical gene Csac_1452. And only *C. kristjanssonii* and *C. hydrothermalis* contain an ortholog of the alcohol dehydrogenase gene, Csac_0407 (Calkr_0097, Calhy_0244) but none of the investigated species contains an ortholog of the alcohol dehydrogenase gene, Csac_0622. Additionally, *C. owensensis* contains only a Rex pseudogene.

For all the investigated orthologs putative Rex regulator binding motif could be identified, except for Csac_1741 and Csac_1452, those orthologs did not contain a putative Rex binding motif (**Supplementary Table S3.D**). This suggests that both *C. saccharolyticus* genes are probably not a part of the Rex regulon and motif resemblance to the Rex binding motif is coincidental.

All the identified motifs have identical sequences with respect to the motifs in *C. saccharolyticus*, except the motif of the *C. kronotskyensis* ortholog of Csac_0407 and the *C. hydrothermalis* ortholog of Csac_1861, which contain a single base substitution and only the *C. owensensis* ortholog Csac_2485 motif contains 3 substitutions

(**Supplementary Table S3.D**). Overall this data reveals that the Rex binding motifs are highly conserved in other *Caldicellulosiruptor* species.

Table 3.2 Putative binding motifs of the Rex transcriptional regulator protein identified in *C. saccharolyticus*, based on transcriptional regulation. +: under high P_{H_2}, significantly upregulated genes are present in the putative operon. ns: no significantly regulated genes were identified in the putative operon. Firmicute related Rex consensus sequences from literature and the putative Rex binding motifs consensus from *C. saccharolyticus* (weblogo, weblogo.berkeley.edu,), based on these binding motifs, are given.

Locus tag		Putative Rex binding motif	Annotation/Function
Csac_1220[a]	ns	ATGTCAAAAAAATATCAA	Redox-sensitive transcriptional regulator Rex
Csac_0622[a]	+	ATGTTAATTTTTTAACAG	Iron-containing alcohol dehydrogenase
Csac_0407[a]	+	ATGTTAAAATTATAACAA	Iron-activated alcohol dehydrogenase
Csac_2425[a]	+	ATGTTATAATATTAACAA	Putative transcriptional regulator, CopG family
Csac_1534[a]	+	TTGTTAACAAATTAACTT	Ech hydrogenase, subunit EchA
Csac_1860[a]	+	ATGTTAAATTTCTAACAA	Hyd hydrogenase, subunit hydC
Csac_1861	+	TTGTCAAGATATTAACTT	Histidine kinase
Csac_2485	+	TTGTTAAAAAATTGACCA	Oxaloacetate decarboxylase alpha chain
Csac_1458	+	ATGTTGATTTTTTAACAT	Pyruvate oxidoreductase, gamma subunit
Csac_1452[b]	+	AAGTTGTACTTATAACAT	Hypothetical protein
Csac_1741[b]	+	AAGTTATAATTTTAACTA	Ferric uptake regulator, Fur family
		(A)TTGTTAnnnnnnTAACAA(T)	[123] (*T. eth*)
		TTGTTwwwTTwTTAACAA	[124](*C. ace*)
		tTGtTAAanantTAACAa	RegPredict [114](*Clostridiaceae*)
		tTGntaaannntTaACaa	[125](*Thermoanaerobacterales*)
		aTGTtAAaatatTAACaa	This study (*C.sac*)

[a] Member of the *C. saccharolyticus* Rex regulon (RegPrecise database, regprecise.lbl.gov).
[b] Sequences were not included for *C. saccharolyticus* Rex consensus motif.

3.3.2. Iron uptake regulator, Fur

Fur is an iron uptake regulator sensitive for [Fe^{2+}]. High Fe^{2+} levels within the cell cause Fur to bind with the corresponding promoter region; when iron levels drop Fe^{2+} dissociates from the Fur protein which then detaches from the promoter, permitting transcription [126]. A Fur regulator protein (FeoC, Csac_1741) binding motif, similar to Fur *Thermotogales* motif, could be identified in the promoter region of *feoA*

(Csac_1739) (**Table 3.3**). The *feoABC* gene cluster is upregulated under high P_{H_2} conditions as discussed above (**Supplementary Table S3.B**). All the other investigated *Caldicellulosiruptor* species contained a *feoA* (Csac_1739) ortholog. The *C. saccharolyticus* and *C. owensensis* FeoC binding motifs contained a single base substitution with respect to the other motifs (**Supplementary Table S3.D**). In addition, results of the regulon inference by known PWM analysis (**Table 3.3**) revealed the presence of putative Fur binding sites located in the operator region of the FeoA family protein coding gene in *T. tengcongensis* MB4, *T. ethanolicus* X514 and *C. thermocellum* ATCC 27405. These results indicate that the FeoABC operon (Csac_1739-1741) is a likely member of the Fur regulon.

3.3.3. Fatty acid biosynthesis regulator, FapR

FapR acts as a repressor of the fatty acid biosynthesis [127]. A binding motif similar to the FapR *Staphylococcus* motif, could be identified in the upstream region of the first gene (*FapR*, Csac_1601) of the fatty acid biosynthesis operon and the acetyl-CoA carboxyl transferase gene (Csac_1881) (**Table 3.3**). Both the fatty acid biosynthesis operon and the acetyl-CoA carboxyl transferase gene cluster were upregulated under high P_{H_2} conditions (**Supplementary Table S3.B**). All the other investigated *Caldicellulosiruptor* species contained a *FapR* (Csac_1739) and acetyl-CoA carboxyl transferase (Csac_1881) ortholog, which binding motifs were all identical to the respective *C. saccharolyticus* motifs (**Supplementary Table S3.D**). Potential FapR binding sites could also be identified in the operator region of the *FapR* and acetyl-CoA carboxyl transferase encoding genes in *C. thermocellum* ATCC 27405 and *C. cellulolyticum* H10. These results indicate that both the fatty acid biosynthesis (Csac_1601-1607) and the acetyl-CoA carboxyl transferase (Csac_1881-1883) operons belong to the FapR regulon in *C. saccharolyticus*.

4. Discussion

The thermophilic bacterium *Caldicellulosiruptor saccharolyticus* is well known for its capacity to produce high levels of hydrogen with yields almost reaching 4 mol of H_2 per mol of hexose [48, 110]. This feature combined with the capacity to degrade a broad range of substrates, including cellulose, have led to its selection as a model organism for studying sustainable biohydrogen production [40, 43, 46, 48]. Nevertheless, depending on the partial hydrogen pressure other reduced end-products like lactate and ethanol are produced as well. The effect of P_{H_2} on fermentation performance has been the subject of various investigations [50, 112, 113], including the regulatory mechanism behind the disposal of reductants via H_2,

lactate or ethanol at the metabolic level [51, 113]. Although the P_{H_2} was suggested to trigger a regulatory response in the case of '*Thermoanaerobacter tengcongensis*' [22], analysis of the effect of the P_{H_2} on gene expression has never been performed in fermentative microorganisms in general, or a H_2 producer in particular.

4.1. Effect of P_{H_2} on fermentation performance

Different gasses (N_2 or H_2), gassing systems (sparging or flushing) and stirring speeds (50 rpm or 300 rpm) were exerted on *C. saccharolyticus* cultures grown in a chemostat setup, with the aim to generate different dissolved H_2 concentrations. For example, headspace flushing combined with a low stirring speed decreases the liquid gas surface interaction, which could lead to an accumulation of H_2 in the liquid phase, subsequently inhibiting the hydrogenase activity and thus inflicting changes in metabolism and gene expression.

Growth inhibition of *C. saccharolyticus* caused by H_2 was demonstrated to depend on the dissolved H_2 concentration [112]. Moreover, when mass transfer is limiting, supersaturation of the medium may occur [112, 128]. Indeed, the high P_{H_2} condition (headspace flushing with H_2 and a low stirring speed) led to supersaturation of the liquid phase, with a dissolved H_2 concentration of 1.49 ± 0.10 mM (unpublished data, AAM Bielen). This concentration exceeds the theoretical maximal H_2 solubility of 0.72 mM (Ostwald coefficient (70°C): 0.02045, V_m (70°C): 28.53 L/mol) [129] and is reaching the critical dissolved H_2 concentration of 2.2 mmol/L identified by Ljunggren *et al.* [112].

When switching the reactor from a low P_{H_2} to a high P_{H_2} condition fermentation is affected, leading to the reduced end-products (lactate, ethanol) and decreased acetate and H_2 yields (**Figure 3.2A**, **Table 3.1**). Also the carbon balance dropped from ~90% to ~75% (**Figure 3.2B**). A comparable drop was observed for the balance of degree of reduction. These data suggest that under high P_{H_2} conditions, a reduced end-product is missing from the balance equations. However, a detailed analysis of the growth medium, using different HPLC systems, did not reveal any obvious fermentation end-products like alanine, succinate, malate or formate. Increased cell lysis may explain the gap in the balance, but total protein determinations of the spent medium could not substantiate this. Low molecular compounds like pyruvate, malate, soluble glucans and extracellular amino acids were proposed to account for the poor recovery for *Clostridum thermocellum*, where more than 10% of the carbon was missing [130]. The fact that we did not observe discernible levels of specific compounds could indicate that carbon is distributed over multiple end-products which are below our detection levels. Moreover, although the off-gas is led through a condenser, the loss of volatile end-products cannot be completely excluded.

Although, we used different gasses in this experiment, we could show that the stirring speed, combined with flushing type are the determining factors for the fermentation profile switch. In other words, sparging with H_2 is almost as efficient as sparging with N_2.

4.2. Effect of P_{H_2} on gene transcription

The effect of a high P_{H_2} cultivation condition on *C. saccharolyticus* was studied at the gene expression level. Overall, the differences in transcript levels between the high and low P_{H_2} condition were moderate, as compared to earlier results obtained from batch cultures fermenting various sugars [40, 41]. This may be a result of the fact that the chemostat cultures were carbon-limited, thereby preventing excessive build-up of reduced electron carriers. Such a build-up of reduced electron carriers is expected to be the consequence of the inhibition of hydrogenase activity caused by the increased dissolved H_2 concentration. For instance, Willquist et al reported NADH/NAD ratios of only ~0.12 in C-limited chemostats [113], whereas in batch cultures this ratio could amount to ~1.2 [51].

Nevertheless, the genes that showed a significant difference in expression were predominantly coding for enzymes involved in the disposal of reductant (**Figure 3.3**). The upregulation of the LDH and ADH genes corresponds to the observed change in end-product profile toward lactate and ethanol formation under high P_{H_2}. However, upregulation of both the ferredoxin-dependent (*ech*, [NiFe]) and the NADH-dependent (*hyd*, Fe-only) hydrogenases appears counter intuitive, since an increase in H_2 formation is not expected under high P_{H_2} and increased hydrogenase enzyme levels will not lead to an increase of proton reduction turnover rates. But apparently, *C. saccharolyticus* reacts to the change in redox state caused by the elevated P_{H_2} by facilitating all possible redox sinks, which includes both hydrogenases.

The upregulation of several iron transport systems under high P_{H_2} conditions corresponds to the upregulation of the hydrogenases and the ADHs, since iron is an important moiety of the active site of both *C. saccharolyticus* hydrogenases [131] and is also required for the iron-dependent ADH activity. In *C. saccharolyticus* the transcription regulation of the iron transport coding operon (*feoABC*) is under control of the iron uptake regulator Fur (**Table 3.3**).

The genes responsible for ethanol formation in *C. saccharolyticus* formation were not known, but ORFs Csac_0407 and Csac_0622 were annotated as iron-containing alcohol dehydrogenases. For Csac_0407 high homology exists with genes coding for lactaldehyde reductase/propanediol dehydrogenase and increased transcription level were observed during growth on rhamnose versus glucose, suggests that this alcohol dehydrogenase may also have a function in rhamnose catabolism [40]. Furthermore,

this alcohol dehydrogenase closely resembles the C-terminal alcohol dehydrogenase domain of a bifunctional acetaldehyde dehydrogenase-alcohol dehydrogenase (AAD), that is assumed to be responsible for ethanol formation in many other thermophilic bacteria like, *Thermoanaerobacter pseudoethanolicus* E39 and *T. tengcongensis* [22, 132, 133]. *C. saccharolyticus* does not contain a gene coding for a complete bifunctional acetaldehyde/alcohol dehydrogenase nor a gene that codes for a acetaldehyde dehydrogenase or a pyruvate decarboxylase. Nor could an alternative candidate for acetaldehyde formation be ascertained from the presented transcription data. Furthermore, no acetaldehyde dehydrogenase enzyme activity could be detected in cell extracts (acetaldehyde and $NAD(P)^+$ as electron acceptor or Ac-CoA and NAD(P)H as electron donor, (unpublished data Bielen)), raising the question how acetaldehyde is produced. Possibly, acetaldehyde is derived from pyruvate by a side-activity of pyruvate:ferredoxin oxidoreductase, as was reported for the hyperthermophilic archaeon *Pyrococcus furiosus* [134]. Remarkably, most other *Caldicellulosiruptor* species do not contain orthologs of the ADH gene Csac_0622, suggesting that this potential metabolic pathway for ethanol formation in *C. saccharolyticus* was likely acquired via horizontal gene transfer.

Bioinformatic analysis indicated both Csac_0407 and Csac_0622 have a NADPH binding domain [40] and alcohol dehydrogenase enzyme activity determined in *C. saccharolyticus* cell extracts showed to have a preference for NADPH as electron donor [51]. Though, in *C. saccharolyticus* the source of NADPH is still rather unclear. No NADPH:ferredoxin oxidoreductase homologs could be identified in the *C. saccharolyticus* genome nor any other NADPH generating transhydrogenase [40]. Interestingly, a NADH-dependent reduced ferredoxin:$NADP^+$ oxidoreductase [135] is present in the genome of all sequenced *Caldicellulosiruptor* species securing NADPH generation, except in *C. saccharolyticus* (data not shown).

The physiological function of upregulation of the fatty acid biosynthesis pathway remains elusive, but it could act as both carbon and reductant sink (**Figure 3.3**). The transcription of the fatty acid biosynthesis operon in *C. saccharolyticus* is under the control of the transcription factor FapR (**Table 3.3**). In *Bacillus subtilis* the FapR acts as a repressor of the fatty acid biosynthesis [127]. When malonyl-CoA binds to FapR a conformational change disrupts proper operator binding, allowing transcription [136]. It was proposed that if the rate of fatty acid synthesis falls below the normal levels, a temporary increase in the intracellular concentration of malonyl-CoA would relieve FapR-mediated repression of lipid biosynthetic genes. This mode of transcriptional regulation was assumed not to be auto-regulated since the *B. subtilis* genes responsible for malonyl-CoA synthesis are not under the control of FapR [136]. Interestingly, in the case of *C. saccharolyticus* a FapR binding motif could also be identified in the promoter region of an acetyl-CoA carboxylase (**Table 3.3**);

Table 3.3 Putative binding motifs of the Fur and FapR transcriptional regulator proteins identified in *C. saccharolyticus* and several closely related species. -: orthologous gene is not present. +: orthologous gene present but no binding motif identified.

	Transcription factor Fur (Thermotogales)	
	Gene: FeoA family protein	
Organism	**Locustag**	**Motif**
Clostridium phytofermentans ISDg	+	
Thermoanaerobacter tengcongensis MB4T	TTE0242	ATTGATAATGATTTTCACT
Thermoanaerobacter sp. X514	Teth514_0187	ATTGATAATTATTTTCAAT
Clostridium thermocellum ATCC 27405	+	
Clostridium cellulolyticum H10	Ccel_1998	ATTGAGAATAGTTATCAAA
Caldicellulosiruptor saccharolyticus DSM 8903	Csac_1739	ATTGAGAATAATTCTCAAT
RegPrecise motif*		aTTGAgAaTnAtTcTCAAt

	Transcription factor FapR (Staphylococcus)	
	Gene: Transcription factor FapR	
Organism	**Locustag**	**Motif**
Clostridium phytofermentans ISDg	-	
Thermoanaerobacter tengcongensis MB4T	+	
Thermoanaerobacter sp. X514	+	
Clostridium thermocellum ATCC 27405	Cthe_0938	ATTAGTACCAGGTAATAAT
Clostridium cellulolyticum H10	Ccel_0680	ATTAGTATCTGGTAATAAT
Caldicellulosiruptor saccharolyticus DSM 8903	Csac_1601	TTTAATACCTGATATTAAT
RegPrecise motif*		tTTAatAcctggTatTaaa

	Transcription factor FapR (Staphylococcus)	
	Gene: Acetyl-coenzyme A carboxyl transferase	
Organism	**Locustag**	**Motif**
Clostridium phytofermentans ISDg	-	
Thermoanaerobacter tengcongensis MB4T	+	
Thermoanaerobacter sp. X514	+	
Clostridium thermocellum ATCC 27405	Cthe_0699	ATTATAACCAGGTAATAAT
Clostridium cellulolyticum H10	Ccel_1738	ATTATTACTAGGTACTAAT
Caldicellulosiruptor saccharolyticus DSM 8903	Csac_1881	ATTAATACCAGATATTAAT
RegPrecise motif*		tTTAatAcctggTatTaaa

* RegPrecise motifs are position weight matrixes (regprecise.lbl.gov), only highest bit score nucleotides are given.

this can also be observed in Clostridia species such as *C. thermocellum* and *C. cellolyticum* suggesting a different mode of regulation versus *B. subtilis*.

Finally, several genes/ gene clusters involved in the central carbon metabolism, such as pyruvate: ferredoxin oxidoreductase, oxaloacetate decarboxylase and acetate kinase, were upregulated under high P_{H_2} (**Figure 3.3**). This might be a response to the accumulation of the glycolytic intermediates glyceraldehyde-3-phosphate and/or pyruvate, due to the limited availability of oxidized electron carriers. While the downregulation of non-essential catabolic pathways or transporters might be a way to save metabolic energy under high P_{H_2} conditions.

4.3. Rex mediated transcription regulation

The regulation of the different reductant sinks at the gene level raised the question by what mechanism *C. saccharolyticus* senses the partial hydrogen pressure, and how this controls gene expression. NAD^+ and ferredoxin are the main electron carriers in carbon metabolism of *C. saccharolyticus*. It is conceivable that an elevated P_{H_2} leads to a change in the ratio of reduced versus oxidized cellular electron carriers by inhibiting the hydrogenase activity. Inhibition of hydrogenase by CO was shown to lead to an increased NADH/NAD ratio in *Clostridium cellolyticum* batch cultures [137]. A transcription factor that is known to respond to the intracellular redox potential is the Rex regulator. For various gram-positive bacteria like *Streptomyces coelicolor* [138], *Staphylococcus aureus* [139], *Bacillus subtilis* [140] and *Streptococcus mutans* [141], this Rex regulator has been shown to negatively control expression of genes involved in fermentative metabolism. The DNA-binding activity of Rex is modulated by the intracellular ratio of NADH/NAD^+ [122]. At a low NADH/NAD^+ ratio, the Rex protein binds to the DNA target sites in the promoter thus repressing transcription, while an increase of NADH/NAD^+ ratio results in the dissociation of Rex from the DNA and de-repression of its target genes [123, 138, 139, 142]. The genome of *C. saccharolyticus* contains a Rex homolog (Csac_1220) suggesting that this type of transcriptional control occurs also here. A detailed description of Rex regulons in the genomes of 119 bacteria from 6 different phyla, including *C. saccharolyticus*, has been provided by Ravcheev *et al.* [125] and the upregulated hydrogenase and iron-dependent alcohol dehydrogenase operons, as discussed above, are members of the curated *C. saccharolyticus* Rex regulon presented in the RegPrecise database ([125], regprecise.lbl.gov, [142]). The presence of Rex binding motifs in their operator regions (**Table 3.2**) and the increase in their transcript levels under the high P_{H_2} cultivation condition suggest a Rex mediated regulation of their transcription.

By using a *rex*-deficient strain [139] a direct regulatory effect of Rex on the expression of lactate dehydrogenase formation was verified for *Staphylococcus aureus.*

The shift in metabolism of *C. saccharolyticus* to lactate formation has been shown to be regulated by the level of intracellular electron- and energy carriers, like the NADH/NAD ratio and the level of pyrophosphate and ATP [51]. Our findings demonstrate that the shift to lactate is also regulated at gene level. However, for *C. saccharolyticus* no Rex binding motif could be identified in the LDH promoter region. In the case of *Clostridium acetobutylicum* where the LDH encoding gene exhibited a Rex binding site, no significant lactate formation could be observed in the *rex*-negative mutant [124]. So, these findings suggest that for both *C. saccharolyticus* and *C. acetobutylicum* another mode of regulation of LDH expression is likely to be operative.

5. Conclusion

In this study, *C. saccharolyticus* was grown in glucose-limited chemostats under low and high hydrogen partial pressure conditions. The effects of these conditions on fermentation performance and gene transcription were studied to investigate the organism's strategy for dealing with elevated H_2 levels.

Fermentation profiles showed that increased P_{H_2} affects the fermentation performance. Switching a chemostat from a low P_{H_2} to a high P_{H_2} led to the formation of reduced end-products (lactate, ethanol) and decreased acetate and H_2 yields. Such a switch was always accompanied by a decrease in carbon recovery. When switching the chemostat back to the original low P_{H_2} condition the system recuperated and fermentation profiles restored to their initial state. The determining factors for the switch in fermentation profile were the stirring speed in combination with type of flushing, while flushing with H_2 was almost as efficient as flushing with N_2.

At the transcriptional level *C. saccharolyticus* reacts to an elevated P_{H_2} by increasing transcription levels of genes involved in the carbon metabolism, transport and redox balancing. With respect to the latter, *C. saccharolyticus* facilitates all its possible redox sinks under elevated P_{H_2}, including the lactate dehydrogenase, ethanol dehydrogenases and hydrogenases. With the exception of lactate dehydrogenase, the transcriptional regulation under high P_{H_2} of these operons can be attributed to the involvement of the redox-sensing transcription regulator Rex.

Supplementary materials

Supplementary Table S3.A, Supplementary Table S3.B
Supplementary Table S3.C, Supplementary Table S3.D

Supplementary materials can be found online.
(http://dx.doi.org/10.1016/j.ijhydene.2012.11.082)

Chapter 4

Biohydrogen production by the thermophilic bacterium *Caldicellulosiruptor saccharolyticus*: Current status and perspectives

This chapter has been published as:

A.A.M. Bielen, M.R.A. Verhaart, J. van der Oost and S.W.M. Kengen. *Biohydrogen Production by the Thermophilic Bacterium Caldicellulosiruptor saccharolyticus: Current Status and Perspectives.* Life, 2013 3(1): 52-85

Abstract

Caldicellulosiruptor saccharolyticus is one of the most thermophilic cellulolytic organisms known to date. This Gram-positive anaerobic bacterium ferments a broad spectrum of mono-, di- and polysaccharides to mainly acetate, CO_2 and hydrogen. With hydrogen yields approaching the theoretical limit for dark fermentation of 4 mol hydrogen per mol hexose, this organism has proven itself to be an excellent candidate for biological hydrogen production. This review provides an overview of the research on *C. saccharolyticus* with respect to the hydrolytic capability, sugar metabolism, hydrogen formation, mechanisms involved in hydrogen inhibition and the regulation of the redox and carbon metabolism. Analysis of currently available fermentation data reveal decreased hydrogen yields under non-ideal cultivation conditions, which are mainly associated with the accumulation of hydrogen in the liquid phase. Thermodynamic considerations, concerning the reactions involved in hydrogen formation, are discussed with respect to the dissolved hydrogen concentration. Novel cultivation data demonstrate the sensitivity of *C. saccharolyticus* to increased hydrogen levels regarding substrate load and nitrogen limitation. In addition, special attention is given to the rhamnose metabolism, which represents an unusual type of redox balancing. Finally, several approaches are suggested to improve biohydrogen production by *C. saccharolyticus*.

1. Introduction

The use of renewable plant biomass for the production of biofuels, chemicals or other biocommodities can provide a realistic alternative for fossil fuel based processes [143, 144]. The implementation of lignocellulosic biomass for biofuel production requires the degradation of recalcitrant substrates like cellulose, hemicellulose or lignin. Lignin is either removed or modified [145], while cellulose and hemicellulose are converted into more readily fermentable mono-, di- and oligo-saccharides. Although this can be achieved by different (thermo)chemical or enzymatic pre-treatments, a more desirable process combines both substrate hydrolysis and fermentation of complex plant biomass. Such a 'consolidated bioprocess' (CBP) circumvents the negative environmental impact inherent to (thermo)chemical pre-treatment and might limit overall process costs [143, 146].

Hydrogen gas (H_2) is considered an alternative for the non-renewable fossil fuels and can be produced in a carbon neutral process. The controlled biological production of H_2 would allow for capturing the CO_2 released during the process, preventing it to dissipate into the environment. In addition, compared to carbon based (bio)fuel types, H_2 has the advantage that i) during its oxidation only H_2O is released and ii) that H_2 fuel cells can be used, which are more energy efficient than the presently used combustion engines [109]. Biohydrogen ($BioH_2$) can be produced from renewable feedstocks in an anaerobic fermentation process, which is often referred to as dark fermentation, to distinguish it from photo-fermentative hydrogen production.

Both plant biomass degradation and biological H_2 formation appear advantageous under thermophilic conditions. Moreover, thermophiles display an extensive glycoside hydrolase inventory aiding in the lignocellulosic biomass breakdown [33-35]. Based on thermodynamic considerations H_2 formation is more feasible at elevated temperatures [27, 147]. Correspondingly, H_2 yields are in general higher for (hyper)thermophiles, reaching the theoretical limit of 4 mol H_2 per mol of hexose, compared to the mesophilic hydrogen producers [27, 86, 148].

Since its isolation in the mid-eighties it has become clear that the thermophilic anaerobic bacterium *Caldicellulosiruptor saccharolyticus* [37] displays both the desirable polysaccharide degrading capabilities (including cellulose) and H_2 producing characteristics, making it an outstanding source for thermostable glycoside hydrolases and an excellent candidate for $bioH_2$ production from renewable biomass.

This review will discuss the available scientific data on *C. saccharolyticus* regarding its lignocellulolytic capability, substrate specificity, catabolism and H_2 producing capacity, which has made *C. saccharolyticus* to become a model organism for the study of fermentative hydrogen formation at elevated temperatures.

2. Isolation and initial characterization

The foreseen commercial value of thermostable cellulolytic enzymes in biotechnological applications triggered the investigation of new sources of these types of enzymes. In search for novel thermophilic cellulolytic micro-organisms, several anaerobic bacteria have been isolated from natural enrichment sites from the Rotorua-Taupo thermal area in New Zealand [36]. One of the isolated strains, TP8.T 6331 [36] also referred to as TP8 [39] or '*Caldocellum saccharolyticum*' [38], revealed thermostable cellulase activity up to 85°C [36, 39] but also lignocellulolytic biomass decomposition capabilities [38]. Strain TP8.T 6331 was assigned to a new genus *Caldicellulosiruptor* as *Caldicellulosiruptor saccharolyticus* and was characterized as a gram-positive, asporogenous, extremely thermophilic and strictly anaerobic bacterium capable of sustaining growth at a temperature range of 45-80°C (T_{opt} =70 °C) and pH range of 5.5-8.0 (pH_{opt} = 7) [37]. Acid production could be detected for a broad substrate range including different pentoses and hexoses, di-saccharides and polysaccharides like cellulose and xylan [36-39]. Especially the capacity to use cellulose at high temperatures was exceptional.

Ever since, several cellulolytic and weakly cellulolytic *Caldicellulosiruptor* species have been identified, all of which are isolated from terrestrial geothermal regions [36, 37, 56, 58-62, 149-154]. The availability of the fully sequenced genomes of 8 of these *Caldicellulosiruptor* species allows the investigation of the possible differences in their cellulolytic traits and the analysis of other remarkable features of this genus [40, 42, 57, 115, 116].

3. Hydrolytic capacity and complex biomass decomposition

For the decomposition of recalcitrant plant polysaccharides *C. saccharolyticus* does not employ cellulosome-like structures, as described for some *Clostridium* species [155], but wields a variety of free-acting endo- and exo- glycoside hydrolases (GH) capable of hydrolyzing the glycosidic bonds of α- and β- glucans like starch, pullulan and cellulose, but also xylan and hetero-polysaccharides like hemicelluloses and pectin [35, 37, 40, 43, 156]. Actually, *Caldicellulosiruptor* species, together with *Thermoanaerobacter* species, are one of the most thermophilic crystalline cellulose-degrading organisms known to date that use free-acting primary cellulases [34]. *C. saccharolyticus* contains 59 open reading frames (ORFs) that include GH catalytic domains [42]. Some of these ORFs code for multifunctional, multi-domain proteins that contain glycoside hydrolase domains, belonging to different GH families, and multiple carbon binding modules [33, 42, 43]. The catalytic properties and structural organization of some of the glycoside hydrolases from *C. saccharolyticus*, have been extensively investigated. A β-glucosidase (BglA) [157, 158], β-xylosidase [159], β-1,4-

xylanase [160] and a type I pullulanase [161] from *C. saccharolyticus* have been cloned into *E. coli* and characterized.

The majority of the genes encoding xylan degradation associated enzymes appear to be clustered on the genome (*xynB-xynF*, Csac_2404-2411) [33, 43]. Both XynA and XynE exhibit endoxylanase and xylosidase activity [43, 162], while XynB only acts as a β-D-xylosidase [163]. XynC does not contain a GH domain and was shown to be an acetyl esterase [164], XynD showed to be active on xylan [43] and although the cloning and expression of intact multi-domain XynF could not be achieved, its N- and C- terminal parts revealed catalytic activity on arabinoxylan. Hence XynF was proposed to be involved in the degradation of the arabinoxylan component of hemicellulose [43]. A second locus covers several genes coding for multidomain proteins involved in glucan and mannan hydrolysis (*celA-manB*, Csac_1076-1080) [33, 43]. *CelA* is coding for a multidomain cellulase [165], the bifunctional cellulase CelB exhibited both endo-β-1,4-glucanase and exo-β-1,4-glucanase activity [43, 166, 167] and CelC was characterized as an endo-1,4-β-D glucanase [168]. ManA was characterized as a β-mannanase [169], but *ManB* codes for an inactive mannase, which after correcting for a frame shift in the nucleotide sequence, exhibited β-mannanase activity [168]. Several ORFs of the described *celA-manB* and *xynB-xynF* loci were differentially transcribed on pretreated poplar and switch grass compared to the monosaccharides glucose and xylose, showing their involvement in the decomposition of complex carbohydrates [43].

While some of the GH proteins act intracellularly, others are excreted, allowing the decomposition of non-soluble substrates to smaller oligo- or mono-saccharides. Early findings by Reynolds *et al.* already indicated that a significant percentage of the cellulolytic activity was found to be associated with insoluble substrate [39]. These interactions, between GH and substrate, are facilitated by carbon binding modules (CBM). CBMs allow the positioning of the GH catalytic domains in the vicinity of the substrate, thus increasing the rate of catalysis. Interestingly, most multi-domain GHs identified in *C. saccharolyticus*, containing one or more CBMs, possess a signal peptide, which mark them for excretion [43]. A relative higher amount of GH related proteins could be observed in the substrate bound protein fraction, from *Caldicellulosiruptor* species grown on Avicel, with respect to the whole cell proteome [42]. Additionally, these proteome studies allowed the identification of specific GHs which interact with crystalline cellulose. CelA, a multi-domain GH consisting of two GH domains (GH9 and GH48) and three CBM3 modules, was found to be the most abundant substrate bound protein for strong cellulolytic *Caldicellulosiruptor* species [42]. The sequenced genomes of *Caldicellulosiruptor* species reveal differences in glycoside hydrolytic capacity, which reflects their difference in biomass degrading capabilities [42, 156]. The secretome of *C. saccharolyticus* grown on glucose contains several carbohydrate-degrading enzymes including CelA, ManA, CelB and CelC (protein sequence ID

A4XIF5/6/7/8 respectively), which indicates that these GH are constitutively expressed even under non-cellulose degrading conditions [170].

In addition to the interactions between substrate and glycoside hydrolases also interactions between whole cells and a substrate have been observed for *C. saccharolyticus.* These interactions appeared to be substrate specific. For instance, a higher degree of cell-to-substrate attachment was observed for cells grown on switch grass, compared to poplar [43]. Several S-layer homology (SLH) domain containing proteins, which have been identified in *Caldicellulosiruptor* species, are proposed to have a role in such cell substrate interactions. These SLH domain proteins contain both glycoside hydrolases domains and non-catalytic carbohydrate binding domains, which are utilized in lignocellulose degradation by both recruiting and degrading complex biomass via cell substrate interactions [171]. In addition, cell immobilization on a support matrix, like pine wood shavings, supports cell survival and improves the H_2 evolving capacity of *C. saccharolyticus* [172].

4. Sugar catabolism and pathway regulation

4.1. Sugar uptake and fermentation

Soluble sugar substrates can enter *C. saccharolyticus* cells either as mono-, di- or oligo-saccharides via several ABC-transporters, which facilitate substrate transport across the membrane at the expense of ATP. In addition *C. saccharolyticus* contains one fructose specific phosphotransferase system (PTS). During PTS-mediated transport the substrate is both transported and phosphorylated at the expense of phosphoenolpyruvate (PEP). The substrate specificity of the 24 sugar ABC transport systems, identified in *C. saccharolyticus*, has been assigned based on bioinformatic analysis and functional genomics [41]. Most of the identified ABC transporters have a broad predicted substrate specificity and for some transporters the annotated substrate specificity was confirmed by transcriptional data obtained from cells grown on different mono-saccharides [41]. Some substrates can be transported by multiple transporter systems (**Figure 4.1**).

The growth of *C. saccharolyticus* on sugar mixtures revealed the co-utilization of hexoses and pentoses, without any signs of carbon catabolite repression (CCR) [40, 41]. The absence of CCR is in principle a very advantageous characteristic of *C. saccharolyticus*, as it enables the simultaneous fermentation of hexoses and pentoses [40]. Substrate co-utilization has also been confirmed for biomass derived hydrolysates [44, 45]. Despite the absence of CCR a somewhat higher preference for the pentose sugars (xylose and arabinose) with respect to the hexose sugars (glucose, mannose and galactose) was demonstrated, but the highest preference was observed for the PTS-transported hexose fructose [41].

Once inside the cell, the sugar substrates are converted into a glycolytic intermediate. NMR analysis of the fermentation end-products of *C. saccharolyticus* grown on ^{13}C-labeled glucose showed that the EMP is the main route for glycolysis [48]. All genes encoding components of the EMP and non-oxidative pentose phosphate pathway have been identified in the genome. There is no evidence for the presence of the oxidative branch of the pentose phosphate pathway or the ED pathway [40] (**Figure 4.1**).

Each sugar substrate, with the exception of rhamnose is completely catabolized to glyceraldehyde 3-phosphate (GAP) (rhamnose catabolism will be discussed separately in more detail below). The subsequent conversion of GAP to pyruvate, via the C-3 part of glycolysis, results in the formation of the reduced electron carrier NADH. Pyruvate can be further oxidized to acetyl-CoA by pyruvate ferredoxin oxidoreductase (POR), which is coupled to the generation of reduced ferredoxin (Fd_{red}). Finally, acetyl-CoA can be converted to the fermentation end-product acetate (**Figure 4.1**).

Both types of reduced electron carriers (NADH and Fd_{red}) can be used by hydrogenases for proton reduction, thus forming H_2. The genome of *C. saccharolyticus* contains a gene cluster coding for an NADH-dependent cytosolic hetero-tetrameric Fe-only hydrogenase (*hyd*) and a cluster encoding a membrane bound multimeric [NiFe] hydrogenase (*ech*), which presumably couples the oxidation of Fd_{red} to H_2 production [40]. Under optimal cultivation conditions, when all reductants are used for H_2 formation, the complete oxidation of glucose yields 4 mol of H_2 per mol of glucose consumed (**Eq. 4.1**).

However, suboptimal growth conditions lead to a mixed acid fermentation with ethanol (**Eq. 4.2**) and lactate (**Eq. 4.3**) as end-products in addition to acetate and H_2. Lactate is produced from pyruvate, using NADH as electron donor and catalyzed by lactate dehydrogenase (LDH). The corresponding *ldh* gene could be easily identified in the genome [40, 51]. However, the identity of the enzymes and genes involved in ethanol formation, are less clear. Two alcohol dehydrogenase (ADH) genes have been identified in the genome which, based on transcriptional data, can be both be involved in ethanol formation from acetaldehyde. The way acetaldehyde is produced, is however, not known. Acetaldehyde can be produced from pyruvate by a pyruvate decarboxylase, as described for yeast, or from acetyl-CoA by an acetaldehyde dehydrogenase, as is commonly seen in fermentative bacteria. In several thermophilic ethanol-producing bacteria, acetaldehyde is produced by a bifunctional acetaldehyde/ethanol dehydrogenase [173, 174]. However, no candidate gene could be identified for any of these alternatives [40, 175]. A third option might be that acetaldehyde is formed from pyruvate in a CoA-dependent side reaction of the pyruvate:ferrodoxin oxidoreductase as described for *Pyrococcus furiosus* [134].

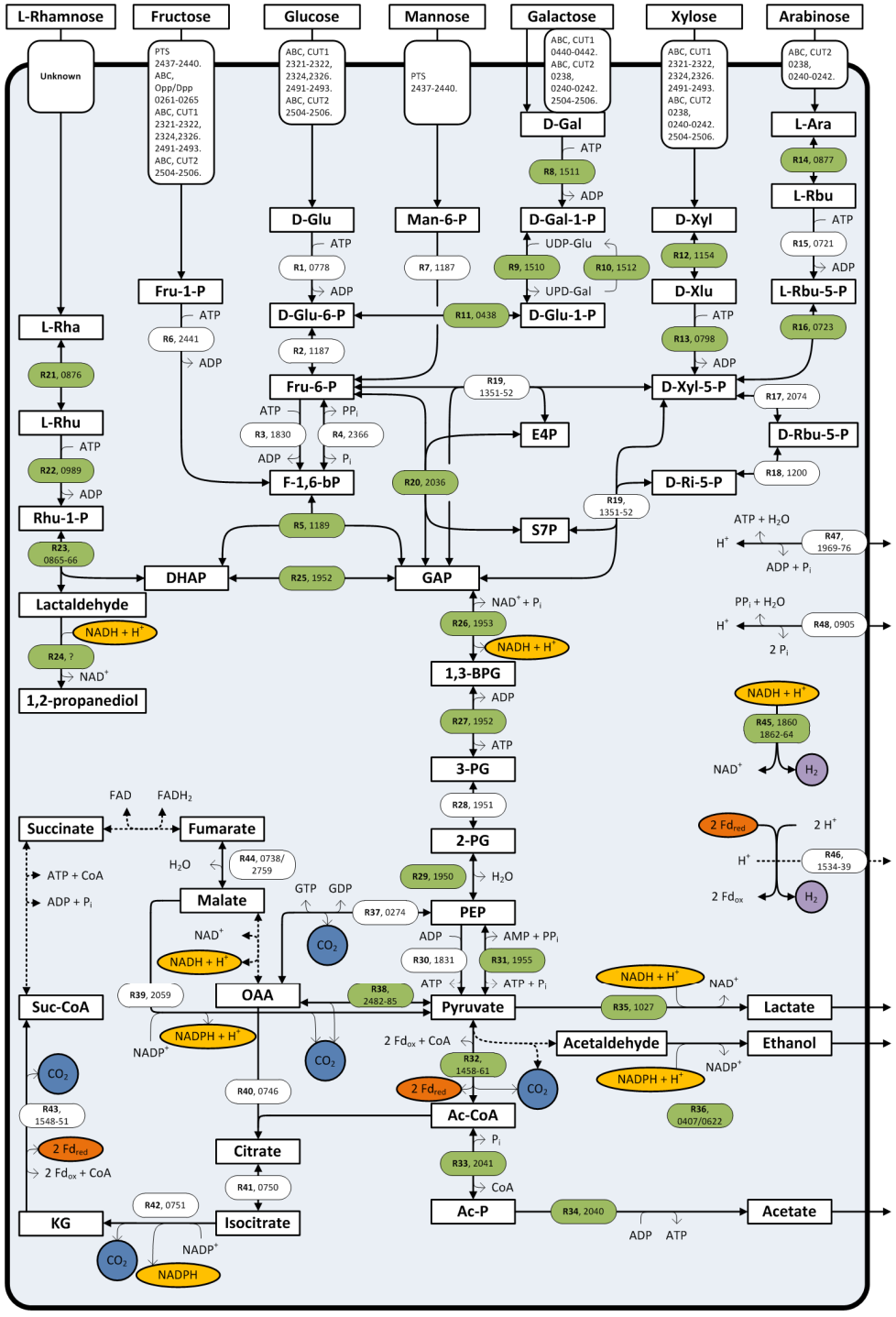

L-Rhamnose
Fructose
Glucose
Mannose
Galactose
Xylose
Arabinose
Unknown
PTS 2437-2440. ABC, Opp/Dpp 0261-0265 ABC, CUT1 2321-2322, 2324,2326. 2491-2493. ABC, CUT2 2504-2506.
ABC, CUT1 2321-2322, 2324,2326. 2491-2493. ABC, CUT2 2504-2506.
PTS 2437-2440.
ABC, CUT1 0440-0442. ABC, CUT2 0238, 0240-0242. 2504-2506.
ABC, CUT1 2321-2322, 2324,2326. 2491-2493. ABC, CUT2 0238, 0240-0242. 2504-2506.
ABC, CUT2 0238, 0240-0242.
D-Gal
L-Ara
R14, 0877
L-Rbu
R15, 0721
L-Rbu-5-P
R16, 0723
D-Xyl
R12, 1154
D-Xlu
R13, 0798
D-Xyl-5-P
R17, 2074
D-Rbu-5-P
R18, 1200
D-Ri-5-P
R8, 1511
D-Glu
D-Gal-1-P
R9, 1510
R10, 1512
UDP-Glu
UPD-Gal
R1, 0778
Man-6-P
R7, 1187
Fru-1-P
R6, 2441
D-Glu-6-P
R11, 0438
D-Glu-1-P
R2, 1187
L-Rha
R21, 0876
L-Rhu
R22, 0989
Rhu-1-P
R23, 0865-66
1,2-propanediol
Fru-6-P
R19, 1351-52
E4P
R3, 1830
R4, 2366
F-1,6-bP
R20, 2036
S7P
R5, 1189
DHAP
R25, 1952
GAP
Lactaldehyde
R24, ?
NADH + H+
NAD+
NAD+ + Pi
R26, 1953
1,3-BPG
R27, 1952
3-PG
R28, 1951
2-PG
R29, 1950
PEP
R37, 0274
R30, 1831
R31, 1955
ATP + H2O
H+
R47, 1969-76
ADP + Pi
PPi + H2O
R48, 0905
2 Pi
R45, 1860 1862-64
H2
2 Fdred
2 H+
R46, 1534-39
2 Fdox
FAD
FADH2
Succinate
Fumarate
R44, 0738/ 2759
H2O
ATP + CoA
ADP + Pi
Malate
GTP
GDP
CO2
OAA
R38, 2482-85
R39, 2059
Suc-CoA
Pyruvate
R35, 1027
Lactate
Acetaldehyde
Ethanol
2 Fdox + CoA
R32, 1458-61
R40, 0746
Ac-CoA
R43, 1548-51
Citrate
R41, 0750
R42, 0751
KG
Isocitrate
R33, 2041
CoA
Ac-P
R34, 2040
Acetate
R36, 0407/0622
NADPH + H+
NADP+
ATP
ADP
AMP + PPi
Pi
PPi

Figure 4.1 Overview of the central carbon metabolism of *Caldicellulosiruptor saccharolyticus*. During sugar catabolism electron carriers (NAD^+ and Fd_{ox}) are reduced. These reduced electron carriers can be re-oxidized by hydrogenases to form H_2 (reaction 45 and 46). For each reaction (numbered) the locus tags of the associated genes are given. When given in green they have been shown to be regulated by a specific mono-saccharide [40, 41]. Dotted lines: no candidate genes can be identified. Reactions: **R1**, glucokinase, EC 2.7.1.2; **R2**, phosphoglucose isomerase, EC 5.3.1.9; **R3**, ATP-dependent 6-phosphofructokinase, EC 2.7.1.11; **R4**, PPi-dependent 6-phosphofructokinase, EC 2.7.1.90; **R5**, fructose-bisphosphate aldolase, EC 4.1.2.13; **R6**, 1-phosphofructokinase, EC 2.7.1.56; **R7**, phosphomannose isomerase, EC 5.3.1.8; **R8**, galactokinase, EC 2.7.1.6; **R9**, galactose 1-phosphate uridyl transferase, EC 2.7.7.12; **R10**, UDP-galactose 4-epimerase, EC 5.1.3.2; **R11**, phosphoglucomutase, EC 5.4.2.6; **R12**, D-xylose isomerase, EC 5.3.1.5; **R13**, xylulokinase, EC 2.7.1.17; **R14**, L-arabinose isomerase, EC 5.3.1.4; **R15**, L-ribulokinase, EC 2.7.1.16; **R16**, L-xylulose 5-phosphate 3-epimerase, EC 5.1.3.22; **R17**, ribulose-phosphate 3-epimerase, EC 5.1.3.1; **R18**, ribose-5-phosphate isomerase, EC 5.3.1.6; **R19**, transketolase, EC 2.2.1.1; **R20**, transaldolase, EC 2.2.1.2; **R21**, L-rhamnose isomerase, EC 5.3.1.14; **R22**, L-rhamnulose kinase, EC 2.7.1.5; **R23**, L-rhamnulose-1-phosphate aldolase, EC 4.1.2.19; **R24**, L-1,2-propanediol oxidoreductase, EC 1.1.1.77; **R25**, triose phosphate isomerase, EC 5.3.1.1; **R26**, glyceraldehyde-3-phosphate dehydrogenase, EC 1.2.1.12; **R27**, phosphoglycerate kinase, EC 2.7.2.3; **R28**, phosphoglycerate mutase, EC 5.4.2.1; **R29**, phosphopyruvate hydratase (enolase), EC 4.2.1.11; **R30**, pyruvate kinase, EC 2.7.1.40; **R31**, pyruvate phosphate dikinase, EC 2.7.9.1; **R32**, pyruvate:ferredoxin oxidoreductase, EC 1.2.7.1; **R33**, phosphotransacetylase, EC 2.3.1.8; **R34**, acetate kinase, EC 2.7.2.1; **R35**, L-lactate dehydrogenase, EC 1.1.1.27; **R36**, alcohol dehydrogenase, EC 1.1.1.1; **R37**, phosphoenolpyruvate carboxykinase, 4.1.1.32; **R38**, oxaloacetate decarboxylase (Na^+ Pump), EC 4.1.1.3; **R39**, malic enzyme 1.1.1.40; **R40**, citrate synthase, EC 2.3.3.1; **R41**, aconitase, EC 4.2.1.3; **R42**, isocitrate dehydrogenase, EC 1.1.1.42; **R43**, 2-oxoglutarate:ferredoxin oxidoreductase, EC 1.2.7.3; **R44**, fumarase, EC 4.2.1.2; **R45**, NADH-dependent Fe-only hydrogenase (*hyd*); **R46**, ferredoxin-dependent [NiFe] hydrogenase (*ech*); **R47**, H^+-ATPase, EC 3.6.3.14; **R48**, V-type H^+-translocating pyrophosphatase. Abbreviations: D-Glu, D-Glucose; D-Glu-6-P, D-Glucose-6-phosphate; Fru-6-P, Fructose-6-phosphate; F-1,6-bP, Fructose-1,6-bisphosphate; Fru-1-P, Fructose-1-phosphate; Man-6-P, Mannose-6-phosphate; D-Gal, D-Galactose; D-Gal-1-P, D-Galactose-1-phosphate; D-Glu-1-P, D-Glucose-1-phosphate; D-Xyl, D-Xylose; D-Xlu, D-Xylulose; D-Xyl-5-P, Xylulose 5-phosphate; L-Ara, L-Arabinose; L-Rbu, L-Ribulose; Rbu-5-P, L-Ribulose-5-P; D-Rbu-5-P, D-Ribulose-5-phosphate; D-Ri-5-P, D-Ribose-5-P; E4P, D-Erythrose 4-phosphate; S7P, D-Sedoheptulose-7-phosphate; DHAP, Dihydroxyacetone phosphate; GAP, Glyceraldehyde 3-phosphate; 1,3-BPG, 1,3-Bisphosphoglycerate; 3-PG, 3-Phosphoglycerate; 2-PG, 2-Phosphoglycerate; PEP, Phosphoenolpyruvate; Ac-CoA, Acetyl-CoA; Ac-P, Acetyl phosphate; OAA, Oxaloacetate; KG, 2-Oxoglutarate; Suc-CoA, Succinyl-CoA; H_2O, Water; H^+, Proton; H_2, Hydrogen; CO_2, Carbon dioxide; Fd_{ox} oxidized Ferredoxin; Fd_{red} reduced Ferredoxin; $NADP^+$, Nicotinamide adenine dinucleotide phosphate; NADPH, reduced Nicotinamide adenine dinucleotide phosphate; NAD^+, Nicotinamide adenine dinucleotide; NADH, reduced Nicotinamide adenine dinucleotide; FAD, Flavin adenine dinucleotide; $FADH_2$, reduced Flavin adenine dinucleotide; ATP, Adenosine triphosphate; ADP, Adenosine diphosphate; AMP, Adenosine monophosphate; GTP, Guanosine triphosphate; GDP, Guanosine diphosphate; PP_i, pyrophosphate; P_i, inorganic phosphate; CoA, Coenzyme A; UPD-Glu, UDP-Glucose; UDP-Gal, UPD-Galactose.

There is no experimental evidence for such a side reaction in *C. saccharolyticus*, but the absence of a dedicated enzymatic acetaldehyde-forming step might explain the low ethanologenic capacity of *C. saccharolyticus*. The difference in the observed lower ethanol to acetate ratio for *C. saccharolyticus* with respect to *Clostridium thermocellum* [38] can be explained by the fact that the *C. saccharolyticus* genome does not have a similar pathway for ethanol formation as identified in *C. thermocellum* [173]. Similar to other high yield ethanol-producing thermophiles, like *Thermoanaerobacter*

ethanolicus [176] or *Thermoanaerobacterium saccharolyticum* [177], *Clostridium thermocellum* has a bifunctional acetaldehyde-CoA/alcohol dehydrogenase, which catalyzes ethanol formation from acetyl-coA [173, 174]. Such a pathway is absent in low level ethanol-producing thermophiles like *C. saccharolyticus* [40], *Thermoanaerobacter tengcongensis* [22] and *Pyrococcus furiosus* [22, 178]. Although *C. saccharolyticus* is able to produce some ethanol, the flux through the ethanol-forming pathway is apparently limited resulting in the lower ethanol to acetate ratio.

NADPH is assumed to be the preferred substrate for the ethanol-forming ADH reaction [40, 51, 178]. A potential source of NAPDH could be the isocitrate dehydrogenase in the oxidative branch of the incomplete TCA cycle (**Figure 4.1**). Alternatively, NADPH can be produced from NADH, but so far no candidate genes coding such transhydrogenase have been be identified in *C. saccharolyticus* [40, 175, 178]. With respect to H_2 production, it is important to realize that only the production of acetate is coupled to H_2 formation since no net reducing power remains for H_2 formation when ethanol or lactate is produced.

(**Eq. 4.1**) $\text{Glucose} + 4\ H_2O \rightarrow 2\ \text{acetate}^- + 2\ HCO_3^- + 4\ H^+ + 4\ H_2$

(**Eq. 4.2**) $\text{Glucose} + 2\ H_2O \rightarrow 2\ \text{ethanol} + 2\ HCO_3^- + 2\ H^+$

(**Eq. 4.3**) $\text{Glucose} \rightarrow 2\ \text{lactate}^- + 2\ H^+$

4.2. Rhamnose fermentation

Compared to glucose, fermentative growth on the deoxy sugar rhamnose is associated with a different carbon and electron metabolism. The proposed pathway for rhamnose degradation (**Figure 4.1**) implies that during rhamnose catabolism half of the generated reduced electron-carriers is used for the reduction of lactaldehyde to 1,2-propanediol, while the other half can be recycled through H_2 formation [40, 178]. Indeed, fermentation of rhamnose by *C. saccharolyticus* results in the production of 1,2 -propanediol, acetate, H_2 and CO_2 in a 1:1:1:1 ratio (**Figure 4.2A**). This ratio suggests that all NADH is used for 1,2 propanediol formation and that all Fd_{red} is used for H_2 formation. However, when *C. saccharolyticus* is grown on rhamnose under a headspace of carbon monoxide (CO), which is an established competitive inhibitor of both NiFe- and Fe-only [179, 180] hydrogenases, H_2 evolution is significantly inhibited (**Figure 4.2B**). The CO cultivation condition does not affect 1,2-propanediol formation, but less H_2 is produced and the remaining reduced electron-carriers are now recycled by the formation of lactate and ethanol, consequently leading to a decrease in the acetate level.

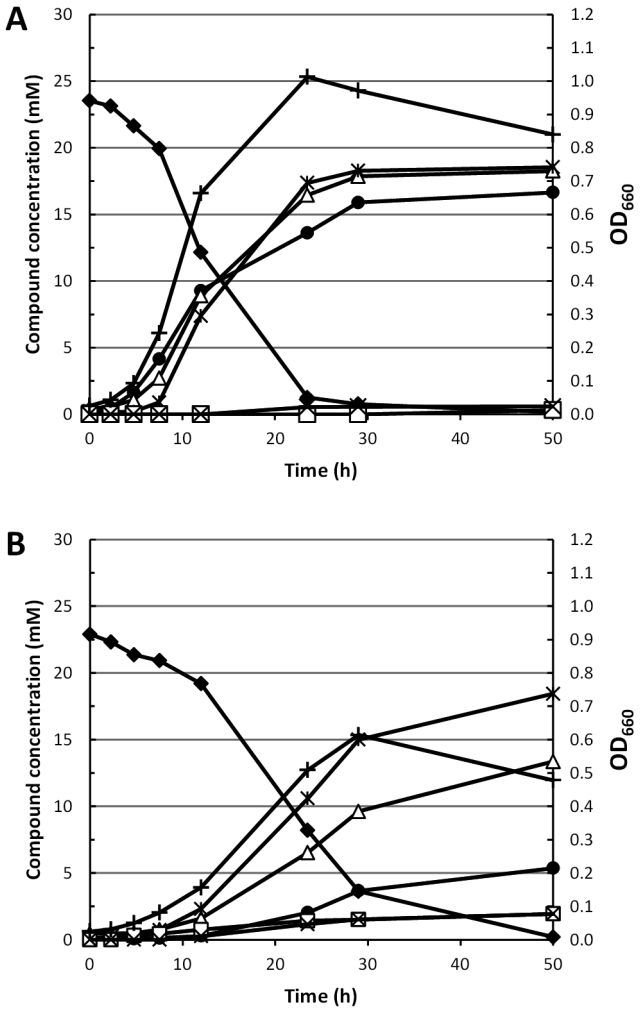


Figure 4.2 Fermentation profile of *C. saccharolyticus* grown on rhamnose batch cultivation: **A)** without CO in the headspace **B)** with a 100% CO headspace. Rhamnose (diamonds), acetate (open triangles), 1,2 propanediol (asterisks), lactate (open squares), ethanol (crosses), H_2 (circles) and OD_{660} (plus sign).

These findings suggest that, based on the substrate specificity of the lactate dehydrogenase (NADH, [51]) and the ethanol dehydrogenase(s) (NADPH, [51]),electron exchange between the Fd_{red} and $NAD(P)^+$ is required. However, no genes have been identified in *C. saccharolyticus* coding for an enzyme capable of catalyzing such a reaction [40, 175]. Transcript levels of the genes from the gene cluster containing both the L-rhamnose isomerase and L-rhamnulose-1-phosphate aldolase are highly upregulated during growth on rhamnose (**Figure 4.3**) [40].

Although the function of the other genes in the cluster with respect to rhamnose catabolism remains unclear, the absence of the gene cluster and the rhamnose kinase gene from the genome strongly correlates with the inability of rhamnose degradation by other *Caldicellulosiruptor* species (**Table 4.1**). This difference in rhamnose degrading capability between *Caldicellulosiruptor* species reflects the open nature of the *Caldicellulosiruptor* pan genome [42].

4.3. Involvement of pyrophosphate in the energy metabolism

From a bioenergetics point of view, glucose fermentation is optimal when acetate is the only end-product, because an additional ATP is generated during the final acetate-forming step (**Figure 4.1**). When it is assumed that the ABC-transporter mediated substrate uptake requires 1 ATP, overall ATP yields become 1.5 ATP per acetate versus only 0.5 ATP per lactate or ethanol. However, ATP yields might even be higher when the involvement of pyrophosphate (PP_i) as an energy carrier is considered.

PP_i is a by-product of biosynthesis reactions like DNA and RNA synthesis or is generated when the amino acids are coupled to their tRNAs during protein synthesis [68]. Since these reactions are close to equilibrium accumulation of PP_i is believed to have an inhibitory effect on growth, and only the effective removal of PP_i drives these biosynthetic reactions forward [69]. When PP_i is hydrolysed to P_i, by a cytosolic inorganic pyrophosphatase (PPase), the free energy just dissipates as heat. *C. saccharolyticus* does not contain a cytosolic PPase but possesses a membrane bound H^+-translocating PPase [107], which allows the free energy released upon PP_i hydrolysis to be preserved as a proton motive force. The high-energy phosphate bond of PP_i can also be used for the phosphorylation of fructose 6-phosphate, catalysed by a PP_i -dependent phosphofructokinase (**Figure 4.1**).

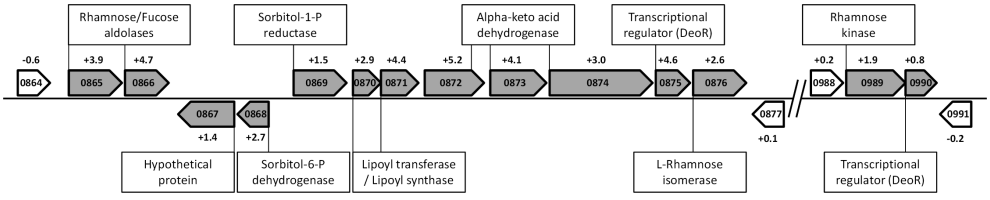


Figure 4.3 Schematic representation of two rhamnose associated gene clusters from *C. saccharolyticus* (Csac_0865-Csac_0876 and Csac_0989-Csac_0990). For each member of the cluster (grey arrows) the proposed function (text box) and locus tag number (4 digit number) is given. Presented log2 values represent the ratio between transcription levels of the specific gene during growth on rhamnose with respect to growth on glucose; (+) upregulated, (-) downregulated on rhamnose versus glucose [40].

Table 4.1 Overview of Caldicellulosiruptor species which are able to grow on rhamnose. There is a correlation between the ability to grow on rhamnose and the presence of gene clusters orthologous to the rhamnose associated gene clusters identified in C. saccharolyticus (**Figure 4.3**). nt, not tested for growth on rhamnose. (+), able to grow of rhamnose / gene cluster is present in genome. (-), no growth observed on rhamnose / gene cluster is not present in genome. ns, not sequenced.

Strain	Growth on Rhamnose	Gene cluster Csac_0865-76	Gene cluster Csac_0989-90	Reference
C. saccharolyticus	+	+	+	[37]
C. bescii	+	+	+	[154]
C. owensensis	+	+	+	[58]
C. obsidiansis	nt	+	+	[151]
C. kronotskyensis	nt	+	+	[59]
C. hydrothermalis	nt	+	+	[59]
C. kristjanssonii	-	-	-	[56]
C. lactoaceticus	-	-	-	[60]
C. acetigenus	nt	ns	ns	[61]

Furthermore, PP_i is consumed during the catabolic conversion of phosphoenolpyruvate to pyruvate, catalysed by pyruvate phosphate dikinase (PPDK). Such a catabolic role for PPDK was proposed based on the increase in transcript level of the *ppdk* gene under increased glycolytic fluxes [40, 107]. Altogether, the use of PP_i as an energy donor could be a way for the organism to deal with the relative low ATP yields which are usually associated with fermentation [181].

4.4. Mechanism involved in mixed acid fermentation

H_2 has been reported as a growth inhibitor for *C. saccharolyticus* [38, 50] and a critical dissolved H_2 concentration, which leads to the complete inhibition of growth, of 2.2 mmol/L has been determined for controlled batch cultivations [112]. In addition, an elevated P_{H2} has been shown to cause a switch in the fermentation profile, leading to increased formation of ethanol and lactate, in both controlled batch and chemostat cultivations [112, 113, 148, 175]. Willquist *et al.* reported that in controlled batch fermentations the initiation of lactate formation coincided with an increment in both the internal $NADH/NAD^+$ ratio and the P_{H2} of the system [51, 113].

An increase of the overall carbon flux through glycolysis results in a higher NADH production rate. To maintain a constant $NADH/NAD^+$ ratio a subsequently higher H_2 formation rate is required. However, when the H_2 formation rate (volumetric H_2 production rate) exceeds the H_2 liquid to gas mass transfer rate, H_2 will accumulate in the liquid phase. In some cases this can even lead to super saturation of the liquid, meaning that the dissolved H_2 concentration exceeds the maximal theoretical H_2 solubility [112, 182]. Such high dissolved H_2 levels inhibit H_2 formation and presumably cause an increase in the $NADH/NAD^+$ ratio. An increased $NADH/NAD^+$

ratio has been shown to have an inhibitory effect on GAPDH activity thus limiting the glycolytic flux [113] and consequently leads to a decrease in substrate consumption and growth rate. A switch to lactate formation alleviates the inhibitory effect of an increased NADH/NAD^+ ratio by causing a decrease of the NADH/NAD^+ ratio through the reduction of pyruvate.

A clear switch to lactate formation caused by an increase in glycolytic flux can be observed for *C. saccharolyticus* grown in a chemostat under high P_{H2} (**Figure 4.4A**). When the glucose load is increased from 20mM to 40mM, end-product formation is switched form mainly acetate to mainly lactate, respectively. After the switch to 40mM glucose some adaptation time is required before a new steady state is achieved. During this adaptation period substantially higher amounts of glucose were detected in the culture effluent, probably indicating an inhibition of the glycolytic flux at the level of GAPDH. When a new steady state was achieved, residual glucose was only slightly higher with respect to the 20mM condition. As a consequence of the increase in glycolytic flux, the lactate concentration dramatically increased while the acetate concentration hardly changed, indicating that during both substrate loads a similar volumetric H_2 productivity was maintained. These data indicate that under these conditions the organism is not capable of dealing with the increased glycolytic flux by increasing its H_2 productivity, but requires a switch to lactate formation. Although oxaloacetate formation was discernible under a 20mM glucose load, significantly higher and therefore quantifiable levels of oxaloacetate were detected under the 40mM glucose load condition. This observation together with the increased flux towards lactate suggests that in the newly achieved steady state, a bottleneck exists at the level of pyruvate.

Ethanol formation can also serve as reductant sink in *C. saccharolyticus*. For some chemostat cultivation conditions a decrease in H_2 yield is only associated with ethanol formation and not with lactate formation [48, 113, 175]. These conditions concern low substrate loads, and the observed increase in ethanol formation was triggered by an increase of the dilution rate [48, 113] or a change in mode of gas flushing [175], both potentially generating a moderate increase in the H_2 level of the system.

4.5. Regulation of reductant disposal pathways

Hydrogen, ethanol and lactate formation are the main routes for reductant disposal in *C. saccharolyticus*. The gene expression of the hydrogenases and both alcohol dehydrogenases involved in ethanol formation are proposed to be under the control of the NADH/NAD^+ sensitive transcriptional regulator Rex [175] (**Figure 4.5**). Cultivation of *C. saccharolyticus* under high P_{H2} conditions was shown to lead to an upregulation of both the hydrogenases and alcohol dehydrogenases [175].

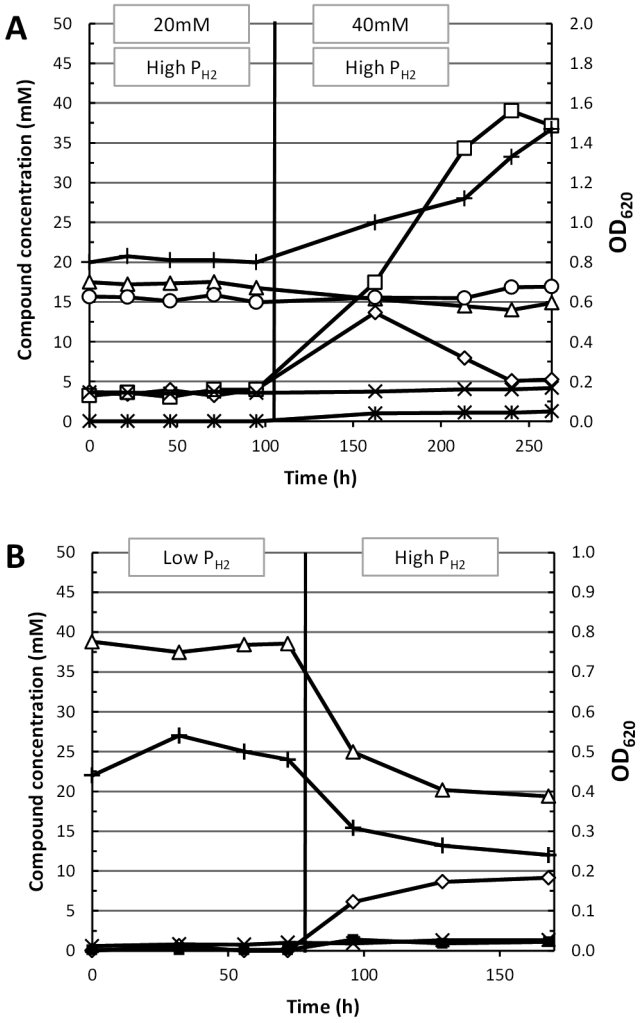


Figure 4.4 Fermentation profile of *C. saccharolyticus* **A)** grown in a chemostat under high H_2 partial pressure at two different glucose concentrations (see section 4.4), **B)** grown in a chemostat without NH_4^+ in the medium under low and high H_2 partial pressure (see section 6.4). The switch from 20mM to 40mM glucose containing-medium and the switch from low to high H_2 partial pressure cultivation condition is indicated by a vertical line. Acetate (open triangles), lactate (open squares), ethanol (crosses), CO_2 (open circles), oxaloacetate (asterisks), glucose out (open diamonds) and CDW (plus sign). Chemostat cultivation parameters (3L reactor, 1L working volume, pH = 7.0 (NaOH), temp 70°C, D = 0.1 h^{-1}, low P_{H2} [sparging (4 L/h) with N_2 gas and stirring speed = 250rpm]), high P_{H2} [headspace flushing (4 L/h) with H_2 gas and stirring speed = 50 rpm]).

In addition, the *ldh* transcript level is also upregulated under high P_{H2} conditions, but in silico analysis of the *ldh* promoter region did not reveal a likely Rex operator binding sequence [175]. Thus the exact regulatory mechanism triggering *ldh* transcription remains therefore unclear. Nonetheless, lactate dehydrogenase activity has been shown to be regulated at the enzyme level, with fructose 1,6 bisphosphate and ATP acting as allosteric activators and both NAD^+ and PP_i as competitive inhibitors [51]. Under high P_{H2} cultivation conditions, enzyme activity assays revealed increased LDH activity with respect to low P_{H2} conditions [51, 113]. A hampered glycolytic flux at the level of GAPDH, potentially triggered by NADH build-up, might lead to the accumulation of fructose 1,6 bisphosphate. In turn, the accumulated level of fructose 1,6 bisphosphate could stimulate lactate formation (**Figure 4.5**), explaining the switch to lactate under high P_{H2} at the enzyme level.

Interestingly, LDH enzyme activity can also be measured under non-lactate producing conditions, which suggest that additional factors control lactate formation [51, 113]. Low levels of lactate formation can be observed at the end of growth during the transition to stationary phase for cultures grown in controlled batch systems [45, 51, 110, 111, 148]. This initiation of the lactate formation coincided with a relative increase in ATP levels and a relative decrease in PP_i levels [51, 107]. These observed changes in ATP and PP_i levels could release the inhibitory effect of PP_i and stimulate LDH activity [51] (**Figure 4.5**). For this latter mode of LDH activity regulation, the lactate formation is proposed to be coupled to the energy metabolism. Accordingly, lactate formation is prevented during exponential growth, which is associated with a high anabolic activity and high PP_i levels. Whereas, during the transition to stationary phase, which is associated with low anabolic activity and relative low PP_i levels, lactate formation inhibition is alleviated [51, 148].

5. Thermodynamic considerations of glucose conversion and H_2 formation

For *C. saccharolyticus* an elevated hydrogen concentration has been shown to affect fermentation performance, leading to a mixed acid fermentation [50, 112, 113, 175]. These observed changes in the fermentation profile as a function of the H_2 concentration can be conceptually explained by considering the thermodynamics of the reactions leading to H_2 formation [27, 147]. The Gibbs energy ($\Delta G'$) for a specific reaction can be calculated from the standard Gibbs energy ($\Delta_r G^{0\prime}$) and the reactant concentrations by the following relation:

(Eq. 4.4) $$\Delta_r G' = \Delta_r G^{\circ\prime} + RT \ln([C]^c[D]^d / [A]^a[B]^b)$$

where $\Delta_r G^{0\prime}$ is the standard Gibbs energy (J/mol, 1 molar concentration of all reactants, at a neutral pH and at a specific temperature), R is the gas constant

(J/mol*K); T is the temperature (K), A and B are the substrate concentrations with respective stoichiometric reaction coefficients a, b; and C and D are the reaction products with respective stoichiometric reaction coefficients c, d.

When H_2 formation occurs in the aqueous phase H_2 supersaturation can occur. For instance, over-saturation of 12 to 34 times the equilibrium concentration has been reported for *C. saccharolyticus* cultivations [112]. This indicates that estimating the dissolved H_2 concentration from the measured H_2 partial pressure by using the equilibrium constant does not always give an accurate representation of the state of the system. So H_2(aq), instead of the H_2 partial pressure (P_{H2}), should be used as a parameter when investigating the effect of H_2 on the metabolism. Therefore, all Gibbs energy calculations discussed herein were performed using the Δ_fG^0 of the dissolved H_2 concentration (H_2(aq)).

Under standard conditions (25°C), acetate formation (**Eq. 4.1**, $\Delta G^{0\prime}$ = -142.6 kJ/reaction) is energetically less favourable then lactate (**Eq. 4.3**, $\Delta G^{0\prime}$ = -195.0 kJ/reaction) and ethanol (**Eq. 4.2**, $\Delta G^{0\prime}$ = -231.8 kJ/reaction) formation. The Gibbs energy ($\Delta G'$) for the complete conversion of glucose to acetate depends, however, on the H_2 concentration (**Eq. 4.1**), whereas the $\Delta G'$ for both ethanol and lactate formation from glucose is independent of the H_2 concentration (**Eq. 4.2** and **Eq. 4.3**, respectively). This means that the complete oxidation of glucose to acetate becomes energetically more favourable when the dissolved H_2 concentration is lowered. For example, acetate formation becomes energetically favourable compared to lactate or ethanol formation when the dissolved H_2 concentration drops below 5.0 or 0.12 mM, respectively (**Figure 4.6A**, **Supplementary Data S4.1**).

However, considering the Gibbs energies of the overall conversions can be misleading, as the thermodynamics of the involved partial reactions can be less favourable. Acetate formation is inevitably linked to H_2 formation and the $\Delta G^{0\prime}$ of the redox reaction coupling the reoxidation of the reduced electron carriers NADH or Fd_{red} to proton reduction are endergonic under standard conditions (**Figure 4.6B**, **Supplementary Data S4.2**). Decreasing the dissolved H_2 concentration will lower the $\Delta G'$ of these H_2 forming reactions, but an extremely low H_2 concentration is required before H_2 formation from NADH becomes exergonic (0.5 µM) (Fd_{red}, 0.2 mM) (**Figure 4.6B**). So, at elevated H_2 concentrations this partial reaction raises a thermodynamic barrier. NADH dependent pyruvate reduction to lactate, on the other hand, has a negative Gibbs energy ($\Delta G^{0\prime}$ = -25.0 kJ/reaction) and is independent of the dissolved H_2 concentration. This means that under certain conditions lactate formation is more feasible, despite the more negative $\Delta G'$ for acetate formation compared to lactate formation (**Eq. 4.1** and **Eq. 4.3**), simply because H_2 formation from NADH is energetically unfavourable.

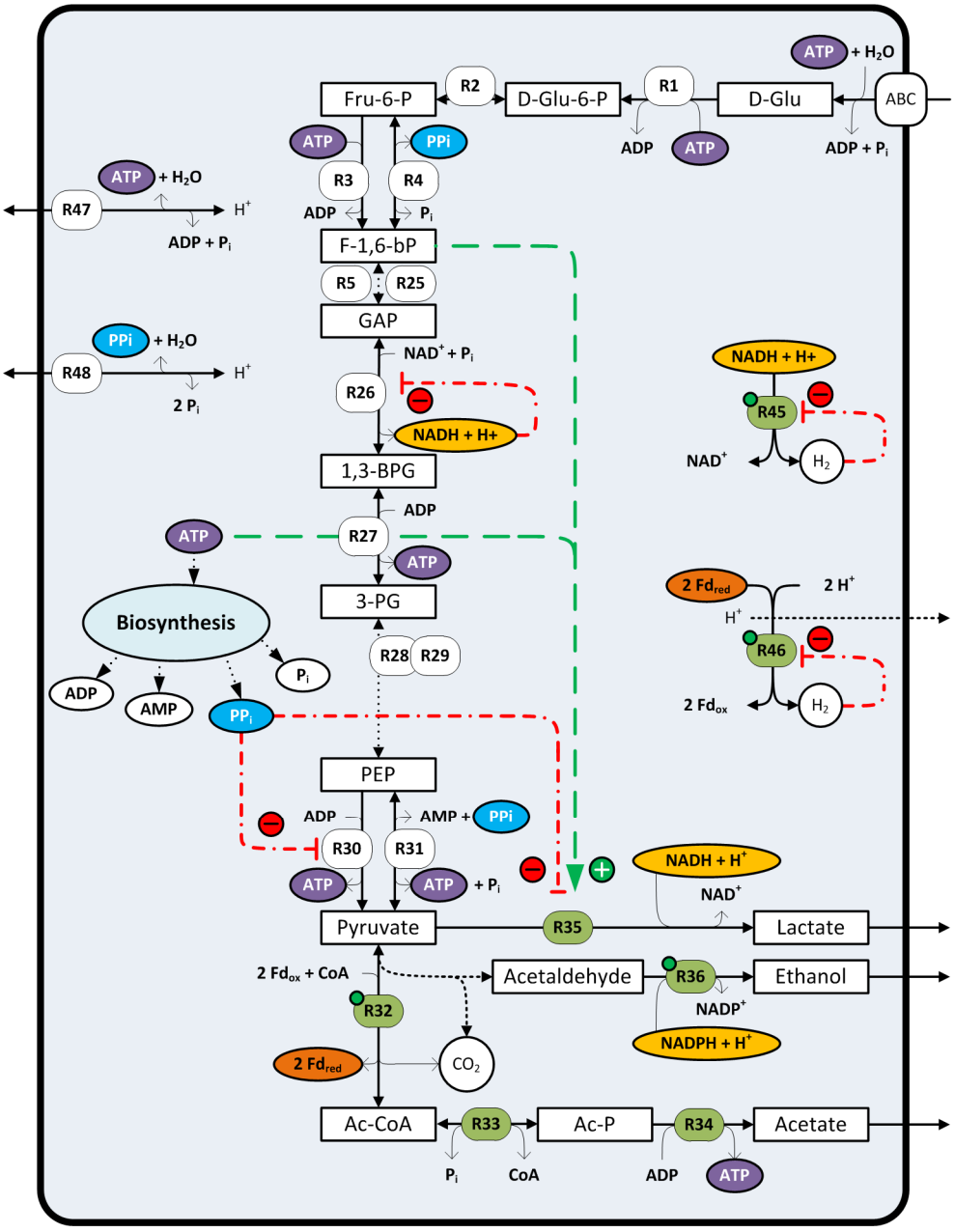

ATP + H2O
ABC
D-Glu
R1
D-Glu-6-P
R2
Fru-6-P
ADP
ATP
ADP + Pi
ATP
PPi
R3
R4
ADP
Pi
ATP + H2O
R47
H+
ADP + Pi
F-1,6-bP
R5
R25
GAP
PPi + H2O
R48
H+
2 Pi
NAD+ + Pi
R26
NADH + H+
1,3-BPG
NADH + H+
R45
NAD+
H2
ADP
ATP
R27
ATP
3-PG
Biosynthesis
ADP
AMP
Pi
PPi
R28
R29
2 Fdred
2 H+
H+
R46
2 Fdox
H2
PEP
ADP
AMP + PPi
R30
R31
ATP
ATP + Pi
NADH + H+
NAD+
Pyruvate
R35
Lactate
2 Fdox + CoA
Acetaldehyde
R36
Ethanol
NADP+
R32
NADPH + H+
2 Fdred
CO2
Ac-CoA
R33
Ac-P
R34
Acetate
Pi
CoA
ADP
ATP

Figure 4.5 Overview of the regulatory mechanisms involved in the central metabolism of *Caldicellulosiruptor saccharolyticus*. The abbreviations of the compounds and reactions (circled numbers) are given in the legend of Figure 4.1. For the enzyme reactions given in green (circled numbers, green) the encoding genes are upregulated under increased P_{H2} [175]. Those which are under the control of the REX transcriptional regulator are marked with a green dot. H_2 inhibits its own formation (R45, R46) and accumulation of NADH inhibits glyceraldehyde-3-phosphate dehydrogenase (R26). PP_i, a by-product of biosynthesis, acts as an inhibitor of both pyruvate kinase (R30) and lactate dehydrogenase (R35) activity. Both ATP and F-1,6-bP are activators of lactate dehydrogenase (R35) activity.

As explained, the thermodynamic barrier associated with H_2 formation from NADH can be lowered by decreasing the dissolved H_2 concentrations, moreover this barrier may also be tackled by an additional input of energy. The latter can be achieved by reverse electron transport catalysed by an NADH: ferredoxin oxidoreductase, where the transfer of electrons from NADH to Fd_{red} is coupled to a proton or ion gradient [106]. Produced Fd_{red} can subsequently be used for H_2 formation. Alternatively, the energetically more favourable oxidation of Fd_{red} can be used to push the less favourable formation of H_2 from NADH in a bifurcating system (**Figure 4.6B**). Such a bifurcating function has been identified for *T. maritima* and was, based on protein sequences, also attributed to the Fe-only hydrogenase (*hyd*) of *C. saccharolyticus* [30]. However, the gene-arrangement of the Fe-only hydrogenase (*hyd*) in *C. saccharolyticus* is identical to that of *Thermoanaerobacter tengcongensis* [40], for which a NADH-dependent hydrogenase activity has been demonstrated [22], thus arguing against a bifurcating systems in *C. saccharolyticus*.

In general, H_2 formation is thermodynamically more favourable at elevated temperatures because i) $\Delta G^{0\prime}$ values of the involved reactions are lower at increased temperature, and ii) the RT coefficient in **Eq. 4.4** is temperature dependent, thus enhancing the effect of a decreased H_2 concentration (**Figure 4.6A** and **4.6B**). Overall, thermophilic organisms have been shown to be able to produce H_2 at higher yields compared to mesophilic organisms. For thermophiles yields approaching the theoretical limit of 4 H_2 per hexose have been reported, while for mesophiles H_2 yields generally do not exceed 2 H_2 per hexose [27, 86, 148]. These higher yields reflect the indicated thermodynamic advantage but also the lower diversity of fermentation end-products observed for those thermophiles. For a specific organism the diversity of available electron acceptors, and formed end-products, depends on the metabolic capabilities of the organism. Whether a specific pathway is operational depends on the regulation of that pathway at the transcription or translation level, but also depends on the kinetic properties and regulation of the enzyme activities of the specific enzymatic steps of that pathway.

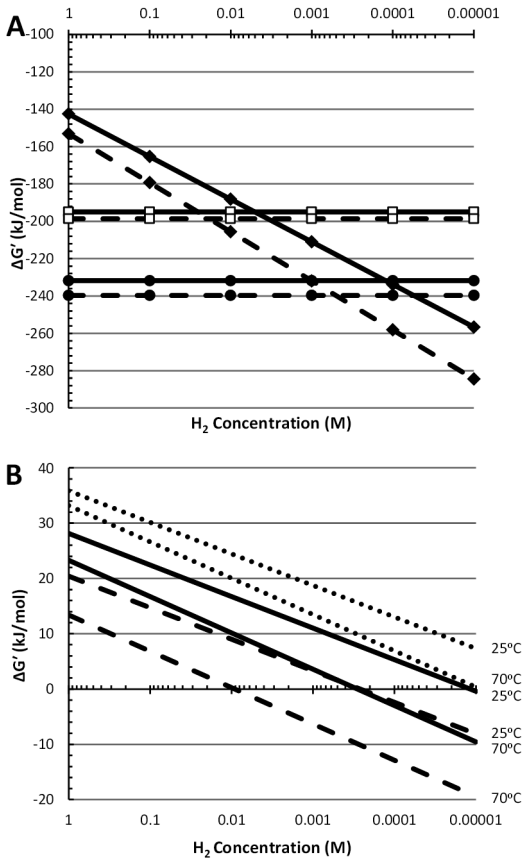


Figure 4.6 Effect of the H_2 concentration on the Gibbs energy change of reactions involved in H_2 formation: **A)** $\Delta G'$ of the complete oxidation of 1 mol of glucose to acetate (closed diamonds), lactate (open squares) and ethanol (closed circles) at 25°C (solid lines) and 70°C (dashed lines); **B)** $\Delta G'$ of H_2 formation from NADH (dotted line), reduced ferredoxin (dashed line) and via the bifurcating system (50% NADH and 50% Fd_{red}) (solid line) at 25°C and 70°C. Values were calculated from data presented in ([4, 28, 29, 31, 32]).

6. Factors limiting H_2 formation

Considering the potential use of complex biomass for H_2 production a multitude of fermentability studies have been performed using *C. saccharolyticus*. An overview of the literature related to fermentability studies on either crop-based feedstock or industrial waste stream derived biomass is given in **Table 4.2**. Those investigations mainly focus on the fermentability of various complex substrates, associated H_2 formation and the effect of pretreatment on substrate accessibility and growth, overall demonstrating the bacterium's broad hydrolytic capacity.

Alternatively, growth experiments on pure sugar substrates can be used to investigate the specific response associated with a certain substrate or to examine specific pathways involved in the metabolism of a substrate. **Table 4.3** gives an overview of the literature related to growth experiments on pure sugars and pure sugar mixes, including the determined H_2 yields and H_2 productivities. The currently available fermentation data are discussed here to highlight the different factors limiting H_2 formation during the fermentative H_2 production by *C. saccharolyticus*.

Table 4.2. Literature overview of fermentability studies on *C. saccharolyticus* using crop-based feedstock or industrial waste stream derived biomass. *Extraction methods and mechanical pre-treatments were excluded in this overview, biological pre-treatment indicated pre-treatment by pre-incubation with *Bacillus amyloliquefaciens*; **Multiple, different pre-treatments used; ^B, batch cultivation and CB, controlled batch cultivation; $Minimal and maximal reported H_2 yields are given.

Reference	Substrate	Pre-treatment*	Cultivation method^	H_2 yields$ / Remarks
[183]	Wheat grains	Enzymatic	B	
	Wheat straw	Acid/Enzymatic	B	
[184]	Barley straw	Acid/Enzymatic**	B	
[42]	Crystalline cellulose	-	B	Proteome data
	Birchwood xylan	-	B	
	Switchgrass	Acid	B	
	Whatman no. 1 filterpaper	-	B	
[185]	Wheat straw	Acid/Enzymatic	B	
	Barley straw	Acid/Enzymatic	B	
	Corn stalk	Acid/Enzymatic	B	
	Corn cob	Acid/Enzymatic	B	
[43]	Poplar	-	B	Microarray data
	Switchgrass	Acid	B	Microarray data
[186]	Beet Molasses	-	CB	0.9-4.2
[47]	Potato steam peels	Enzymatic	CB	1.7-3.4
	Potato steam peels	-	CB	1.1-3.5
[46]	Filter paper	-	B	
	Wheat straw	Biological	B	
	Silphium perfoliatum leaves	Biological	B	
	Maize leaves	Biological	B	
	Sugar cane bagasse	Biological	B	
	Sweet sorghum whole plant	Biological	B	

Table 4.2. continued

[187]	Sugar beet	-	CB	3.0
[188]	Sweet sorghum bagasse	Alkaline/Enzymatic**	B/CB	1.3-2.6
[45]	Carrot pulp	Enzymatic	CB	1.3-2.8
	Carrot pulp	-	CB	
[156]	Crystalline cellulose	-	B	Proteome data
	Cellobiose	-	B	Microarray data
[189]	Switchgrass	-	B	
	Poplar	-	B	
[41]	Xylan	-	B	Microarray data
	Xyloglucan	-	B	Microarray data
	Xyloglucan-oligosaccharides	-	B	Microarray data
[190]	Barley straw	Acid/Enzymatic	B	
	Corn stalk	Acid/Enzymatic	B	
	Barley grain	Enzymatic	B	
	Corn grain	Enzymatic	B	
	Sugar beet	-	B	
[191]	Sweet sorghum plant	-	B	
	Sweet sorghum juice	-	B	
	Dry sugarcane bagasse	-	B	
	Wheat straw	-	B	
	Maize leaves	-	B	
	Maize leaves	Biological	B	
	Silphium trifoliatum leaves	-	B	
[44]	*Miscanthus giganteus*	Alkaline/Enzymatic	B/CB	2.4-3.4
[172]	Agarose	-	B	With different support matrixes
	Alginic acid	-	B	
	Pine wood shavings	-	B	
[192]	Jerusalem artichoke	-	B	Co-fermentation with natural biogas-producing consortia
	Fresh waste water sludge	-	B	
	Pig manure slurry	-	B	
[110]	Paper sludge	Acid/Enzymatic	CB	
[193]	Paper sludge	Acid/Enzymatic	B	

6.1. Comparison between hydrolysates and pure sugar mixtures

Studies on biomass hydrolysates and mono-saccharide mixtures, mimicking the biomass hydrolysates, showed that, while at low substrate loads fermentation performances were comparable, at higher substrate loads H_2 yields were higher on the mixed mono-saccharides compared to the biomass hydrolysates. The difference in yields was caused by a shift to lactate formation during the growth on high substrate load hydrolysates. Interestingly, for the higher substrate concentrations, the total sugar consumption was higher during growth on hydrolysates compared to growth on sugar mixtures [45, 47]. For *C. saccharolyticus,* grown on carrot pulp hydrolysate (20 g/L), a lower cumulative H_2 production was found, compared to growth on a glucose/fructose mixture (20 g/L), while a relatively higher maximal H_2 productivity was observed for the hydrolysate compared to the sugar mixture. During the growth on carrot pulp hydrolysates the relative higher H_2 productivity preceded the switch to lactate formation [45]. These results demonstrate the relation between high H_2 productivity, lactate formation and an overall low H_2 yield. However, these described phenomena were not observed for fermentations on *Miscanthus* hydrolysates. For each tested substrate load (10, 14, and 28 g/L) fermentation performance during growth on the *Miscanthus* hydrolysate was similar to the glucose/xylose sugar mix and only moderate levels of lactate were formed even at high substrate loads [44]. The differences in H_2 production characteristics between the discussed hydrolysates might be related to the difference in sugar composition of the hydrolysates or the differences in pretreatment applied prior to hydrolysis.

In general, biomass derived hydrolysates might contain substances which negatively affect growth or fermentation performance. For example, the dilute-acid pretreatment of lignocellulosic biomass releases undesirable inhibiting compounds like 5-hydroxymethylfurural (HMF), furfural, phenolic compounds and acetate [185] and a growth inhibition of 50% was reported for *C. saccharolyticus* in the presence of 1-2 g/l HMF or furfural [44].

6.2. Incomplete substrate conversion

For chemostat cultivation on glucose (4.4 and 5 g/L) it was shown that a dilution rate (D) exceeding 0.1 h^{-1} gave rise to an incomplete substrate conversion but also to a lower H_2 yield. The concomitant decrease in both biomass level and H_2 yield caused the volumetric H_2 productivity (mmol/L*h) to level off at higher dilution rates [48, 113]. The observed lower H_2 yield reflects the shift in end-product formation, from mainly acetate at a low dilution rate of 0.05 h^{-1} to a mix of acetate and ethanol at a D = 0.15 h^{-1} [113]. Interestingly, no lactate was produced during these chemostat cultivations [48, 113]. The observed incomplete substrate conversion indicated that

another factor was limiting under those conditions. Indeed, increasing the yeast extract to glucose ratio, from 0.25 to 1 g/g, resulted in the almost complete consumption of glucose and also a doubling of cell density (D = 0.3 h^{-1}), which led to an increase in volumetric H_2 productivity (20mmol/L*h). The H_2 yield was, however, not affected [48]. This finding indicated that these observed changes in H_2 yields were a function of the growth rate and did not depend on the substrate conversion efficiency.

6.3. End-product inhibition and osmotolerance

Incomplete substrate conversions can also be observed in controlled batch fermentations with high initial substrate levels. Growth on 10 g/L of both glucose and fructose resulted in complete substrate consumption, while higher initial substrate levels of glucose (20 and 31 g/L) and fructose (20 g/L) resulted in incomplete conversions [45, 47]. Similar observations were done for different sugar mixtures [44, 45, 188]. These incomplete conversions were attributed to the inhibitory effect of accumulating organic acids like acetate or lactate. Inhibition experiments showed that an acetate concentration of 200 mM and higher prevented acid production by *C. saccharolyticus* grown on glucose (10 g/L) in batch [47], which is in line with the earlier findings of van Niel *et al.*, who observed critical sodium acetate and potassium acetate concentrations of 192 mM and 206 mM, respectively, for batch growth on sucrose [50]. However, similar inhibitory effects were observed for NaCl and KCl, with critical concentration of 216 mM and 250 mM respectively [50], suggesting that increased osmolarity is the cause of inhibition and not the end-products per se. This relative low osmotolerance in comparison to marine organisms, like *Thermotoga neapolitana* [47], probably reflects the terrestrial origin of *C. saccharolyticus*. Ljunggren *et al.* designed a kinetic model for the growth of *C. saccharolyticus* incorporating the inhibitory effect of a high osmolarity and determined a critical osmolarity (no growth) in the range of 270 to 290 mM. They also showed that osmolarity is of minor influence on fermentations with low initial glucose levels [112]. This low tolerance to osmotic pressure also prevents the application of CO_2 as a cheaper and more convenient stripping gas than N_2. The use of CO_2 as stripping gas during *C. saccharolyticus* cultivations negatively affects growth rate and hydrogen productivity. CO_2 sparging led to a higher dissolved CO_2 concentration, which required addition of extra base to maintain a constant pH, overall leading to an increase in osmotic pressure [111].

6.4. Medium requirements

Controlled batch cultivations were used to investigate the influence of NH_4^+ on the performance of *C. saccharolyticus* grown on molasses. These experiments revealed that the omission of NH_4^+ gave rise to a higher H_2 yield and maximal H_2 productivity [186]. Although *C. saccharolyticus* is able to grow on a medium without NH_4^+, containing only YE as a nitrogen source, it becomes very sensitive to changes in P_{H2} (**Figure 4.4B**). Chemostat cultivations showed complete glucose (20.7 mM) consumption and high acetate yields (1.87 ± 0.02 mol/mol) under low P_{H2}. However, when the cultivation condition changed to a high P_{H2} a new steady state could be achieved, but substrate consumption was incomplete (55%) and acetate yields decreased (1.68 ± 0.01 mol/mol).

Omission of yeast extract (YE) during growth of *C. saccharolyticus* on molasses did not affect the H_2 yield but led to a lower volumetric H_2 productivity [186]. Similar observations were made for controlled batch fermentations on glucose, where the absence of YE did not affect the H_2 and biomass yields [194]. Contrary to the molasses study the volumetric productivity was not affected [194] but this might be due to the lower substrate load used. *C. saccharolyticus* is able to grow on a defined minimal medium with additional vitamins, but without additional amino acids [194]. Growth in the absence of YE helps to reduce the production costs. However, increased biomass levels and especially growth on high substrate loads might augment medium requirements. So fine-tuning of the medium composition with respect to the specific substrate and substrate load is required.

Table 4.3 Literature overview of fermentation studies on *C. saccharolyticus* grown on pure sugar substrates. *Yields in mol H_2/mol hexose; nd, not determined; **For batch cultivations the maximal productivity is given; nd, not determined.^B, batch cultivation and CB, controlled batch cultivation.

Reference	Substrate	Substrate load (g/l)	H_2 Yield*	Productivity** (mmol/(L*h))	Cultivation method^	Dilution rate (h^{-1})
[194]	Glucose	10	3.0	20.0 mol/(g*h)	CB	
		10	3.4	23.6 mol/(g*h)	CB	
		4	3.5	10.1 mol/(g*h)	Chem	0.05
		4	3.5	10.4 mol/(g*h)	Chem	0.05
[113]	Glucose	5	3.5	5.2	Chem	0.05
		5	2.9	11.0	Chem	0.15
		5	1.8	2.5	Chem	0.05
		5	wash out	wash out	Chem	0.15

Table 4.3 continued

[195]	Glucose	5	nd	nd	B
[47]	Glucose	10	3.4	12.0	CB
		31	2.8	12.9	CB
[187]	Sucrose	10	2.9	7.1	CB
[188]	Glucose/ Xylose /Sucrose (6:2.5:1.5, w/w/w)	10	3.2	10.7	CB
		20	2.8	9.4	CB
[45]	Glucose	10	3.2	11.2	CB
		20	3.4	12.2	CB
	Fructose	10	2.6	13.2	CB
		20	2.4	13.4	CB
	Glucose/ Fructose (7:3, w/w)	10	3.0	13.2	CB
		20	2.6	12.2	CB
[156]	Xylose	5	nd	nd	B
	Glucose	5	nd	nd	B
[170]	Glucose	5	nd	nd	B
[111]	Sucrose	4	2.7	23.0	CB
		4	3.1	11.8	CB
	Glucose	4	3.0	20.0	CB
		4	2.7	12.0	CB
[41]	Glucose	0.5	nd	nd	B
	Mannose	0.5	nd	nd	B
	Arabinose	0.5	nd	nd	B
	Xylose	0.5	nd	nd	B
	Fructose	0.5	nd	nd	B
	Galactose	0.5	nd	nd	B
	mix (0.5 g/L each)	3	nd	nd	B
[44]	Glucose/ Xylose (7:3 w/w)	10	3.4	12.0	CB
		14	3.3	10.1	CB
		28	2.4	9.7	CB

Table 4.3 continued

[196]	Sucrose	10.3	2.8	22.0	Trickle bed	0.2-0.3
[40]	Glucose	4	nd	nd	CB	
	Xylose	4	nd	nd	CB	
	Rhamnose	4	nd	nd	CB	
	Glucose/ Xylose (1:1, w/w)	4	nd	nd	CB	
[48]	Glucose	4.4	3.3	4.2	Chem	0.05
		4.4	3.6	8.9	Chem	0.10
		4.4	2.9	9.5	Chem	0.15
		4.4	2.9	9.1	Chem	0.20
		4.4	3.1	11.0	Chem	0.30
		4.4	3.0	12.4	Chem	0.35
		1.9	4.0	4.0	Chem	0.09
		1.9	3.3	9.9	Chem	0.30
		4.1	3.5	7.7	Chem	0.09
		4.1	3.1	11.6	Chem	0.30
[110]	Glucose	10	2.5	10.7	CB	
	Xylose	10	2.7	11.3	CB	
	Glucose/ Xylose (11:3, w/w)	8.4	2.4	9.2	CB	
[21]	Sucrose	10	3.3	8.4	CB	

7. Future prospects for improving bioH_2 production

7.1. Improving H_2 yields and H_2 productivity

C. saccharolyticus has many properties that make it an excellent candidate for bioH_2 production via dark fermentation. However, for bioH_2 production to become economically feasible major improvements should be made with respect to the H_2 productivity [197-199]. Productivity is maximized by improving the substrate consumption rate but also the H_2 yield.

C. saccharolyticus is able to produce H_2 close to the theoretical maximum of 4 H_2 per hexose. In this respect it is important to realize that the theoretical yield refers to

the pure catabolic component of glucose conversion. The glucose used for anabolism should not be incorporated. Generally this distinction is not considered in literature, and reported experimental data therefore reveal H_2 yields lower than 4. For example, a yield of 3.5 H_2 per consumed glucose has been reported by de Vrije *et al.* (chemostat cultivation with a 23.0 mM glucose load and a dilution rate of 0.1 h^{-1}) [48]. According to their data 16% of the consumed glucose is used only for biomass formation indicating that only ~19.4 mM glucose was available for ATP generation. Given the reported H_2 production this results in a theoretical conversion efficiency of 4.15 mol H_2 per mol glucose, which approximates the theoretical maximum of 4 H_2 per hexose as indicated by Thauer *et al*. [4]. These high yields are only achievable when the organism ferments the substrate solely to acetate. However, non-ideal growth conditions lead to a mixed fermentation profile (acetate, lactate and ethanol), and a consequently lower H_2 yield. From the organism's perspective switching to lactate or ethanol is profitable since it allows the organism to continue to grow under elevated P_{H2} conditions, albeit with a lower growth rate because ATP yields are lower under lactate and ethanol forming conditions. Obviously, the lower H_2 yield under a mixed end-product fermentation, is not desirable from a biotechnological point of view. To maximize the H_2 yield and productivity, the dissolved H_2 concentration should be kept as low as possible, which requires an optimization of the liquid to mass transfer rate [4, 112, 182], which is mainly a matter of reactor design. In continuous stirred-tank reactor systems low dissolved H_2 concentrations could be achieved by increasing the sparging rate [200, 201] or the stirring speed [202, 203]. In addition, reduction of internal reactor pressure [204-206] and enforced bubble formation [206, 207] could potentially lead to a lower dissolved H_2 concentration. Addition of zeolite particles, which enhances bubble formation, allowed a reduction of the N_2 stripping rate from 5 L/(h*L) to 1 L/(h*L), without affecting the H_2 productivity and H_2 yield of *C. saccharolyticus*. N_2 stripping could even be completely omitted when an internal reactor pressure of 0.3 bar was used, which sustained a similar fermentation performance compared to cultivation at atmospheric pressure (1bar) using a N_2 gas stripping rate of 5 L/(h*L) [206].

Most research on *C. saccharolyticus* has been performed in serum bottles and suspended continuous stirred-tank reactor systems, where only relatively low cell biomass levels can be achieved. Higher cell biomass levels would cause an increase in substrate consumption rates leading to an increase in H_2 productivity when H_2 yields are maintained. To realize higher biomass levels a deeper insight into growth limiting medium compounds should be acquired. Additionally, other reactor types, like a trickle bed reactor or a fluidized bed system might allow cell biomass accumulation. *C. saccharolyticus* could be cultivated in a non-axenic 400L trickle bed reactor [196] out-competing other organisms, with H_2 yields around 2.8 mol H_2/mol hexose and a

productivity of 22 mmol H_2/L*h. These results showed that *C. saccharolyticus* can be used in a large scale non-sterile industrial setting.

Combining different organisms in a co-cultivation setup allows the exploitation of the hydrolytic capability of each individual species and could enhance the overall range of useable substrates. Batch co-cultivations of *C. saccharolyticus* with either *C. owensensis*, *C. kristjanssonii* or an enriched compost microflora were performed on a glucose-xylose mixture [208]. The co-cultivation with the enriched compost microflora resulted in a fast, simultaneous consumption of both glucose and xylose with a relatively high specific hydrogen production rate, but with a lower H_2 yield [208]. A stable co-culture consisting of two closely related *Caldicellulosiruptor* species, *C. saccharolyticus* and *C. kristjanssonii*, could be established in a continuous cultivation system [209]. These findings demonstrate the possibility to create co-cultures for H_2 formation and reveal an apparent synergistic effect of the strains which lead to improved fermentation performances.

Overall H_2 yields from biomass derived substrates can be increased when the dark fermentation is coupled to a second stage like electrohydrogenesis or photofermentation [11, 210]. The former system uses a microbial electrolysis cell (MEC), in which electricity is used to convert acetate or other organic acids to hydrogen. In the latter case the main end-product of the dark fermentation, acetate, is further converted by an anaerobic non-sulfur purple photosynthetic bacterium, forming a maximum of 4 mol H_2 per mol acetate, giving an overall H_2 yield of 12 mol H_2 per mol glucose. The effluent of *C. saccharolyticus* has been successfully used as a feed for photofermentative growth and H_2 production [186, 211, 212]. Alternatively, dark fermentation end-products H_2 and acetate could serve as substrates for hydrogenothropic methanogens in a biogas generating system. The addition of *C. saccharolyticus* to natural biogas-producing consortia led to an improvement of biogas production and a stable co-cultivation could be maintained for several months [192].

7.2. Genetic engineering of Caldicellulosiruptor species

The first steps in the development of a genetic system for *Caldicellulosiruptor* species have been made. Chung *et al.* have shown that methylation with an endogenous unique α-class N4-Cytosine methyltransferase is required for transformation of DNA isolated from *E. coli* into *Caldicellulosiruptor bescii* [213]. Furthermore, an uracil auxotrophic *C. bescii* mutant strain was generated by a spontaneous deletion in the *pyrBCF* locus [213]. This nutritional deficiency was exploited as a selection marker of *C. bescii* transformants [213]. A similar strategy might be applied to develop a genetic system for the other members of this genus.

To improve the H_2 producing capabilities of *C. saccharolyticus* or other *Caldicellulosiruptor* species metabolic engineering strategies could focus on improving

the H_2 yields. Additionally it could be aimed at altering intrinsic properties of *Caldicellulosiruptor* species limiting H_2 productivity like the enhancement of their H_2 tolerance or osmotolerance. For example, for industrial applications increased substrate concentrations are favored since it reduces the fresh water demand, thus reducing the overall costs and the environmental impact of the process. However, for *C. saccharolyticus* the maximum substrate load is limited by its sensitivity to osmotic stress [50, 112, 148].

With respect to the mixed acid fermentation, both lactate and ethanol formation lead to a lowered H_2 yield. Because ethanol formation is not the major reductant sink under redox stress and in general is only produced at low levels, knocking out the alcohol dehydrogenase responsible for ethanol formation will probably not significantly alter the fermentation performance of *C. saccharolyticus*. However, lactate formation can be seen as a the main mechanism to alleviate redox stress. Targeting the lactate formation pathway for complete knockout will probably make *C. saccharolyticus* less resilient to fluctuations in dissolved H_2 concentration and is inadvisable.

Alternatively, one could alter glycolysis in such a way that substrate conversion is less energy efficient. So to generate the same amount of ATP, essential for biosynthesis and maintenance, a higher glycolytic flux to acetate is required. This would result in a higher H_2 yield because the glycolytic flux to acetate is increased with respect to the carbon flux to biomass. A less energy efficient glycolysis can be achieved by eliminating some ATP generating steps from the central metabolic pathway. For example, exchanging the NADH-dependent GAPDH in *C. saccharolyticus* with the ferredoxin-dependent GAPOR, would decrease the overall ATP yield of glycolysis. In addition, since the H_2 formation from the generated Fd_{red} is energetically more favourable than H_2 formation from NADH, the organism would become less sensitive to increased H_2 levels [27, 148].

H_2 yields on rhamnose and fucose could be increased if the carbon flux from the intermediate lactaldehyde is redirected via methylglyoxal to pyruvate, by the insertion of a lactaldehyde dehydrogenase and a methylglyoxal dehydrogenase. When rhamnose is completely oxidized to acetate, via this pathway, the H_2 yield increases from 1 to 5 H_2 per rhamnose. With respect to alternative substrates, glycerol could serve as a good substrate for H_2 production because of the relative high reduced state of its carbon atoms. A maximal H_2 yield of 3 mol/mol glycerol is achieved if glycerol is completely oxidized to acetate. So far growth or co-consumption on/of glycerol has, however, not been observed for *C. saccharolyticus* (unpublished results, A.A.M. Bielen).

8. Conclusions

The bacterium *Caldicellulosiruptor saccharolyticus* possesses several features that make it an excellent candidate for biological hydrogen production. With an optimal growth temperature of 70°C *C. saccharolyticus* is one of the most thermophilic cellulose degrading organisms known to date. The organisms diverse inventory of endo- and exo- glycoside hydrolases allow it to degrade and grow on a variety of cellulose- and hemicellulose-containing biomass substrates. Some of these glycosidases are multi-domain proteins that contain both glycoside hydrolase domains and carbon binding modules which facilitate the efficient degradation of recalcitrant plant polysaccharides into mono-, di- or oligo- saccharides. The high diversity of transport systems present in the genome confirm the broad substrate preferences of *C. saccharolyticus* and its ability to co-utilization hexoses and pentoses, without any signs of carbon catabolite repression, is a desirable trait for any consolidated bioprocess.

For *C. saccharolyticus* sugar substrates are primarily fermented to acetate, CO_2 and H_2, via the EMP pathway. Typically, the fermentation of hexose and pentose lead to the generation of the reduced electron carriers NADH and Fd_{red} in a 1:1 ratio. These reduced electron carriers can be reoxidized during two distinct H_2 generating steps, respectively catalyzed by an NADH-dependent cytosolic Fe-only hydrogenase (*hyd*) and the Fd_{red}-dependent membrane-bound [NiFe] hydrogenase (*ech*). Alternatively, rhamnose catabolism is coupled to a different type of redox balancing, where the generated NADH is used for 1,2-propanediol formation and only the Fd_{red} is available for H_2 formation.

Fermentation data reveal that *C. saccharolyticus* is capable of producing H_2 with yields close to the theoretical limit of 4 H_2 per hexose. However, under non-ideal conditions both ethanol and lactate formation act as alternative redox sinks, thus reducing H_2 yields. All possible redox sinks, including the hydrogenases, are upregulated during cultivation under an increased partial hydrogen pressure. The mechanism underlying transcription of the lactate dehydrogenase gene remains elusive, but the transcription of the genes coding for both hydrogenases and the alcohol dehydrogenases, potentially involved in ethanol formation, seem to be under the control of an $NADH/NAD^+$-sensing transcriptional regulator REX.

An increased intracellular $NADH/NAD^+$ ratio, putatively caused by the inhibition of hydrogenase activity at elevated H_2 levels, can hinder glycolysis at the level of glyceraldehyde-3-phosphate dehydrogenase, resulting in the inhibition of growth. Lactate formation serves as an alternative redox sink, alleviating redox stress. Lactate dehydrogenase activity is enhanced by the glycolytic intermediate fructose-1,6-bisphosphate but also modulated by the energy carriers ATP and pyrophosphate. The latter mechanism couples lactate formation to the energy metabolism, where lactate

formation is inhibited during exponential growth and inhibition is alleviated during the transition to the stationary phase.

Overall, maintaining low dissolved H_2 levels in the system appeared to be one of the most important factors for optimizing H_2 production. In addition, improvements should be made with respect to the H_2 productivity and osmotolerance of the organism to allow bioH_2 production by *C. saccharolyticus* to become economically feasible.

Supplementary materials

Supplementary Data S4.1
Supplementary Data S4.1

Supplementary materials can be found online.
(doi:10.3390/life3010052)

Chapter 5

Pyrophosphate as a central energy carrier in the hydrogen producing extremely thermophilic *Caldicellulosiruptor saccharolyticus*

This chapter has been published as:

A.A.M. Bielen[#], K. Willquist[#], J. Engman, J. van der Oost, E.W.J. van Niel, S.W.M. Kengen. *Pyrophosphate as a central energy carrier in the hydrogen producing extremely thermophilic Caldicellulosiruptor saccharolyticus*. FEMS Microbiology Letters, 2010 307(1): 48-54

[#] These authors contributed equally to this paper

Abstract

The role of inorganic pyrophosphate (PP_i) as an energy carrier in the central metabolism of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* was investigated. In agreement with its annotated genome sequence, cell extracts were shown to exhibit PP_i-dependent phosphofructokinase and pyruvate phosphate dikinase activity. In addition, membrane-bound pyrophosphatase activity was demonstrated, while no significant cytosolic pyrophosphatase activity was detected. During the exponential growth phase high PP_i levels (approx. 4±2 mM) and relatively low ATP levels (0.43±0.07 mM) were found, and the PP_i/ATP ratio decreased 13-fold when the cells entered the stationary phase. Pyruvate kinase activity appeared to be allosterically affected by PP_i. Altogether, these findings suggest an important role for PP_i in the central energy metabolism of *C. saccharolyticus*.

1. Introduction

The extreme-thermophilic and strictly anaerobic bacterium *Caldicellulosiruptor saccharolyticus* belongs to the class of the Clostridia. This bacterium has potential for industrial applications because of its ability i) to produce high hydrogen levels [48], ii) to grow on complex lignocellulosic material [44, 172], and iii) to co-metabolize a number of monosaccharides without revealing any form of carbon catabolite repression [40, 41]. For these reasons *C. saccharolyticus* recently became the subject of various research projects focusing on renewable energy production [21, 44, 172].

The classical EMP pathway is the main route of glycolysis in this organism [48], and analysis of the *C. saccharolyticus* genome sequence has revealed the presence of all the EM-pathway enzymes [40]. However, the authors of this study indicated further that the *C. saccharolyticus* genome contains genes coding for a PP_i-dependent pyruvate phosphate dikinase (PPDK) in addition to the pyruvate kinase (PK). Genes coding for typical gluconeogenic enzymes like pyruvate water dikinase (or PEP synthase) and fructose bisphosphatase are absent [40]. Interestingly, recent studies on the acetate-lactate metabolic shift in *C. saccharolyticus* revealed that PP_i is a strong modulator of the lactate dehydrogenase [51]. These observations motivated us to investigate the role of inorganic pyrophosphate in the energy metabolism of *C. saccharolyticus*. PP_i-dependent reactions have regularly been described for plants and primitive eukaryotes [68]. However, little is known about PP_i-dependency in heterotrophic prokaryotes.

2. Material and methods

2.1 Organism and growth condition

C. saccharolyticus DSM 8903 [37] was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen. For the enzyme and nucleotide measurements cell extracts were prepared from *C. saccharolyticus* cells, which were cultured batch-wise in pH-controlled reactors and in a medium as previously described [40, 111], using glucose as carbon source (4 g/L, for determination of enzyme levels and 10g/L for determination of nucleotide levels). For the determination of nucleotide levels the working volume was 1.7 L to minimize the effect of sampling on the culture.

2.2 Analytical methods

Growth was monitored by measuring the optical density (OD) at 620 nm (Hitachi U-1500, Tokyo, Japan). Cell dry weight (cdw) was determined using a filtration method

as previously described [51]. Protein levels were determined according to Bradford [214] using bovine serum albumin as standard.

2.3 Bioinformatics

Nucleotide sequences of investigated genes were retrieved either from the IMG database (http://img.jgi.doe.gov) or GenBank (http://www.ncbi.nlm.nih.gov/). Sequence alignments were performed using CLUSTAL X (V1.83) [215]. Molecular phylogenetic analysis was done by using the distance (neighborhood-joining) method. Gene-neighborhood analysis was performed using the ortholog neighborhood viewer available at the IMG site.

2.4 Enzyme assays

Cells (OD_{660} 0.3-0.4) were harvested (50 mL) during the mid-logarithmic growth phase by centrifugation (10 min, 4570 × g). Cell pellets were suspended in Tris-HCl buffer (100 mM, pH 7.2) containing $MgCl_2$ (5 mM) and NaCl (40 mM)) [51]. Cells were disrupted by sonification and cell extracts (CE) were prepared by centrifugation (10 min, 16,000 × g). Membrane and cytosolic fractions were obtained by additional centrifugation (1 h, 100,000 × g) of the CE, where the membranes were resuspended in the indicated buffer. Extracts were stored at -20°C until further use. The determination of enzyme activities were carried out with two biological and four technical replicates.

Enzyme activities of PPDK (EC 2.7.9.1), PK (EC 2.7.1.40), ATP- and PP_i-dependent PFK (ATP-PFK, EC 2.7.1.11, PP_i-PFK, EC 2.7.1.90) were determined in the described Tris-buffer which was additionally reduced with dithiothreitol (5 mM). Assays were performed at 50°C, to secure auxiliary enzyme activity. All auxiliary enzymes were purchased from Sigma Aldrich. Substrate conversions were coupled to the oxidation of NADH (ε_{334} = 6.18 $mM^{-1}cm^{-1}$). The PPDK and PK assays were coupled to the conversion of pyruvate to lactate using L-lactate dehydrogenase (LDH) as auxiliary enzyme. To determine the influence of the PP_i concentration on PK activity, low molecular weight compounds were first removed from CEs (M_W below 5 kDa) by using a PD10 desalting column (GE health care; [51]). These later assays were performed at 70°C using a thermostable LDH as auxiliary enzyme. The PFK activity was assayed by coupling to the reduction of dihydroxyacetone phosphate (DHAP) to glycerol 3-phosphate, catalyzed by glycerol-3-phosphate dehydrogenase (GPDH), using fructose 1,6-bisphosphate aldolase (FBA) and triose-phosphate isomerase (TPI) as auxiliary enzymes. One unit (U) of enzyme activity is defined as that amount of the enzyme that catalyzes the conversion of 1 µmole of substrate per minute. The reaction mixtures contained for PPDK: LDH (6.8 U, from chicken heart), NADH (0.25 mM), NH_4Cl (25

mM), phosphoenolpyruvate (PEP) (2 mM), AMP (2 mM), PP_i (0.4 mM); for PK: LDH (6.8 U), NADH (0.25 mM), PEP (2 mM), ADP (2 mM); for PP_i-PFK: GPDH (1.3 U, from rabbit muscle), FBA (0.8 U, from rabbit muscle), TPI (0.8 U, from rabbit muscle), NADH (0.25 mM), fructose 6-phosphate (F6P) (2 mM) and PP_i (0.4 mM); and for ATP-PFK: GPDH (1.3 U), FBA (0.8 U), TPI (0.8 U), NADH (0.25 mM), F6P (2 mM) and ATP (2 mM). At the end of each assay, the auxiliary enzymes were checked to be non-limiting by the addition of pyruvate (5mM) for the PPDK and the PK assays and fructose 1,6-bisphosphate (5mM) for the PFK assays.

Pyrophosphatase (inorganic diphosphatase, PPase, EC 3.6.1.1) activity was determined at 70°C in the indicated buffer. Hydrolysis of PP_i (0.4 mM) was followed by measuring the formation of P_i in time, in a discontinuous spectrophotometric assay (630 nm) using a malachite green detection method [216]. As a negative control either PP_i or the extract was omitted from the assay.

2.5 Determination of ATP, ADP and pyrophosphate levels

To determine the intracellular concentrations of ATP, ADP and PP_i, cell suspensions (15 mL) were collected from the fermentor at different points during growth. Three biological and six technical replicates were performed for each condition. The cell suspensions were quenched with 10 g ice (distilled H_2O) and centrifuged (3 min, 18,000 × g), pellets were washed with cold NaCl solution (0.91% w/v, 0°C). After the second centrifugation step (3 min, 18,000 × g), the pellet was resuspended in 500 µl $HClO_4$ (30%) and immediately frozen (-80°C) until further analysis. The supernatants from both centrifugation steps were analyzed for ATP to determine possible cell leakage. The nucleotides and PP_i were extracted using a method adapted from Cole *et al.* [217]. The extraction recovery of ATP, determined according to Meyer and Papoutsakis [218], was 74±4 %. Based on the findings of Meyer and Papoutsakis the extraction recovery for ADP was assumed to be the same as determined for ATP. For PP_i it was assumed that losses during extraction were negligible [219]. ADP was converted to ATP using pyruvate kinase (1.98 U/mL) (Sigma, St. Louis, USA), PEP (240 µM), KCl (100 mM) and $MgCl_2$ (1 mM). The ATP concentration was determined using an ATP bioluminescent assay kit (Sigma, St. louis, USA). Substantial amounts of ATP leaked out of the cell during extraction, i.e. after the first and the second centrifugation step the leakage was 68% and 3% of total ATP, respectively. Therefore, the total level of ATP and ADP (AXP) were estimated according to **Eq. 5.1**:

(**Eq. 5.1**) $$AXP = 3.5 \times \frac{AXP_{measured}}{0.74}$$

Where $AXP_{measured}$ is the ATP or ADP concentration determined after extraction and 3.5 is the correction factor, which accounts for leakage. The level of PP_i was determined using a Pyrophosphate Assay kit (P_iPER™, Invitrogen, Carlsbad, USA). Because of a relatively high P_i concentration of the growth medium, leakage of PP_i could not be determined so PP_i levels were not corrected for possible leakage. The nucleotide and PP_i intracellular concentrations were calculated on the basis that 1 g cdw (~5.5 g/L wet weight), corresponds to an intracellular volume of 4.58 mL. The cell dimension of *C. saccharolyticus* is 0.35×3.5 μm [37] and 1 g cdw equals ca 1.36×10^{13} cells [21].

3. Results and discussion

3.1 Genome analysis

All the enzymes making up the classical EMP pathway have been identified in *C. saccharolyticus* [40], including two genes that were annotated to code for a phosphofructokinase (Csac_1830 and Csac_2366). A sequence alignment of these kinases against PFK-A family members (data not shown) and a more detailed analysis of their amino acid sequences (**Figure 5.1**), revealed that Csac_1830 belongs to the B1 group and contains the typical ATP-dependent PFK amino acid combination G_{104} and G_{124} [220, 221]. On the other hand, Csac_2366 belongs to the B2 group and contains the typical PP_i-dependent PFK amino acid residues D_{104} and K_{124} [221, 222]. These results strongly suggest that Csac_1830 is an ATP-dependent PFK and that Csac_2366 is a PP_i-dependent PFK. The genome also contains the genes encoding a PK (Csac_1831) and a PPDK (Csac_1955), which are both able to perform the catabolic conversion of phosphoenolpyruvate to pyruvate [40]. PPDK catalyzes a PP_i-dependent reaction, whereas PK requires ADP (**Figure 5.2A**). The anticipated PP_i-dependence of the specified glycolytic steps, prompted us to look for other PP_i-dependent reactions. The *C. saccharolyticus* genome lacks any homolog to cytosolic pyrophosphatases (COG0221, COG1227 (Bacillus type; [223]). Instead, it was found to contain a gene coding for a V-type proton pumping membrane-bound pyrophosphatase (H^+-PPase, COG 3808, Csac_0905). The anticipated H^+-PPase has 14 predicted membrane helices and is expected to be an integral membrane protein. Sequence-based phylogenetic analysis of the H^+-PPase revealed it as a member of the K^+-insensitive group of H^+-PPase protein family (data not shown; [224]).

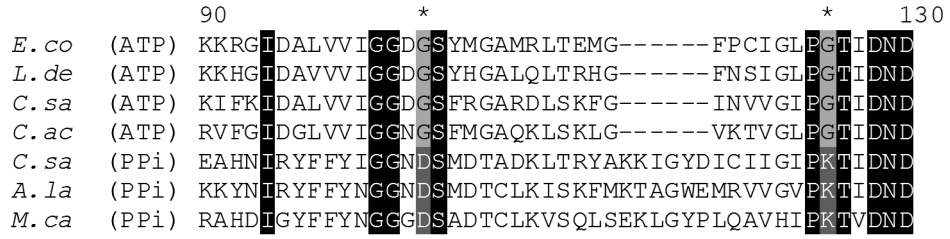


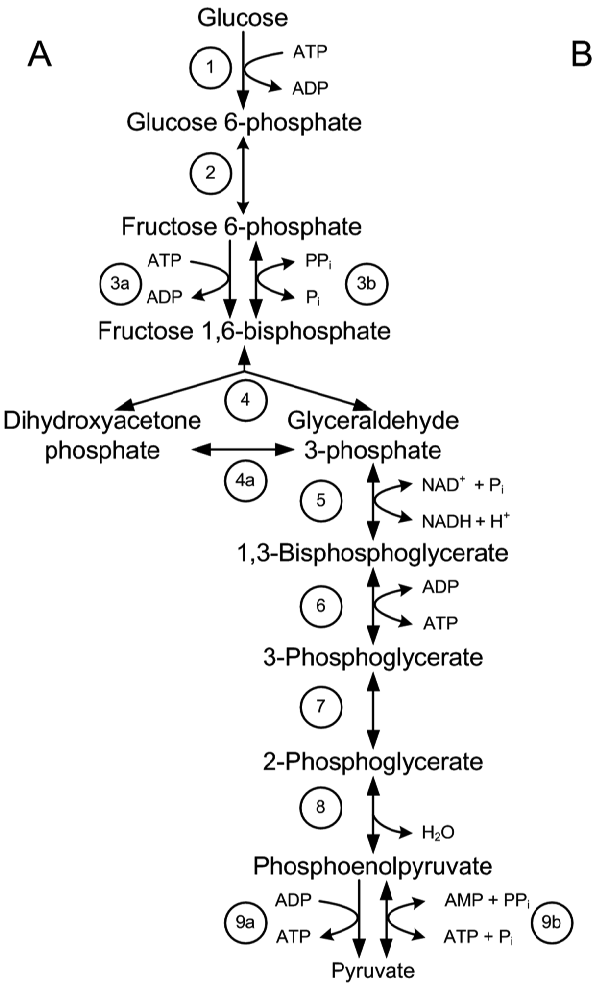
Figure 5.1 Alignment of PFK partial sequences from *Caldicellulosiruptor saccharolyticus* (*C.sa*, YP_001181135 (PP_i), YP_001180606 (ATP)) with B1 group members: *Escherichia coli* (*E.co*, NP_418351), *Lactobacillus delbrueckii* (*L.de*, YP_812811), *Clostridium acetobutylicum* (*C.ac*, NP_347157) and B2 group members: *Acholeplasma laidlawii* (*A.la*, YP_001621078) and *Methylococcus capsulatus* (*M.ca*, YP_113714). All corresponding proteins of the given sequences have been biochemically characterized and their phosphor-donors are given. (*) indicates the 104 and 124 amino acid position of the *E. coli* PFK protein.

3.2 Enzyme analysis

To confirm the genome sequence based predictions, we performed enzyme assays on crude cell extracts. Activities of PK, PPDK, PP_i-PFK, ATP-PFK and membrane-bound PPase could all be detected in CEs of *C. saccharolyticus* (**Table 5.1**). Under the specified assay conditions the average PPDK activity (0.4 U/mg) was twice the PK activity. For the ATP- and PP_i-dependent PFKs no significant difference was observed in the activity levels (~0.05 U/mg). In line with the presence of a membrane-associated pyrophosphatase, high levels of PPase activity (~0.15 U/mg) could be demonstrated in the membrane fraction, while membrane-free cell extract barely showed PPase activity. Whether the membrane-bound PPase in *C. saccharolyticus* couples pyrophosphatase activity to the translocation of protons across the membrane remains to be investigated.

Table 5.1 Enzyme activities (U/mg) of CE derived from exponentially growing *C. saccharolyticus* cultures PPase activity was determined in cell extract (CE), in the membrane fraction (MF) and in the cytosolic fraction (CF).

PPDK	PK	PPi-PFK	ATP-PFK	PPase (CE)	PPase (MF)	PPase (CF)
0.42 ± 0.09	0.18 ± 0.05	0.041 ± 0.006	0.049 ± 0.004	0.067 ± 0.017	0.15 ± 0.06	0.004 ± 0.003


A
Glucose
1
ATP
ADP
Glucose 6-phosphate
2
Fructose 6-phosphate
3a
ATP
ADP
PPi
Pi
3b
Fructose 1,6-bisphosphate
4
Dihydroxyacetone phosphate
Glyceraldehyde 3-phosphate
4a
5
NAD+ + Pi
NADH + H+
1,3-Bisphosphoglycerate
6
ADP
ATP
3-Phosphoglycerate
7
2-Phosphoglycerate
8
H2O
Phosphoenolpyruvate
9a
ADP
ATP
AMP + PPi
ATP + Pi
9b
Pyruvate

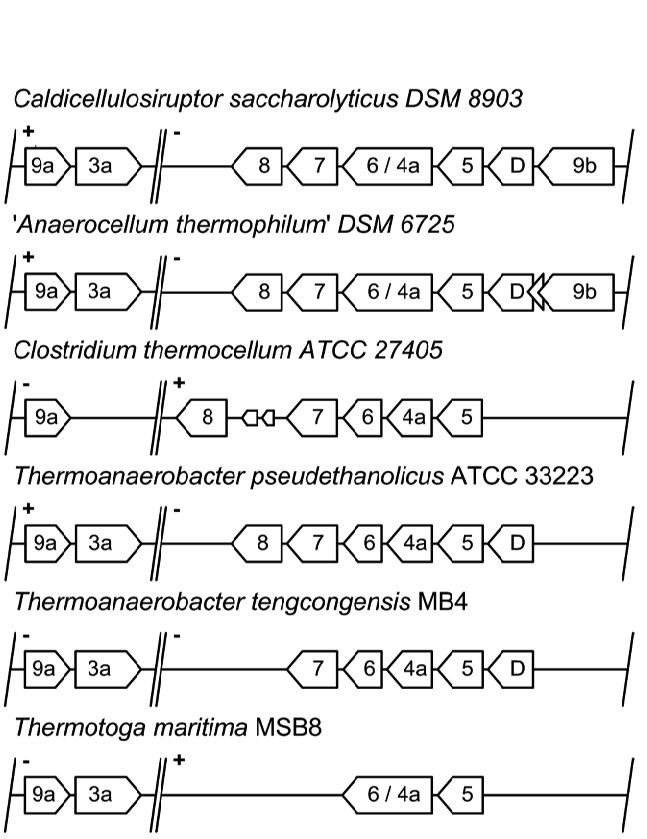

B
Caldicellulosiruptor saccharolyticus DSM 8903
+
-
9a
3a
8
7
6 / 4a
5
D
9b
'Anaerocellum thermophilum' DSM 6725
Clostridium thermocellum ATCC 27405
Thermoanaerobacter pseudethanolicus ATCC 33223
Thermoanaerobacter tengcongensis MB4
Thermotoga maritima MSB8

Figure 5.2 A) The central glycolytic pathway of *Caldicellulosiruptor saccharolyticus*. Consisting of the enzymes: (**1**) glucokinase; (**2**) glucose-6-phosphate isomerase; (**3a**) ATP-dependent 6-phosphofructokinase; (**3b**) PP_i-dependent 6-phosphofructokinase; (**4**) fructose-bisphosphate aldolase; (**4a**) triose-phosphate isomerase; (**5**) glyceraldehyde-3-phosphate dehydrogenase (phosphorylating); (**6**) phosphoglycerate kinase; (**7**) phosphoglycerate mutase; (**8**) phosphopyruvate hydratase (enolase); (**9a**) pyruvate kinase; (**9b**) pyruvate, phosphate dikinase. **B)** Gene ortholog neighborhoods of pyruvate kinase (**9a**, locus tag: Csac_1830) and glyceraldehyde 3-phosphate dehydrogenase (**5**, locus tag: Csac_1953) from *Caldicellulosiruptor saccharolyticus* and several functionally related species. Numbers designate the genes corresponding to the enzyme depicted in **A**. **D**, Transcriptional regulator of DeoR-family. Signs (**+/-**) indicate the gene cluster orientation on the genomes.

3.3 Analysis of ATP, ADP and PP_i

The presence of PP_i-dependent enzymes in the central metabolic pathway suggested the involvement of PP_i as an energy carrier in the metabolism of *C. saccharolyticus*. Therefore, the levels of both ATP and PP_i were determined during growth (**Figure 5.3**). PP_i levels were relatively high (approx. 4±2 mM), while ATP levels were remarkably low (0.43±0.07 mM). The levels of PP_i and ATP are in the same range as observed in three different methanotrophic bacteria, i.e. 5 mM PP_i and 0.5 mM ATP [225]. Since we were unable to determine the leakage of PP_i during extraction (see material and methods), it is possible that the PP_i levels are even higher.

In the transition to the stationary phase the ATP levels increased 2-fold and the PP_i levels decreased 6.5 fold, hence the PP_i/ATP ratio dropped 13-fold in the transition to the stationary phase. A similar PP_i dynamic has been reported for *Moorella thermoacetica* where the PP_i levels also peaked during the exponential phase, albeit with a substantially lower concentration (1.44 mM; [226]). Heinonen and Drake also observed that in *M. thermoacetica* the high growth-phase dependent cytosolic PP_i levels corresponded to a low cytosolic PPase activity, while in *E.*coli PP_i levels were low and static (0.3 mM) corresponding with a high PPase activity. Interestingly, *M. thermoacetica* contains two V-type H^+ translocating pyrophosphatases and no cytosolic PPases (IMG database; not shown), which might be the basis for the similarity in PP_i dynamics with respect to *C. saccharolyticus*. The decrease of the PP_i concentration in the transition to the stationary phase could be due to a declining anabolism, since PP_i is a product of various biosynthetic pathways. The increase in ATP levels upon the transition to the stationary phase is consistent with the notion that the turnover of ATP-demanding biosynthetic pathways or sugar uptake systems decreases when growth rate declines.

Interestingly, in all samples the ADP concentrations were higher than the ATP concentrations (**Figure 5.3**), which is generally considered to be a sign of starvation. Nevertheless *C. saccharolyticus* is growing exponentially suggesting that in addition to ATP other energy carriers, like PP_i, may contribute to the energy charge.

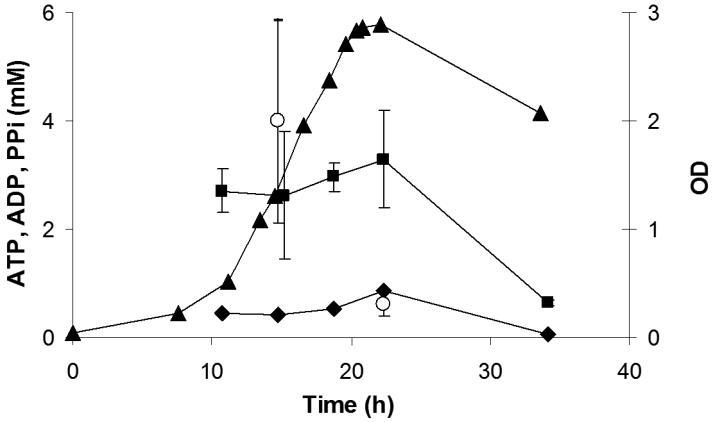


Figure 5.3 Growth profile (OD; ▲) and the concentrations of ATP (♦), ADP (■), and PP_i (o) during controlled batch growth. Bars indicate the standard deviation of three batch cultures.

3.4 Role of PP_i-PFK and PPDK in C. saccharolyticus

An overview of the glycolytic pathway of *C. saccharolyticus* (**Figure 5.2A**), reveals the absence of the standard gluconeogenic enzymes: fructose bisphosphatase (EC 3.1.3.11) and pyruvate water dikinase (EC 2.7.9.1). Their absence might suggest that the PP_i-PFK and the PPDK, which are both reversible enzymes, have an anabolic rather than a catabolic role. However, growth on carbon-three substrates like pyruvate and glycerol have so far not been reported for *C. saccharolyticus*. Moreover, microarray analysis have shown that during growth on glucose and xylose compared to growth on rhamnose PPDK is more than 7 times upregulated, together with the other C3-branch EM pathway enzymes, suggesting that the PPDK fulfills a catabolic role [40]. Interestingly, the gene coding for PPDK clusters together with all enzymes of the C3 branch with the exception of PK which is located elsewhere on the genome. The PPDK cluster also includes a DeoR-type transcriptional regulator. An overview of the neighborhood of the genes of the C3 branch of *C. saccharolyticus* and other functionally related bacteria is presented in **Figure 5.2B**. Although all investigated bacteria possess a PPDK, only '*Anaerocellum thermophilum*' (recently reclassified as *Caldicellulosiruptor bescii* [154]) reveals the same gene arrangement as *C. saccharolyticus*. For '*A. thermophilum*' an additional ORF, coding for a hypothetical protein, can be found overlapping both the PPDK and DeoR ORF. Whether the PPDK gene clusters of *C. saccharolyticus* and '*A. thermophilum*' are transcribed as a single polycistronic mRNA remains to be investigated. In contrast, PPDK from *Thermotoga maritima* clusters with the glycolytic enzyme fructose-bisphosphate aldolase, an acetate kinase and a GntR-type transcription regulator (data not shown). Furthermore, except for C*lostridium thermocellum* which lacks a PK, all investigated

organisms revealed the PK gene to be clustered with the gene coding for the ATP-PFK, suggesting coregulation [227]. In *Lactococcus lactis* the ATP-PFK and PK operon additionally contains the gene coding for lactate dehydrogenase and is known as the *las* (lactic acid synthesis) operon [228].

If PPDK acts in the catabolic direction, *C. saccharolyticus* has two options for converting PEP to pyruvate and ATP. It is therefore plausible that some type of regulation might occur. Therefore, the influence of PP_i levels on PK activity in *C. saccharolyticus* was investigated. PP_i was found to inhibit PK activity in *C. saccharolyticus* with an apparent Ki value of 2.9±0.9 mM PP_i (**Figure 5.4**). Consequently, when the PP_i levels are high during exponential growth (approx. 4 mM; **Figure 5.3**), the PK is inhibited by ~60%, again suggesting a catabolic role for PPDK in this growth phase. Consistently, in *Trypanosoma cruzi,* where PPDK is also working in the direction of ATP generation, and PP_i is also a strong inhibitor of PK [229]. Furthermore, PPDK has been shown to be used in the direction of ATP synthesis in some other organisms [230, 231]. The role of PP_i as allosteric effector has recently also been described for the lactate dehydrogenase (LDH) of *C. saccharolyticus* [51]. PP_i acts as inhibitor of the LDH, while ATP stimulates the enzyme. The estimated kinetics of the LDH explains the switch from a metabolism producing mainly acetate to a metabolism producing less acetate and more lactate.

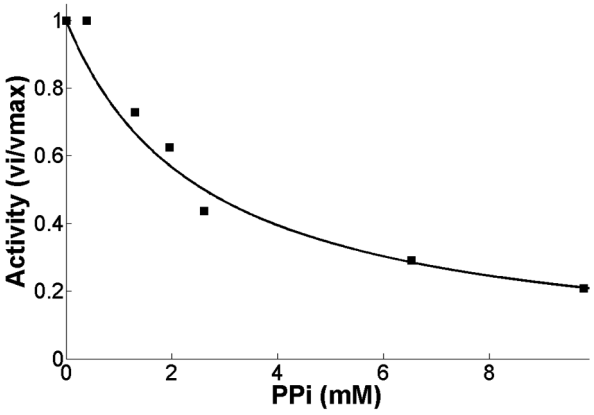


Figure 5.4 PP_i inhibition kinetics of pyruvate kinase in CE. The line represents the linear competitive inhibition model (with Ki= 2.9±0.9 as estimated with the least square method.

3.5 Energy conservation by using pyrophosphate

The hydrolysis of PP_i is generally regarded as an indispensable reaction of a cell's metabolism. PP_i is a by-product of various energy-requiring biosynthetic reactions, e.g. DNA and RNA synthesis and during the formation of precursors for protein and polysaccharide synthesis [68]. These reactions are often near equilibrium and only the effective removal of PP_i drives these reactions forward. Therefore, the coupling of these reactions to PP_i hydrolysis is crucial to maintain growth [232]. It is unknown what levels of PP_i still allow the cellular metabolism to proceed, but apparently *C. saccharolyticus* is capable of maintaining relatively high levels of PP_i during exponential growth without hampering biosynthesis. When PP_i is hydrolyzed to inorganic phosphate (P_i) by a cytosolic pyrophosphatase, the free-energy change is not preserved, but dissipated as heat. However, the involvement of a membrane-bound H^+-translocating pyrophosphatase makes it possible to use the energy released upon PP_i hydrolysis by establishing a proton motive force [224, 233]. The energy in PP_i also drives the PP_i-dependent reactions in the glycolytic pathway of *C. saccharolyticus*. This has also been shown for some other organisms, which conserve the free energy by using a PP_i-PFK [234-236] or a PPDK [230, 231, 237] in their central carbon catabolism. Mertens [181] argued that a PP_i-dependent glycolysis could be a way for fermentative bacteria to cope with a lower ATP yield.

Overall, the presented results indicate that pyrophosphate has a central role in the metabolism of the hydrogen-producing *C. saccharolyticus*. This type of metabolism agrees well with the observed physiology with respect to its sugar utilization [41]. The wide range of high-affinity sugar uptake systems and the absence of carbon catabolite repression suggests that *C. saccharolyticus* is not fastidious, but rather has evolved to conserve energy in many different ways.

Chapter 6

The involvement of pyrophosphate in glycolysis and gluconeogenesis: Remarkable singularities or wide spread phenomena?

Abstract

The involvement of pyrophosphate in glycolysis and gluconeogenesis has been reported for several bacterial, archaeal and eukaryotic organisms. However, an overview addressing the variations in pyrophosphate-dependent reactions and the genomic distribution of the related enzyme subtypes over the three domains of life is not available. In this study the enzyme diversity and genomic distribution of the glycolytic/gluconeogenic enzymes: phosphofructokinase, fructose bisphosphatase, pyruvate kinase, pyruvate water dikinase, pyruvate phosphate dikinase, as well as the soluble- and membrane-integral pyrophosphatases were investigated by using the fully sequenced genomes of 495 organisms. For this purpose the related chemical reactions were associated to genes using COG identifiers. Subsequent phylogenetic and hierarchical clustering analysis of the corresponding protein sequences allowed the classification of the sequences to specific enzyme subtypes. Based on the genomic distribution of the identified enzyme subtypes the correlations between the presence/absence of these enzymes subtypes was investigated. The investigation of the enzyme diversity allowed the identification of several novel enzyme-subtypes alongside the previously known subtypes, including a novel pyrophosphate-dependent phosphofructokinase, a membrane-integral-pyrophosphatase subgroup and several PK subtypes with non-canonical domain architectures. Although some organisms possess only a single alternative for a specific glycolytic/gluconeogenic step, it is not uncommon that multiple genes, coding for alternative enzymes catalysing the same step, are present in a single genome. Twenty-six percent of the investigated genomes contain the classical glycolysis-associated enzymes while lacking the non-classical enzymes pyrophosphate-dependent phosphofructokinase (PP_i-PFK) and pyruvate phosphate dikinase (PPDK). Twenty-eight percent of the genomes possess the classical gluconeogenic enzymes and lack both PP_i-PFK and PPDK encoding genes. Alternatively, 52% of all investigated genomes additionally contain the non-classical enzymes PP_i-PFK and/or PPDK. For genomes solely containing the classical glycolytic/gluconeogenic enzyme subtypes, a low correlation with respect to the presence of a membrane-integral pyrophosphatase is observed. On the other hand, 85% of the investigated genomes that contain a membrane-integral pyrophosphatase, additionally contain a PP_i-PFK and/or PPDK, which represents 80% of the organisms containing a PP_i-PFK and 78% containing a PPDK gene. Furthermore, the physiological relevance of pyrophosphate energy metabolism is discussed. Overall, the presented data indicate that the involvement of pyrophosphate in glycolysis/gluconeogenesis is a wide spread phenomenon.

1. Introduction

The Embden-Meyerhof-Parnas (EMP) pathway is the major metabolic route involved in the conversion of glucose to pyruvate, whereas the widely distributed gluconeogenic pathway covers the reverse reactions, i.e. the conversion of pyruvate up to glucose 6-phosphate. Although both pathways are highly conserved with respect to their intermediates, considerable diversity occurs in terms of catalytic conversions. Enzymes that catalyse a specific glycolytic step may require different co-substrates [238]. Moreover, enzymes catalysing the same chemical reaction can often be divided into different (sub)types based on their biochemical properties e.g. regulatory or structural characteristics [239, 240].

The interconversion of the reaction pairs fructose 6-phosphate (F6P) / fructose 1,6-bisphosphate (F1,6bP) and phosphoenolpyruvate (PEP) / pyruvate (PYR) display the highest diversity with respect to the associated phosphoryl-donors (**Figure 6.1A**). The catabolic conversion of F6P to F1,6bP is typically catalysed by the ATP- or ADP-dependent phosphofructokinase (PFK) [238, 240], while the fructose bisphosphatase (FBP) is obviously linked to the anabolic reaction [241]. However, the conversion of F6P to F1,6bP catalysed by the pyrophosphate-dependent PFK (PP_i-PFK) is proposed to be reversible under physiological conditions [181] and may either play a catabolic or anabolic role [240]. Pyruvate kinase (PK) catalyses the catabolic conversion of PEP to PYR [239], while pyruvate water dikinase (PWDK, PEP synthase) is commonly associated with anabolism [231], although, in some cases PWDK has been described to have an essential role in glycolysis [242]. Likewise, both catabolic as anabolic functions have been reported for pyruvate phosphate dikinases (PPDK) [231]. ATP-PFK and PK-catalysed reactions are virtually irreversible and are generally prone to allosteric regulation, thus providing strong control points for the fluxes through the EMP pathway and gluconeogenesis [239, 240]. The involvement in glycolysis of either PP_i-PFK as PPDK, which are less prone to allosteric regulation [240], might result in the loss of a proper control point in this catabolic route.

When operating in the glycolytic direction both PP_i-PFK and PPDK reactions consume pyrophosphate (PP_i). PP_i is a by-product of many biosynthesis reactions, including the synthesis of proteins, polysaccharides and polynucleotides (DNA and RNA) [68]. Accumulation of PP_i has an inhibitory effect on cell growth and its effective removal is essential for the biosynthetic reactions to proceed [69]. While the hydrolysis of PP_i by soluble inorganic pyrophosphates (sPPases) releases the associated free energy as heat, membrane-integral pyrophosphates (M-PPases) can couple the free energy release to ion translocation, thereby creating an electrochemical gradient (**Figure 6.1B**) [243]. Alternatively, PP_i can be removed through PP_i-mediated glycolytic steps, which in some cases even have been shown to effectively replace pyrophosphatase activity [229].

Although the involvement of PP_i as a phosphoryl-donor in glycolysis has been reported for organisms belonging to all the three domains of life [231, 234, 237, 244-246], it is uncertain whether such PP_i-mediated types of glycolysis should be considered as organism specific rarities or should be seen as a widespread phenomenon i.e. representing subtypes of the classical EMP-pathway. This work investigates the diversity of the PP_i-dependent glycolytic enzymes, their distribution over the three domains of life and their correlation with the other enzymes associated with the glycolytic/gluconeogenic interconversions of F6P/F1,6-bisP and PEP/PYR and the PPase-catalysed reactions (**Figure 6.1**).

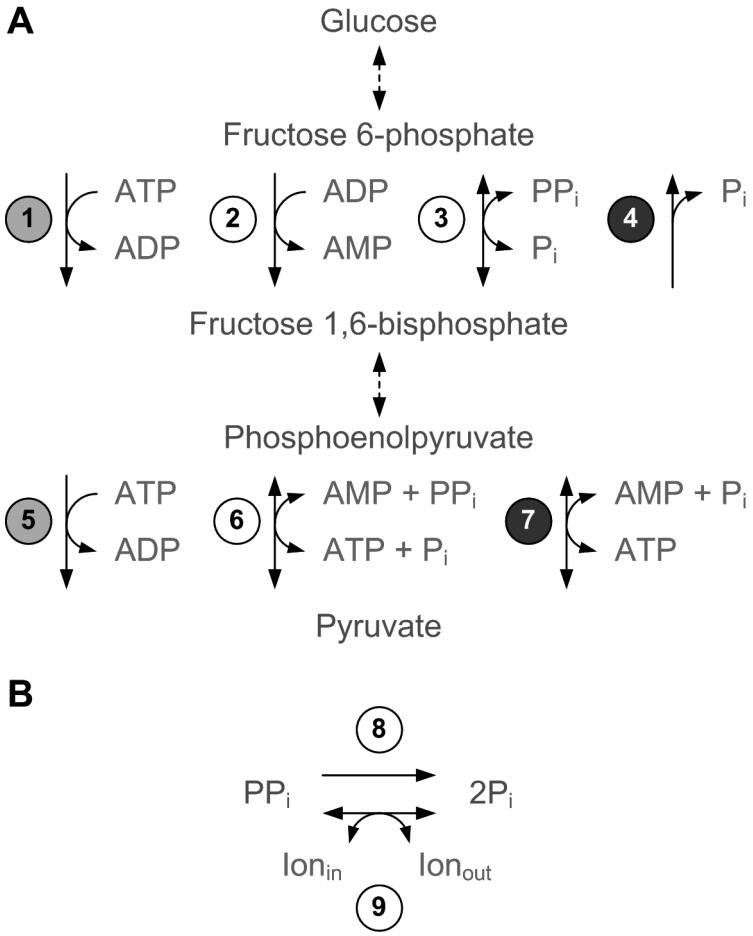


Figure 6.1 A) Chemical conversions at the F6P/F1,6bP and PEP/PYR interconversion nodes of the EMP pathway and gluconeogenesis. Reaction associated enzymes: (**1**) ATP-PFK EC 2.7.1.11; (**2**) ADP-PFK EC 2.7.1.146; (**3**) PP_i-PFK EC 2.7.1.90; (**4**) FBP EC 3.1.3.11; (**5**) PK EC 2.7.1.40; (**6**) PPDK EC 2.7.9.1; (**7**) PWDK EC 2.7.9.2. Classical EMP-pathway reactions are indicated in green, classical gluconeogenic reactions are indicated in purple. **B)** Hydrolysis of inorganic pyrophosphate. Reaction associated enzymes: (**8**) soluble PPase EC 3.6.1.1; (**9**) Membrane bound PPase EC 3.6.1.1.

For this purpose, the publically available genomes of 70 archaea, 30 eukarya and 395 bacteria were analysed for the presence of genes related to the indicated reactions. COG identifiers [247, 248]were used to link the genes, via their corresponding protein sequences, to the chemical reactions. The identified gene products were assigned to specific enzyme subtypes, based on their protein sequences, information from public databases and available literature on relevant features such as protein signatures or specific catalytic/allosteric site residues. Results of the subtype classification were used to investigate correlations between the presence/absence of a specific enzyme subtype with respect to another enzyme subtype, e.g. the correlation between the genomic presence of PPDK and PP_i-PFK encoding genes.

Our findings reveal that, although some organisms possess only a single alternative for a specific glycolytic/gluconeogenic step, it is not uncommon that multiple genes coding for alternative enzymes catalysing the same step are present in a single genome. In addition to the known enzyme subtypes, our approach allowed us to identify novel enzyme subtypes which, to the best of our knowledge, have not yet been the subject of any characterization studies. Based on the genomic content of the investigated organisms, our observations support the statement that the involvement of PP_i in glycolysis/gluconeogenesis is a wide spread phenomenon, and that from an enzymatic point of view there is no standard glycolysis.

2. Methods

2.1. Selection of chemical interconversion reactions

A schematic overview of the data flow used during this study is given in **Figure 6.2**. The sets of interconversion reactions, associated with the reaction pairs: F6P/F1,6bP, PEP/PYR and PP_i/P_i, were identified using the KEGG database (http://www.genome.jp/kegg/, [249]) (**Table 6.1**). In addition, Enzyme Class numbers (ECs), KEGG Orthology numbers (KOs), related Cluster of Orthogolous Groups of proteins identifiers (COGs) and Gene Onthology identifiers (GOs) related to these interconversion reactions, were obtained via the KEGG database. The interconversion reactions indicated in **Table 6.1** were used for further analysis.

2.2. Genome selection

Only completely sequenced genomes were analysed during this study. Based on the finished genomes present in the IMG database (November 2012, (http://img.jgi.doe.gov/cgi-bin/w/main.cgi), [250]), a subset was composed containing bacterial, archaeal and eukaryotic genomes. This subset included a maximum of one genome per genus (random selection at the genus level).

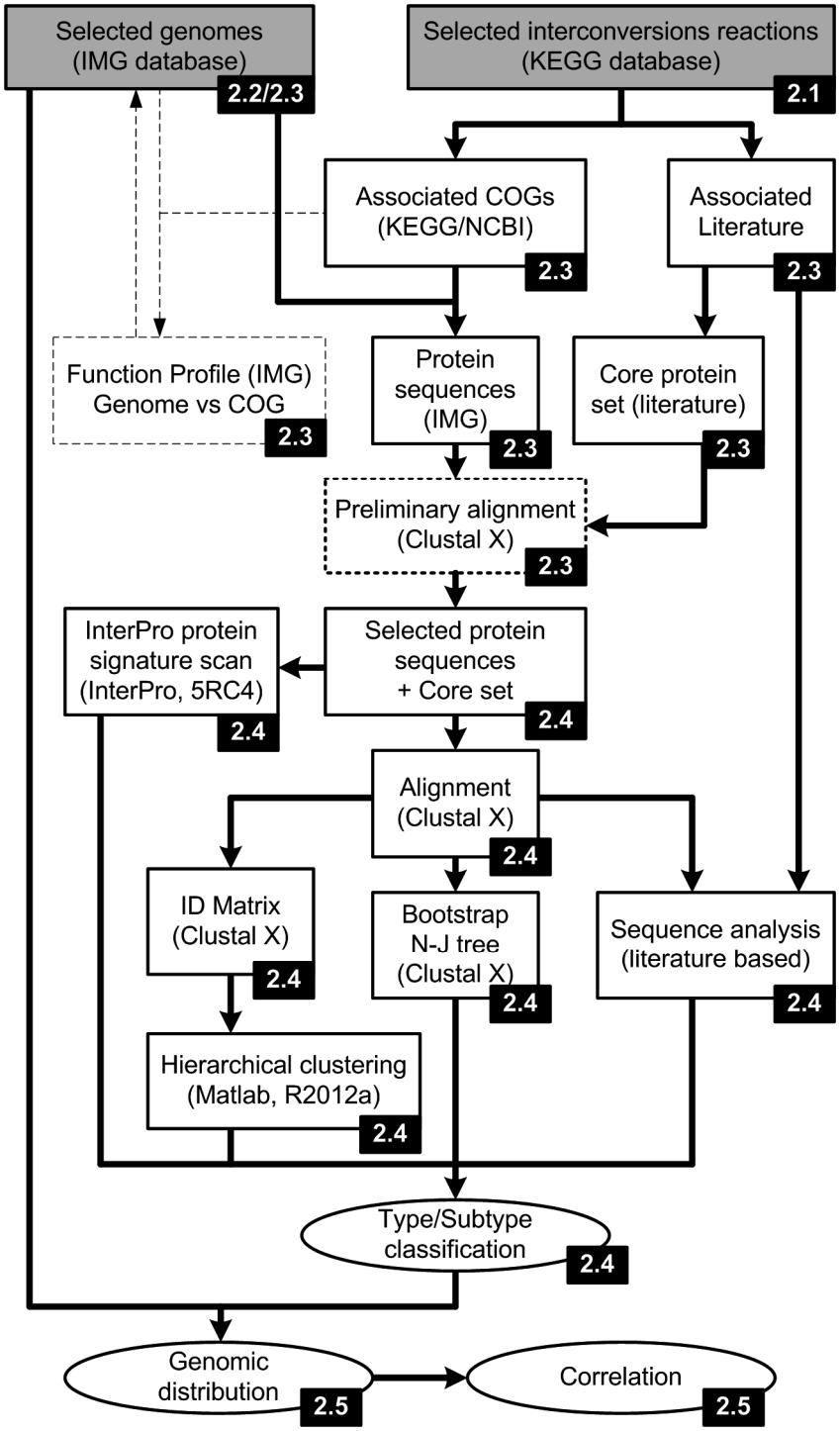

Selected genomes (IMG database)
2.2/2.3
Selected interconversions reactions (KEGG database)
2.1
Associated COGs (KEGG/NCBI)
2.3
Associated Literature
2.3
Function Profile (IMG) Genome vs COG
2.3
Protein sequences (IMG)
2.3
Core protein set (literature)
2.3
Preliminary alignment (Clustal X)
2.3
InterPro protein signature scan (InterPro, 5RC4)
2.4
Selected protein sequences + Core set
2.4
Alignment (Clustal X)
2.4
ID Matrix (Clustal X)
2.4
Bootstrap N-J tree (Clustal X)
2.4
Sequence analysis (literature based)
2.4
Hierarchical clustering (Matlab, R2012a)
2.4
Type/Subtype classification
2.4
Genomic distribution
2.5
Correlation
2.5

Figure 6.2 Schematic overview of the data flow, software and public databases used during this study. Numbers represent the sections where indicated proceedings are discussed.

In addition, the bacterial subset was further constrained at the family level. When multiple organisms from the same family displayed a similar COG function profile (for details see **section 2.3**) only one of those organisms was incorporated into the final genome set (random selection at the family level). A list of the 495 organisms selected for analysis, including their IMG genome and NCBI taxon IDs, can be found in **Supplementary Table S6.A**.

2.3. COG function profiles and protein sequence selection

COG identifiers [247, 248] were used to link the indicated chemical reactions (**Table 6.1**), via the gene associated protein sequences, to the genes. Selected genomes where scanned for the presence of the COG related genes, using the Integrated Microbial Genomes (IMG) platform [250]. The generated COG function profiles revealed the present or absence of COG related genes for each analysed genome. These COG function profiles were used to further limit the subset of bacterial genomes as described in **section 2.2**. For each of the 11 investigated COGs all the associated protein sequences, identified in the selected genomes, were retrieved via the IMG database, thus generating 11 protein sequence sets. For each COG related protein sequence set a core sequence set was designed, consisting of the sequences of functionally characterized proteins and/or proteins whose structure has been determined (**Supplementary Table S6.B**). Each core set was combined with the COG related protein sequence. A preliminary alignment allowed the manual trimming of each COG associated protein sequence subset to exclude truncated (incomplete) sequences and false positives. The resulting protein sequence sets were used for further analysis. In addition, based on the manual investigation of the highly conserved functional relevant regions, e.g. the catalytic site, sequences leading to potentially non-functional enzymes were identified.

2.4. Protein subtypes identification and classification

Investigated protein sequences were classified based on information obtained from 1) sequence alignments, 2) phylogenetic analysis, 3) hierarchical clustering analysis, 4) InterPro signatures scans and 5) available literature. All protein sequence alignments were performed using Clustal X 2.1 (default settings, [215]). The resulting alignments were used to manually investigate the presence of functionally relevant residues determining substrate specificity, e.g. the amino acids at the catalytic site, or other functional characteristics. For each COG related protein subset a Bootstrap N-J Tree

was generated (Clustal X 2.1, random number generator seed 111, number of bootstrap trails 1000). Phylogenetic trees were used to identify monophyletic clusters within the COG associated protein sequence subsets. Hierarchical clustering analyses were performed in Matlab (software package R2012a), using the built-in function: linkage, with the unweighted average distance (UPGMA) as linkage criterion. Identity matrices, derived from the aligned protein sequences using Clustal X 2.1, were used as input for the hierarchical clustering (HC). Cluster threshold values were determined manually for each COG data set to achieve a relevant resolving resolution. Identified HC clusters where arbitrarily numbered. All protein sequences were re-analysed using InterProScan to acquire the update InterPro based functional classification of the investigated proteins (InterProScan 5RC4, using default thresholds [251]). Identified InterPro IDs which had a limited number of hits, below 2% of the examined protein subsets, were neglected. Identified protein subsets-linked InterPro IDs and their associated signature IDs are given in **Supplementary Table S6.C**. Based on the acquired results each investigated protein was manually classified into a specific protein subtype. In addition, all investigated proteins were classified with respect to the chemical reaction they catalysed. Classification and associated nomenclature elaborated on literature findings.

2.5 Genomic distributions and correlations

The genomic distribution of the genes associated with the different chemical reactions and the classified protein subtypes (**section 2.4**) was investigated based on the gene presence/absence profiles. Correlation studies were performed with respect to 1) identified protein subtypes within a protein family (COG) (e.g. correlation between the different subtypes associated to COG0205 (ATP- and PP_i-PFK subtypes)), 2) identified protein subtypes associated with a specific chemical reaction at a specific interconversion node (e.g. correlation between PP_i-PFK and FBP) and 3) between specific chemical reaction at the three interconversion nodes e.g. correlation between PPDK and PP_i-PFK.

Table 6.1 Overview of interconversion reactions, enzyme classes and COGs associated with reaction pairs: F6P/F1,6bP,PEP/PYR and PP_i/P_i. Reactions indicated with an * were **NOT** investigated in this study. Information was acquired via the KEGG database (http://www.genome.jp/kegg/, **Kanehisa 2000**). ID indicates KEGG identifiers: RPAIR or COMPOUND; Reaction ID indicates KEGG identifier: REACTION; EC: Enzyme Class; KO: KEGG ORTHOLOGY; COG: Cluster of Orthologous Group; GO: gene ontology.

Interconversion fructose 6-phosphate/fructose 1,6-bisphosphate

	ID	KEGG name	Common name	Abbreviation
reaction pair	RP00191	C05345_C05378		
compound	C05345	beta-D-Fructose 6-phosphate	fructose 6-phosphate	F6P
compound	C05378	beta-D-Fructose 1,6-bisphosphate	fructose 1,6-bisphosphate	F1,6bisP

RP00191 associated reactions

reaction ID	EC	Systematic name (IUBMB Enzyme Nomenclature)	Common name	Abbreviation
R04779	2.7.1.11	ATP:D-fructose-6-phosphate 1-phosphotransferase	ATP-dependent phosphofructokinase	ATP-PFK
R09084	2.7.1.146	ADP:D-fructose-6-phosphate 1-phosphotransferase	ADP-dependent phosphofructokinase	ADP-PFK
R02073	2.7.1.90	diphosphate:D-fructose-6-phosphate 1-phosphotransferase	PP_i-dependent phosphofructokinase	PP_i-PFK
R04780	3.1.3.11	D-fructose-1,6-bisphosphate 1-phosphohydrolase	fructose-bisphosphatase	FBP

reaction ID	KO	description	COG	GO
R04779	K00850	PFK, pfk; 6-phosphofructokinase [EC:2.7.1.11]	COG0205	GO: 0003872
	K16370***	pfkB; 6-phosphofructokinase 2 [EC:2.7.1.11]	COG1105	GO: 0003872
R09084	K00918	ADP-dependent phosphofructokinase/glucokinase [EC:2.7.1.146/ 2.7.1.147]	COG4809	GO: 0043844 & GO: 0043843
R02073	K00895	E2.7.1.90, pfk; pyrophosphate--fructose-6-phosphate 1-phosphotransferase [EC:2.7.1.90]	COG0205	GO: 0047334
R04780	K03841	FBP, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11]	COG0158	GO: 0042132
	K02446	glpX; fructose-1,6-bisphosphatase II [EC:3.1.3.11]	COG1494	GO: 0042132
	K11532	glpX-SEBP; fructose-1,6-bisphosphatase II / sedoheptulose-1,7-bisphosphatase [EC:3.1.3.11/ 3.1.3.37]	COG1494	GO: 0042132 & GO: 0050278
	K04041	fbp3; fructose-1,6-bisphosphatase III [EC:3.1.3.11]	COG3855	GO: 0042132
	K01622	K01622; fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13/ 3.1.3.11]	COG1980	GO: 0042132 & GO: 0004332

Table 6.1 continued

Interconversion phosphoenolpyruvate/pyruvate

	ID	KEGG name	Common name	Abbreviation
reaction pair	RP00036	C00022_C00074		
compound	C00074	phosphoenolpyruvate	phosphoenolpyruvate	PEP
compound	C00022	pyruvate	pyruvate	PYR

RP00036 associated reactions

reaction ID	EC	Systematic name (IUBMB Enzyme Nomenclature)	Common name	Abbreviation
R00200	2.7.1.40	ATP:pyruvate 2-O-phosphotransferase	pyruvate kinase	PK
R02628*	2.7.3.9	phosphoenolpyruvate:protein-L-histidine Nπ-phosphotransferase	phosphotransferase system, enzyme I	PTS, enzyme I
R00208*	3.1.3.60	phosphoenolpyruvate phosphohydrolase	phosphoenolpyruvate phosphatase	PEPase
R00206	2.7.9.1	ATP:pyruvate, phosphate phosphotransferase	pyruvate, phosphate dikinase	PPDK
R00199	2.7.9.2	ATP:pyruvate, water phosphotransferase	pyruvate, water dikinase (PEP synthetase)	PWDK (PPS)
R01012*	2.7.1.121	Phosphoenolpyruvate:glycerone phosphotransferase	-	-

reaction ID	KO	description	COG	GO
R00200	K00873	PK, pyk; pyruvate kinase [EC:2.7.1.40]	COG0469	GO: 0004743
R00206	K01006	ppdK; pyruvate,orthophosphate dikinase [EC:2.7.9.1]	COG0574	GO: 0050242
R00199	K01007	pps, ppsA; pyruvate, water dikinase [EC:2.7.9.2]	COG0574	GO: 0008986

Table 6.1 continued

Interconversion pyrophosphate/phosphate

	ID	KEGG name	Common name	Abbreviation
reaction pair	RP00190	C00009_C00013		
compound	C00013	Diphosphate	(inorganic) pyrophosphate	PP_i
compound	C00009	Orthophosphate	(inorganic) phosphate	P_i

RP00190 associated reactions**

reaction ID	EC	Systematic name (IUBMB Enzyme Nomenclature)	Common name	Abbreviation
R00004	3.6.1.1	diphosphate phosphohydrolase	pyrophosphatase	PPase
R00320*	2.7.2.12	pyrophosphate-acetate phosphotransferase	diphosphate-acetate kinase	-
R00346*	4.1.1.38	diphosphate:oxaloacetate carboxy-lyase	diphosphate-phosphoenolpyruvate carboxykinase	-
R00584*	2.7.1.80	diphosphate:L-serine O-phosphotransferase	diphosphate-serine phosphotransferase	-
R01044*	2.7.1.79	diphosphate:glycerol 1-phosphotransferase	diphosphate-glycerol phosphotransferase	-
R03699*	2.7.1.143	diphosphate:purine nucleoside phosphotransferase	diphosphate-purine nucleoside kinase	-

reaction ID	KO	description	COG	GO
R00004	K01507	E3.6.1.1, ppa; inorganic pyrophosphatase [EC:3.6.1.1]	COG0221	GO: 0004427
	K06019***	E3.6.1.1X; pyrophosphatase PPaX [EC:3.6.1.1]	COG0546	GO: 0004427
	K11725***	LHPP; phospholysine phosphohistidine inorganic pyrophosphate phosphatase [EC:3.6.1.1/ 3.1.3.-]	-	GO: 0004427
	K11726***	NURF38; nucleosome-remodeling factor 38 kDa subunit [EC:3.6.1.1]	-	GO: 0004427
	K15986	PPaC: manganese-dependent inorganic pyrophosphatase [EC:3.6.1.1]	COG1227	GO: 0004427
	K15987	hppA; K(+)-stimulated pyrophosphate-energized sodium pump [EC:3.6.1.1]	COG3808	GO: 0004427

*Indicated reactions were not analysed in this study.
** Also included R00206 and R02073, which are given at the interconversion phosphoenolpyruvate/pyruvate and fructose 6-phosphate/fructose 1,6-bisphosphate sections, respectively.
***Indicated KO related gene products were not analysed in this study.

3. Results

This work set out to investigate the diversity of enzymatic steps for which PP_i-dependent alternatives exist and whether the presence of such PP_i-dependent steps are somehow correlated. For this endeavour the enzyme diversity at the interconversion steps of F6P/F1,6bP, PEP/PYR and PP_i/P_i (**Table 6.1**) of 495 fully sequenced genomes was investigated. Protein sequences associated to the interconversion reactions were identified using COG identifiers, and identified sequences were subsequently classified into specific enzyme subgroups. This COG based approach allowed the identification of all previously reported enzyme subtypes and in some cases led to the identification of novel subtypes hitherto not mentioned in literature. In addition, the taxonomic distribution of these enzyme subgroups was examined. Some protein sequences could not be assigned to any specific enzyme subtypes (indicated as U for unclassified). Since these sequences could usually be classified with respect to their function, they were included in the genomic distribution studies regarding the enzyme function, however, they were not included in the distribution and correlation studies of the enzyme subtypes. The specific approaches and results for each investigated interconversion reaction are discussed in the following sections. Results from all alignments performed during this study are available upon request and a complete overview of all classified enzyme subgroups and their taxonomic distribution is given in **Supplementary Tables S6.A & S6.D**.

3.1 Interconversion of fructose 6-phosphate and fructose 1,6-bisphosphate

The glycolytic conversion of F6P to F1,6bP can be catalysed by phosphofructokinases. These enzymes use either ATP, PP_i or ADP as phosphoryl-donor. Based on their sequences PFKs have been classified into three different families (PFK family A, B and C) [240]. Family A PFKs (PFK1), which are either ATP- or PP_i-dependent, have been subdivided into eight PFK subgroups (group LONG, SHORT, B1, E, III, P, B2 and X) [220]. Family B PFKs (PFK2) are considered as non-glycolytic PFKs and have been shown to phosphorylate a large variety of substrates [240]. For this reason Family B PFKs were not incorporated in this study. ADP-dependent PFKs identified in archaea [252, 253], representing the family C PFKs, have evolved from ADP-dependent glucokinases and belong to the ribokinase super family [254].

FBPases, which are typically associated with a gluconeogenic function, catalyse the dephosphorylation of F1,6bP to F6P. Five different classes of FBPase have been reported (Class I-V) [241, 255-257]. Of these FBPases, Class IV FBPases, which structurally resemble Class I FBPases [258] and are classified as inositol monophosphatases, were initially proposed to represent the archaeal FBPase [259]. However, a knockout of the Class IV FBPase in *Thermococcus kodakarensis* did not lead

to phenotypic changes, while the knockout of the Fbp_{Tk} gene, disrupted growth under gluconeogenic conditions. Thus Fbp_{Tk} orthologs (class V FBPases) were identified as the true gluconeogenic FPBase for archaeal hyperthermophiles [241]. For this reason Class IV FBPases were not included in this survey.

3.1.1 ATP- and PPi-dependent phosphofructokinases

The investigation of the ATP-PFK crystal structures from *Geobacillus stearothermophilus* (group B1), *Escherichia coli* (group B1), *Trypanosoma brucei* (group X) and *Saccharomyces cerevisiae* (group E) has allowed for the identification of the family A PFK catalytic site, consisting of the ATP- and F6P-binding pocket [260-263]. Previous analysis of family A ATP- and PP_i-PFK sequences revealed similarity with respect to their active sites, though, conserved differences within active site motifs allow for the discrimination between the required phosphoryl-donor [220, 221, 240].

In *G. stearothermophilus* and *E. coli* ATP-PFKs (B1) the conserved ^{101}GGDGS^{105} motif (*G. stearothermophilus* ATP-PFK (B1) AA numbering (GenBank: BAA02405.1)) forms a part of the ATP-binding pocket [261, 263]. The glycine at position 104 of the ATP binding pocket motif ^{101}GGDGS^{105} interacts with ATP via its backbone amine group, leaving no room for a side-chain [263]. This glycine residue is conserved in all characterized ATP-PFKs but is replaced by an aspartic acid in all characterized PP_i-dependent PFKs [220]. A point mutation G104D in the *E. coli* ATP-PFK (B1) resulted in the loss of ATP-dependent PFK activity [221], confirming the suggestion that the presence of a Asp(D) residue at this position would prevent the binding of ATP by steric and electrostatic hindrance [222]. Moreover, a single point mutation D104G in the PP_i-dependent PFK of *E. histolytica* (group LONG) changed the phosphoryl-donor specificity of this type of PFK form PP_i to ATP, thus indicating the presences of a latent ATP binding site in group LONG PP_i-PFKs [221].

The conserved motifs ^{124}GTIDND^{129} and ^{169}MGR^{171} are parts of the F6P binding site of *G. stearothermophilus* and *E. coli* ATP-PFKs (B1) [261, 263]. For some PFK sequences the glycine at position 124 in the ^{124}GTIDND^{129} motif is replaced by a lysine(K). This lysine, together with Asp(D)127, is proposed to be involved in the double proton shuttle mechanism linked to PP_i-PFK activity [264]. It was initially suggested that the Lys(K)124 residue effectively prevented the binding of nucleotide triphosphate [264] and indeed the single mutation G124K in the *E. coli* ATP-PFK (B1) removed ATP-dependent PFK activity [221]. However, the corresponding lysine residue identified in Kinetoplastea PFKs (group X), which are strictly ATP-dependent, is required to maintain stability of the active site structure, thus indicating the importance of the identity of this residue for this type of PFKs [265, 266].

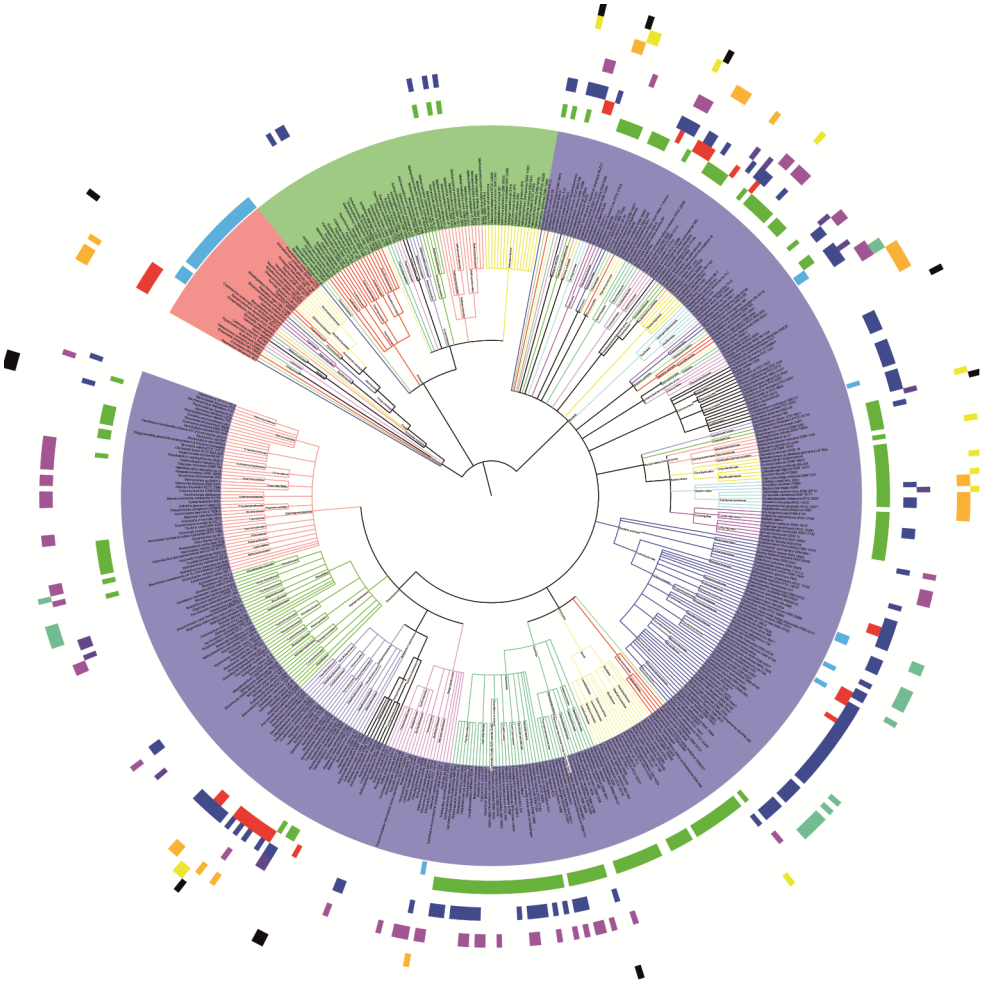


Figure 6.3 Taxonomic distribution of the family A PFK subgroups. Subgroups (from inner to outer circle), **light blue**, E; **green,** B1; **red**, X; **blue**, III; **purple**, Y; **pink**, B2; **turquoise**, P; **orange**, LONG; **yellow**, SHORT; **black**, PFKs not classified to a specific subgroup (U). Taxonomic tree labels (given clockwise). Leaf labels: light (l.) red, ***Eukaryotes***; l. green, ***Archaea***; l. blue, ***Bacteria***. Clade labels ***Eukaryotes*** (Class): **blue**, Aconoidasida; **l. green**, Ectocarpales (Order); **red**, Entamoeba (species); **l. yellow**, Prasinophyceae; **turquoise**, Brassicales (Order); **l. pink**, Kinetoplastida (Order); **black**, Mammalia; **l. blue**, Chromadorea; **green**, Insecta; **l. red**, Unikaryonidae (Family); **yellow**, Chytridiomycetes; **l. turquoise**, Tremellomycetes; **pink**, Agaricomycetes; **blue**, Schizosaccharomycetes; **l. green**, Eurotiomycetes; **red**, Sordariomycetes; **l. yellow**, Saccharomycetes. Clade labels ***Archaea*** (Class): **blue**, Nanoarchaea; **l. green**, Thaumarchaeota; Thermoprotei, **red**; Methanopyri, **l. yellow**; **turquoise**, Thermoplasmata; **l. pink**, Thermococci; **black**, Archaeoglobi; **l. blue**, Methanococci; **green**, Methanobacteria; **l. red**, Methanomicrobia; **yellow**, Halobacteria. Clade labels ***Bacteria*** (Class): **blue**, Dictyoglomia; **l. green**, Chrysiogenetes; **red**, Deferribacteres; **l. yellow**, Nitrospira; **turquoise**, Gemmatimonadetes; **l. pink**, Caldisericia; **black**, WWE1; **l. blue**, Thermobaculae; **green**, Thermodesulfobacteria; **l. red**, Elusimicrobia; **yellow**, Endomicrobia; **l. turquoise**, Synergistia; **pink**, Mollicutes; **blue**, Deinococci; **l. green**, Thermi; **l. yellow**, Kueneniae;

turquoise, Planctomycetia; **l. pink**, Fusobacteria; **black**, Fibrobacteres; **l. blue**, Solibacteres; **green**, Acidobacteriales; **l. red,** Chloracidobacteria; **yellow**, Thermotogae; **l. turquoise**, Aquificae; **pink**, Anaerolineae; **blue**, Thermomicrobia; **l. green**, Dehalococcoidetes; **red**, Chloroflexi; **l. yellow**, Verrucomicrobiaceae; **turquoise**, Opitutae; **l. pink**, Chlamydiia; **black**, Cyanobacteria (Phylum); **l. blue**, Ignavibacteria; **green**, Chlorobia; **l. red** ,Sphingobacteriia; **yellow**, Flavobacteriia; **l. turquoise**, Bacteroidia; **pink**, Cytophagia; **blue**, Actinobacteria; **l. green**, Erysipelotrichi; **red**, Negativicutes; **l. yellow**, Bacilli; **turquoise**, Clostridia; **l. pink**, Betaproteobacteria; **black**, Epsilonproteobacteria; **l. blue**, Deltaproteobacteria; **green**, Alphaproteobacteria; **l. red**, Gammaproteobacteria. Taxonomic distribution trees were prepared using the online tool: Itol (http://itol.embl.de/index.shtml, **Letunic 2011**).

Predictions of phosphoryl-donor specificity based on the identity of the amino acid residues at positions 104 and 124 matched PFK characterisation data, with PFKs containing the residue pair Gly(G)104 + Gly(G)124 or Gly(G)104 + Lys(K)124 being ATP-dependent, while the residue pair Asp(D)104 + Lys(K)124 gave PP_i-dependence [220]. Group B1 and group E PFKs, whose sequences all contain the residues pair Gly(G)104 + Gly(G)124, were identified as ATP-PFKs. All Group X sequences possess the Gly(G)104 + Lys(K)124 residue pair and characterized members are all ATP-dependent. Group III contains both ATP- and PP_i-PFKs with the respective residue reaction pairs Gly(G)104 + Lys(K)124 and Asp(D)104 + Lys(K)124. While groups LONG, SHORT, P, B2 PFKs, typically possess the residue pair Asp(D)104 + Lys(K)124, associated with PP_i-dependence [220].

All family A PFK sequences belong to COG0205. Identified sequences were analysed with respect to the presence of the catalytic site residues [261, 263], including the indicated motifs representing the ATP/PP_i and F6P binding pockets. Based on the alignment, phylogenetic analysis and hierarchical clustering 9 different subgroups could be identified, covering the 8 groups as defined by Bapteste *et al.* [220] and a previously unidentified PFK group, group Y (**Supplementary Table S6.D**). Specific Interpro IDs have previously been assigned for each subgroup with the exception of the novel group Y PFKs, for which no discriminating protein signature exists. Group Y sequences notably differ from all other PFK sequences with respect to the ^{169}MGR^{171} motif, which forms a part of the F6P binding site. This motif is highly conserved among all the other identified PFK subgroup members, while in all group Y sequences Met(M)169 is strictly replaced by a Phe(F)169. Moreover, only 55% of the sequences classified as Group III PFKs give a hit with the Group III specific IPR012829 ID related signature TIGR02483. For the other Group III sequences no specific signatures were identified, however, these sequences did not form separate hierarchical clusters (**Supplementary Table S6.D**).

Phosphoryl-donor specificity was assigned based on the identity of the residues at positions 104 and 124 (*G. stearothermophilus* ATP-PFK (B1) AA numbering) (**Supplementary Table S6.D**). As expected all group E, B1 and X PFK sequences could be classified as ATP-dependent. Most members of the groups LONG, SHORT, P, B2 and Y classified as PP_i-dependent PFKs, although for each of those groups several

members possessed the ATP-related residue pair Gly(G)104 + Lys(K)124. Group III contained a mixture of ATP- (75%) and PP_i-dependent (25%) PFKs.

Overall 70% of the 495 investigated organisms contained genes coding for a family A PFK. All eukaryotic species, most bacteria but only a few archaeal species have a family A PFK. The taxonomic distribution of the identified family A PFK subgroups is shown in **Figure 6.3** and **Supplementary Table S6.A**. The investigated eukaryotic genomes only contained group E, X, and LONG PFKs, while in archaea only group III and B1 PFKs were observed. All subgroups are represented in bacterial species. Although most organisms possessed only a single PFK subgroup, 25% of the investigated genomes contained two or more genes encoding PFKs belonging to different subgroups (**Table 6.2**). Overall 35 different family A PFK related genotypes (different PFK subgroup combinations) could be identified. Results showed that the genotypes of species from the same family can be conserved, e.g. all genomes of the bacterial family Cytophagaceae (5 genomes included in this study) only have a B1 PFK, however, there can also be remarkable differences, e.g. for the bacterial family Thermoanaerobacteriaceae (5 genomes included in this study) three different genotypes were observed (B1 & III, B1 & B2 and B1 & B2 & III, **Supplementary Table S6.A**). More than 50% of the organisms containing a LONG, SHORT, III, Y, P, B2 or X PFK subgroup have an additional gene coding for a different PFK group. The pairwise cross correlations between ATP-PFK groups E, B1 and X are low (**Table 6.2**), indicating that the simultaneous genomic presence of these ATP-PFK subgroups is a rarity. Likewise, the simultaneous genomic presence of PP_i-PFK subgroups LONG, SHORT, Y, P and B2 is hardly observed.

Of all investigated organisms only 40% contain an ATP-dependent PFK, while 18% have both an ATP- and PP_i-dependent PFK and 12% of the organism have a PP_i-PFK only. The taxonomic distribution of the ATP- and PP_i-dependent PFKs is shown in **Supplementary Figures S6.6** and **Supplementary Table S6.A**, and will be discussed with respect to the ADP-PFK and FBPase distributions in **section 3.1.7**.

Many studies have addressed the diversity of the allosteric control of family A PFKs. The ATP-PFKs from *G. stearothermophilus* and *E. coli* are allosterically activated by ADP and inhibited by PEP, with both regulators binding at the same site [261, 263, 267]. AMP, identified as an allosteric activator for the Group X PFK from *T. brucei*, is proposed to bind to the same ADP/PEP effector site [262]. All PFKs from Group E appear to be the product of a gene duplication followed by a fusion event of an ancestral PFK similar to the Group B1 PFKs. The subsequent changes in the active and regulatory sites accounted for the high degree of allosteric regulation observed in most eukaryotic members of Group E [262, 268]. The F2,6bP binding site as identified for the PFK from *Saccharomyces cerevisiae* [260] is only present in some of the group E members. Likewise, the presence of the ATP-binding site "N1" as identified in *Pichia pastoris* (group E) [269] and a second ADP-binding site "N2" identified only in

mammalian PFKs [260] is limited to some group E members. Contrary, the eukaryotic group E PFK from *Dictyostelium discoideum* is not under the control of known allosteric regulators [270] and it is unclear if the bacterial members of group E identified in this study are under allosteric control or not. PP_i-PFKs generally lack allosteric control (**Table 6.3**), although some exceptions have been reported [271, 272]. In higher plants, like *Solanum tuberosum*, a gene duplication event in combination with a subsequent change in the F6P-binding site gave rise to an allosteric PFK (group LONG), which is affected by F2,6bP [262, 273], while other characterized bacterial and eukaryotic members of group LONG PFKs lack such type of allosteric control. Although there seems to be a general correlation with respect to the allosteric regulators F2,6bP, AMP, ADP, ATP and PEP and the classified PFK subgroups, there appears to be no strict conservation of allosteric properties within the different subgroups (**Table 6.3**).

Table 6.2 Family A PFK subgroup pairwise cross correlation. For each classified Family A PFK subgroup (columns) the percentage of genomes containing another PFK subgroup (rows) is given. Percentages are given with respect to the number of organisms containing the PFK subgroup (indicated between parentheses in column headers). All the values between brackets indicate the percentage of organisms containing a specific PFK subgroup with respect to the organism investigated (495). The bottom row shows the number (between parentheses) and percentage (between brackets) of organism containing only a single specific PFK subgroup. ATP-specific PFK subgroups B1, E and X are given in bold. Subgroups which predominantly contained PP_i-dependent PFKs are in italic. Group III PFKs consisted of 75% ATP- and 25% PP_i-dependent PFKs (see also **section 3.1.1**).

	LONG *(27)* *[6]*	*SHORT* *(13)* *[3]*	**B1** **(146)** **[30]**	**E** **(30)** **[6]**	III (131) [27]	*Y* *(12)* *[2]*	*P* *(21)* *[4]*	*B2* *(63)* *[13]*	**X** **(33)** **[7]**
LONG		*0*	**4**	**3**	7	*0*	*0*	*0*	**33**
SHORT	*0*		**5**	**0**	2	*8*	*0*	*2*	**3**
B1	*22*	*54*		**0**	27	*33*	*0*	*35*	**0**
E	*4*	*0*	**0**		3	*8*	*19*	*0*	**3**
III	*33*	*23*	**25**	**13**		*33*	*29*	*33*	**45**
Y	*0*	*8*	**3**	**3**	3		*0*	*0*	**0**
P	*0*	*0*	**0**	**13**	5	*0*		*3*	**3**
B2	*0*	*8*	**15**	**0**	16	*0*	*10*		**6**
X	*41*	*8*	**0**	**3**	11	*0*	*5*	*3*	

Single sub-group	*LONG* *(7)* *[1]*	*SHORT* *(2)* *[<1]*	**B1** **(84)** **[17]**	**E** **(22)** **[4]**	III 54 [11]	*Y* *(5)* *[1]*	*P* *(10)* *[2]*	*B2* *(26)* *[5]*	**X** **(10)** **[2]**

Table 6.3 Allosteric properties of family A PFKs with respect to the PFK subgroups and their phosphoryl donor.

				Allosteric effectors					
Organism	Domain	PFK subgroup	PFK P-donor	activator F2,6bP	activator AMP	activator ADP	inhibitor ATP	inhibitor PEP	Reference
Geobacillus stearothermophilus	B	B1	ATP	-	-	yes	-	yes	[261]
Escherichia coli	B	B1	ATP	-	-	yes	-	yes	[274]
Thermus thermophilus	B	B1	ATP	-	-	yes	-	yes	[275]
Thermotoga maritima	B	B1^	ATP	no	no	yes	-	yes	[276]
Pichia pastoris	E	E	ATP	yes	yes	-	yes	-	[277]
Schistosoma mansoni	E	E	ATP	yes	yes	-	yes	-	[278]
Dictyostelium discoideum	E	E	ATP	no	no	no	-	no	[270]
Amycolatopsis methanolica	B	III	PP_i	-	-	-	-	-	[279]
Streptomyces coelicolor	B	III	ATP	-	-	-	-	yes	[280]
Thermoproteus tenax	A	III	PP_i	no	no	no	no	no	[281]
Methylomicrobium alcaliphilum	B	P^^	PP_i	-	no	no	no	no	[282]
Methylobacterium nodulans	B	P^^	PP_i	-	no	no	no	no	[283]
Thermotoga maritima	B	SHORT	PP_i*	-	-	-	-	-	[284]
Trichomonas vaginalis	E	SHORT	PP_i	no	-	-	-	-	[285]
Naegleria fowleri	E	SHORT	PP_i	no	yes	-	-	-	[271]
Giardia lamblia	E	LONG	PP_i	no	no	-	-	-	[286]
Spirochaeta thermophila	B	LONG	PP_i	no	no	no	no	no	[287]
Solanum tuberosum	E	LONG	PP_i	yes	-	-	-	-	[272]
Trypanosoma brucei	E	X	ATP	no	yes	yes	-	yes	[288]
Leishmania donovani	E	X	ATP	-	yes	no**	-	-	[266]
Amycolatopsis methanolica	B	X^	ATP	no	yes	no	-	no	[234]
Acholeplasma laidlawii	B	B2	PP_i	no	no	no	no	no	[289]
Methylococcus capsulatus Bath	B	B2^^	PP_i	-	no	no	no	no	[290]
Methylosinus trichosporium	B	B2^^	PP_i	no	no	no	no	no	[283]

Table 6.3 additional notes

* Polyphosphate is reported as preferred phosphoryl-donor.
** Only GDP tested not ADP.
^ PFKs have been reported to be inhibited by PP_i [234, 284].
^^ Also active on sedoheptulose-7-phosphate.

3.1.2 ADP-dependent phosphofructokinases

Several archaeal species use ADP as a phosphoryl-donor for the phosphorylation of F6P and/or glucose. These reactions are catalysed by the homologous ADP-dependent phosphofructokinase (ADP-PFK) and ADP-dependent glucokinase (ADP-GK), respectively [252, 253, 291]. Alternatively, in some species the ADP-dependent phosphorylation of these substrates is catalysed by the same enzyme i.e. a bi-functional ADP-GK/PFK [292]. ADP-GKs have also been identified in some higher eukaryotic species [293]. These ADP-dependent sugar kinases are closely related and belong to the ribokinase super family [254]. Sequence and structural analysis revealed a high similarity of the archaeal ADP-GKs and ADP-PFKs with respect to their sugar and ADP/AMP binding pockets [254, 294, 295]. Subtle differences within the sugar binding pockets allow the distinction between the three different types of ADP-sugar kinase [254, 294]. Site directed mutagenesis of the *Pyrococcus horikoshii* ADP-PFK substrate binding pocket identified Asn(N)15, Asn(N)160, Lys(K)158 and Arg(R)191 as critical residues involved in recognition of the F6P phosphate group (*Pyrococcus horikoshii* ADP-PFK AA frame, GenBank: AAL80436.1). In ADP-GKs the lysine residue is absent whereas Asn(N)160 and Arg(R)191 are replaced by a His(H) and Asp(D), respectively [294]. The ability to use glucose as substrate is characterized by the presence of a conserved Glu(E) at position 71; this essential glutamate residue forms a hydrogen bond with the C2 hydroxyl group of the bound glucose and is typically replaced by an Ala(A) in all known ADP-PFKs [254, 294]. All of the discussed residues critical for glucose and F6P binding are also present in the bifunctional ADP-PFK/GKs [294].

All known ADP-PFK, ADP-GK and ADP-GK/PFKs sequences belong to COG4809. Identified COG4809 sequences were scanned for the presence of the sugar and adenosine binding pockets. Distinction between the ADP-dependent PFK, GK and the GK/PFKs was made based on the indicated key residues. ADP-GKs could be identified in archaeal, bacterial and eukaryotic species, whereas ADP-PFKs and ADP-GK/PFKs were only identified in archaea (**Supplementary Table S6.D**). The glucose binding pocket of bacterial and eukaryotic ADP-GKs differs from the sugar binding pocket as identified in archaea, notably the conserved Glu(E)71 was absent in eukaryotic species. All ADP-PFK sequences cluster together forming a monophyletic clade, ADP-

PFK/GK sequences form three monophyletic clades and ADP-GKs four distinct monophyletic clades (**Supplementary Table S6.D**). ADP-GKs were not included in the distribution studies.

Overall 3% of the 495 investigated organisms contain genes coding for a ADP-PFK (including the bi-functional ADP-GK/PFK). The taxonomic distribution of the ADP-PFK subgroups is shown in **Supplementary Figures S6.1** and **Supplementary Table S6.A**. The identified 4 ADP-PFK subgroups represented the ADP-PFKs from 3 different archaeal orders and 1 archaeal family. Interestingly, some but not all Halobacteriaceae (family) contain a bi-functional ADP-GK/PFK.

3.1.3 Class I fructose bisphosphatases

All class I fructose bisphosphatases belong to COG0158. This COG also contains the closely related sequences of the class I SBPases [296, 297], which specifically catalyse the hydrolysis of sedoheptulose 1,7-bisphosphate (S1,7bP). Furthermore, some class I FBPases are bi-functional catalysing both the de-phosphorylation of F1,6P and S1,7bP (Class I F/SBPase) [296, 298].

Class I FBPases are known to be under the control of several allosteric regulators. In mammalian class I FBPase AMP and F2,6bP synergistically inhibit FBPase activity [299, 300]. The binding of AMP to its allosteric site induces a conformational change from the active R-state to the inactive T-state, while the competitive inhibitor F2,6bP binds to the active site subsequently leading to an intermediate conformation with disengaged active site loops (I_T-state) [299, 301]. F2,6bP also inhibits the *E.coli* class I FBPase, however its binding does not lead to any conformational change [299]. In general F2,6bP appears not to be present in *E. coli* and for this reason presumed not to be a physiological inhibitor. Alternatively, *E.coli* class I FBPase is synergistically inhibited by the allosteric regulators AMP and Glucose 6-P (G6P) [302]. In addition, PEP acts as an allosteric activator in *E. coli*. The binding of PEP (or citrate) to the anionic binding site located at the subunit interface stabilizes the active tetrameric R-state, and the related conformational change causes the aromatic side chain of Phe(P)15 (*E. coli* Class I FBPase AA frame, PDB: 2GQ1) to occupy the AMP binding pocket [303, 304]. However, PEP only becomes an activator when G6P levels are sufficiently low [302].

Selected sequences were analysed with respect to 1) the active site including the residues involved in the binding of substrate/products (F1,6bP / F6P + P_i), 2) the residues creating the three metal binding sites, and 3) the additional serines involved in F2,6bP binding [299, 302, 305]. Furthermore, the presence of identified allosteric sites of AMP, PEP and G6P within these sequences were investigated [301, 302, 304].

Based on phylogenetic analysis and hierarchical clustering selected COG0158 sequences were classified into 8 different groups (**Supplementary Table S6.D**).

Group classification is in line with the phylogenetic analysis and domain analysis performed by Jiang *et al.*, with the exception of the additional identification of a novel monophyletic group containing archaeal Halobacteria (Class) sequences (FBP-I-G) [296]. All core class I F/SBPases were classified as group FBP-I-A FBPases. However, bi-functionality could not be distinguished based on the investigated catalytic site residues and it is unknown whether this is a general feature of group FBP-I-A FBPases.

Sequence analysis showed that the active site and metal coordinating residues are highly conserved. However, the aromatic residues at position 257 and 259 (*E. coli* Class I FBPase AA frame), interacting with the 6-phosphate group of F6P/F1,6bP, are absent in all SBP_I sequences. Similar observations were made by Raines *et al.* who proposed that these variations determined the substrate specificity [297]. In addition, the Asn(N) residue at position 206, which interacts with the 6-phosphate group of (F6P/F1,6bP), is strictly conserved within groups FBP-I-A, FBP-I-B, FBP-I-C and FBP-I-E, whereas it is replaced by a Gly(G) or Ala(A) in all sequences of group FBP-I-D, FBP-I-F, FBP-I-G, SBP-I. Interestingly, some sequences of the latter groups have hits with the sedoheptulose-1,7-bisphosphatase signature (PR01958) (**Supplementary Table S6.D**). However, the lack of characterisation data prevented proper functional assignment and FBP-I-D, FBP-I-F, FBP-I-G group members were presumed to function as FBPase. Only the SBP-I group sequences were omitted in the distribution analysis of class I FBPases. For group FBP-I-D members, containing only Deltaproteobacteria and Methanomicrobia sequences, the catalytic site residue at position 243, which is generally a Met(M) or Leu(L), is typically substituted by a Phe(F).

Based on the conservation of the residues involved in the binding of F2,6bP, as identified for Porcine and *E. coli* class I FBPase [299], all investigated sequences are potentially targets for F2,6bP inhibition. AMP binding pockets appear to be absent in the FBPase groups FBP-I-C, FBP-I-D, FBP-I-F, FBP-I-G, SBP-I and from the majority of group FBP-I-A sequences, suggesting these proteins lack the AMP inhibition mechanism as identified for the porcine and *E. coli* class I FBPase [301, 302]. The overall low conservation of the AMP allosteric site within group FBP-I-A, FBP-I-B and FBP-I-E members probably reflect the differences in sensitivity towards AMP inhibition [301]. The presence of the allosteric sites of PEP and G-6P appear to be limited to some members of group FBP-I-B FBPase, indicating that the synergistic inhibition by AMP/G-6P and allosteric activation by PEP are not commonalities for prokaryotic organisms. Whereas, analysis of the residues related to the subunit interaction of porcine FBPase [304] showed this subunit interaction only to be present in group FBP-I-E, FBP-I-C, some members of group FBP-I-A and the group FPB-I-B members not containing the PEP allosteric site.

Forty-one percent of the 495 investigated organisms contain genes coding for a Class I FBPase. There is no cross correlation between the identified subgroups, indicating that these organisms only contain a gene or genes coding for a single Class I

FBPase subgroup. The taxonomic distribution of the Class I FBPase subgroups is shown in **Supplementary Figures S6.2** and **Supplementary Table S6.A**. Observed distributions are in line with the findings of Jiang *et al.* [296].

3.1.4 Class II fructose bisphosphatases

Characterisation of the *E. coli glpX* gene revealed it to encode a FBPase with sequential, structural and catalytic properties different from the *E. coli* Class I FBPase [255]. Mutagenesis studies of GlpX and the analysis of its crystal structure identified the catalytic site of this Class II FBPase and the residues critical for FBPase activity [306]. All Class II FBPases belong to COG1494. This COG also includes the Class II bi-functional F/SBPase [296]. These bi-functional enzymes, which hydrolyse both F 1,6bP and S1,7-bP, are only identified in cyanobacteria, [307, 308].

Identified COG1494 sequences were analysed with respect to their active site residues (*E. coli* Class II FBPase GlpX AA frame, UniprotKB: P0A9C9.1, [306]). The three signatures related to the Class II FBPase Interpro ID (IPR004464) were present in all selected sequences, and enabled the classification into 6 different groups (**Supplementary Table S6.D**). So far, no F/SPBase specific signatures have been generated, however, Class II F/SBPase formed a monophyletic group (F/SBPase_II), which allowed them to be distinguished from the other Class II FBPase. *E. coli* glpX and YggF class II FBPases represent group FBP-II-B1, while group FBP-II-B2 included the sequences of *Corynebacterium glutamicum* [309] and *Mycobacterium tuberculosis* [310]. The Class II FBPase identified in *Bacillus subtilis* (gene *YwjI*) [311] is a member of Group FBP-II-B3. FBPase-II-B4 did not contain any characterized protein while group FBP-II-A encompassed all identified archaeal sequences (Methanosarcinaceae).

Although little characterisation data of Class II FBPase is available it is noteworthy that, while PEP inhibits the Class II FBPase of *B. subtilis* (YwjI) and *C. glutamicum* [309, 311] and AMP was shown to inhibit the *C. glutamicum* and *Synechocystis* sp. strain PCC7492 Class II FBPase [307, 309], the *E. coli* GlpX and YggF FBPases and the *M. tuberculosis* class II FBPases are not inhibited by AMP or PEP [306, 310]. These observations demonstrate the different modes of regulation present within Class II FBPase subgroups.

Overall 33% of the 495 investigated organisms contain genes coding for a Class II FBPase. There is no cross correlation between the identified subgroups, indicating that these organisms only contain a gene or genes coding for a single Class II FBPase subgroup. The taxonomic distribution of the Class II FBPase subgroups is shown in **Supplementary Figures S6.3** and **Supplementary Table S6.A**.

3.1.5 Class III fructose bisphosphatase

So far only the Class III fructose bisphosphatases of *Bacillus subtilis* has been characterized (UniprotKB: Q45597). It requires PEP for activation and stability and is inhibited by AMP [256, 312], however, its catalytic residues have not been identified.

Class III FBPases belong to COG3855. Identified sequences revealed to have two or three of the three Fructose-1,6-bisphosphatase, Bacillus type (IPR009164) associated signatures and were all classified as Class III FBPases. Of the 495 investigated organisms, 8% contains genes coding for a Class III FBPase. Class III FBPase are distributed over several bacterial clades (**Supplementary Figures S6.4** and **Supplementary Table S6.A**).

3.1.6 Class V fructose bisphosphatase

Class V fructose bisphosphatases, orthologs of the FBPase gene from *Thermococcus kodakaraensis* (Fbp_{Tk}), had been identified as the true gluconeogenic FPBase for archaeal hyperthermophiles [241]. However, rather than being archaea specific enzymes, these FBPases were later considered as hyperthermophile specific, where acquisition of the gene by several bacteria was proposed to have occurred by lateral gene transfer [241, 313]. Characterisation of the Class V FBPase from *Thermococcus onnurineus* NA1 identified PEP, ATP and ADP as inhibitors [314], while AMP does not seem to affect this class of FBPases [314, 315]. Furthermore, Class V FBPases have been shown to act as a bi-functional FBPase/aldolase. Its bi-functionality ensures that heat labile triosephosphates are quickly removed and trapped in the more stable fructose 6-phosphate, rendering gluconeogenesis unidirectional [313]. Crystal structures of the *Sulfolobus tokodaii* and *Thermoproteus neutrophilus* Class V FBPase have allowed the identification of the residues involved in the coordination of Mg^{2+}-ions, dihydroxyacetone (DHAP), F6P and F 1,6bisP [316-318].

Class V FBPase, Fbp_{Tk} orthologs, are COG1980 members. Analyses of the identified COG1980 sequences showed that all residues involved in binding of the divalent metal ions, F1,6bP and DHAP are extremely conserved, classifying all these sequences as bi-functional Class V FBPase/aldolase enzymes (*Sulfolobus tokodaii* Class V FBPase AA frame, UniprotKB: F9VMT6). Hierarchical clustering and phylogenetic analysis identified 3 monophyletic groups (**Supplementary Table S6.D**), similar to the findings of Say and Fuchs [313]. FBP-V-B is only found in two of the investigated achaeal genomes while FBP-V-A occurs in both archaeal and bacterial species. Group FBP-V-C can only be found in several apparently unrelated bacterial genomes. Overall 14% of the 495 investigated organisms contain genes coding for a Class V FBPase. There is no cross correlation between the identified subgroups. The taxonomic

distribution of the Class V FBPase subgroups is shown in **Supplementary Figures S6.5** and **Supplementary Table S6.A**.

3.1.7 Genotype diversity at the F6P and F1,6bP interconversion node

Of the investigated organisms 83% has one or more genes coding for the investigated FBPases. As expected there is a high cross correlation between FBPases FBP_I_C and F/SBP_II (100%, vica versa 83%), which represent the two types of FBPases identified in Cyanobacteria. Overall 16% of the FBPase-containing organisms possess two or more different FBPase classes, which suggests an apparent functional redundancy for these organisms at the FBPase level (**Supplementary Figures S6.6**). The data show that 25% of the investigated organism, representing the majority of the archaeal genomes and the α,β and ε proteobacteria species, have a FBPase but lack any PFK.

All ADP-PFK containing organism also have one or more genes coding for a FBPase (Class I FBP-I-G, 20%; FBP-II-A, 40%; FBP-V-A, 47%), however, no ATP- or PP_i-dependent PFK encoding genes.

Fifty-eight percent of the investigated organisms have an ATP-PFK and 30% a PP_i-PFK (**Supplementary Figures S6.6**). Several organisms only possess an ATP- or a PP_i-PFK while 18% have both types of PFK (**Table 6.4**). In almost every organism that contains an ATP-PFK gene, but lacks a PP_i-PFK gene, a FBPAse encoding gene can be found (**Table 6.4**). These genomes, 37% of the investigated organisms, represent the classical glycolysis/gluconeogenic genotype at theF6P/F1,6bP level. Thirty-seven percent of the organisms that contain both ATP- and PP_i-PFK do not have a gene coding for a FBPase. For these organisms the PP_i-PFK might provide an alternative for the missing FBPase (7% of all investigated genomes). Alternatively, in 12% of all investigated genomes a PP_i-PFK can be found alongside an ATP-PFK and a FBPase,

Table 6.4 The genomic presence of FBPase in relation to ATP- and PP_i-PFKs containing genomes. Percentage between brackets is given with respect to the total number of genomes investigated (495).

	Only ATP-PFK	Both ATP- & PP_i-PFK	Only PP_i-PFK
Number of genomes	(195)	(91)	(57)
Number (% of total genomes)	[40]	[18]	[12]
Without FBP (% of total genomes)	[2]	[7]	[5]
With FBPase (% of total genomes)	[37]	[11]	[7]
With FBPase (%)	94	63	56
With class I FBPase (%)	50	21	21
With class II FBPase (%)	45	24	33
With class III FBPase (%)	12	12	5
With class V FBPase (%)	6	14	7

revealing an apparent redundancy of the PP_i-PFK gene. Fifty-six percentof the organisms that only contain a PP_i-PFK have a FBPase, representing 7% of the total investigated genomes. In these organisms the PP_i-PFK could replace the classical function of ATP-PFK. Furthermore 5% of all genomes contain only a PP_i-PFK but no ATP-FPK or FBPase (**Table 6.4**).

3.2 Interconversion of phosphoenolpyruvate and pyruvate

The interconversion of PEP and pyruvate is an vital reaction in most organisms. The high energy phosphate bond of PEP allows for substrate-level phosphorylation when converted to pyruvate, and congruently, the reverse reaction (gluconeogenesis) requires input of energy. This glycolytic/gluconeogenic interconversion of PEP into PYR can be catalysed by essentially different enzymes, namely pyruvate kinase, pyruvate phosphate dikinase and pyruvate water dikinase. The PK catalysed reaction is irreversible under physiological conditions and its activity is generally under the control of allosteric regulation. So far sequence analysis allowed the subdivision of PK into 6 different clusters (I-VI) [239]. PPDK and PWDK, which share the same structural organisation and catalyse their reactions via a very similar mechanism [319, 320], appear to lack any allosteric control. Alternatively, regulation by phosphorylation has been described for some plant PPDKs and bacterial PWDKs [321].

PEP can be also converted to PYR via the PEP:carbohydrate phosphotransferase system (PTS) [322]. The PTS catalyses the concomitant translocation and phosphorylation of a carbohydrate. In the first step of the catalytic cascade PTS enzyme I is phosphorylated using PEP as a phosphoryl-donor. The system could theoretically be responsible for up to 50% of the flux through the PEP/PYR node. Whether the PTS is involvemed in metabolism depends on the availability of the corresponding carbohydrate substrate. For this reason the genomic distribution of the PTS enzyme I was not incorporated in this study.

3.2.1 Pyruvate kinase

The PK protein is comprised of three or four distinct domains. The active site is located at the interface of the A (IPR015813, IPR015793) and B (IPR011037, IPR015806) domains, where the B domain is located in-between the two parts of the A domain (A_1-B-A_2). The C-terminal (C) domain generally contains the allosteric binding site (IPR015795, IPR015794). Some eukaryotic PKs possess a small additional N-terminal domain [239, 323]. Furthermore, an extra C-terminal sequence (ECTS) (C') was identified in some *Bacilli* and a cyanobacterium. These ECTSs are homologous to the central swivelling domain (IPR008279) as identified in PPDK, PWDK and the sugar

phosphotransferase system enzyme I, especially the region around the catalytic histidine residue [239] (see also **section 3.2.2**).

Some PKs have an absolute requirement for a monovalent cation like K^+ or NH_4^+, while other PKs display catalytic activity without the addition of any monovalent ion. In the K^+-dependent Rabbit muscle PK, K^+ is coordinated by Asn(N)74, Ser(S)76, Asp(D)112, Thr(T)113 (PK AA numbering, *O. cuniculus* M-PK, GenBank: AAB61963.1) [324]. However, substitution of the Glu(E) at position 177 by a lysine eliminated the requirement for a monovalent ion [325]. Moreover, extensive analysis of available PK sequences revealed a dichotomy with respect to K^+-requirement, where K^+-dependence correlates with a glutamic acid Glu(E) at position 117, while in K^+-independent PKs this residue is generally substituted by a Lys(K)117 [326]. The NH_3^+ moiety of the lysine is proposed to substitute for the monovalent ion [325-327]. However, the K^+-independence is not completely explained by the presence of a charged residue at position 117, for in the K^+-independent PKs of *T. tenax* and *P. aerophilum* the corresponding position is occupied by a serine [323, 326].

Analysis of the crystal structures of the ATP and pyruvate bound rabbit muscle PKs allowed the identification of the residues involved in the binding of these compounds [323, 324, 326, 328]. Sequence analysis revealed that the catalytic site residues involved in the binding of ATP [328], pyruvate [324] and the amino acids coordinating the required Mn^{2+}/Mg^{2+}-ions [324, 327] are highly conserved in eukaryotic, bacterial and archaeal PKs [239].

Many PKs are prone to allosteric regulation, where the binding of the regulator induces the transition between the inactive T-state and the active R-state [327, 329, 330]. Differences in response to the major allosteric activators F1,6bP, F2,6bP, AMP and Ribose 5-P (R5P) allows the distinction of different PK subtypes based on their allosteric control [239, 327, 331-333]. Moreover, reports on PKs lacking allosteric control have been made as well [323, 327, 334, 335].

All known PK sequences belong to COG0469. Identified sequences were analysed with respect to the K^+-binding pocket [324, 326], residues involved in the coordination of the Mn^{2+}/Mg^{2+}-ions [324, 327] the catalytic site [239, 324, 327, 328, 336] and the allosteric binding site [239, 327, 330, 332, 333]. Based on the identity of the residue at position 117 [326] subgroups either were classified as K^+-dependent or K^+-independent PKs. Although a low resolving resolution did not allow the classification of 19% of the PK sequences (2% K^+-dependent, 17% K^+-independent), analysis identified 16 PK subgroups (**Supplementary Table S6.D**). Group nomenclature elaborated on the findings of Muñoz *et al.* [239].

The canonical PK domain organisation (A_1-B-A_2-C), could be identified in all subgroups with the exception of group CX sequences. As expected, additional N-terminal domains (N) could be identified in several eukaryotic sequences in group CI, CII and CV_B. Group CX sequences have both the A and B domains but lack the C-

domain. Instead, all Group CX sequences have an alternative N-terminal sequence (ANTS, designated N'), which generally consists of more than 120 amino acids. In addition, 55% of Group CX contain an ~100 AA insertion in the B domain (designated as B'). Based on the presence of active site residues in these sequences all Group CX members were classified as a novel type of PKs ($N'\text{-}A_1\text{-}(B/B')\text{-}A_2$), however the function of the N' and B' domains is not known (no Interpro scan hits have been reported for these specific sequence stretches).

Extra C-terminal domains (~110-130 AA) could be identified in group CIV (73%), group CV_C (80%), group CVI_C (94%) and in group CY (100%) sequences. For the majority of the ECTSs of group CIV and CVI_C the TSH motif, containing a catalytic histidine(H) of the central swivelling domain (IPR008279, signature PF00391) as identified in PPDK, PWDK and the sugar phosphotransferase system enzyme I [239] and the regulatory threonine(T) residue [337], could be identified. This type of ECTS (designated C') appears to be involved in protein stability/integrity, however, it is unknown whether it has a specific function in regulation [338] (also see **section 3.2.4**). A different type of ECTS could be detected in group CY and CV_C sequences, designated C'' and C''' respectively. None of these ECTSs contained the TSH motif and showed lower sequence similarity with respect to C' sequences and each other (within C' ECTS median similarity 41%, within C'' ECTS 55%, within C''' ECTS 55%, between C' and C'' median similarity 24%, between C' and C'' median similarity 25%, between C'' and C''' median similarity 20%). The function of type C'' and C''' ECTS is not known.

Ninety percent of the investigated organisms contain genes coding for a PK. The taxonomic distribution of the PK subgroups is shown in **Supplementary Figures S6.7A** and **S6.7A** and **Supplementary Table S6.A**. In general organisms either have a K^+-dependent or a K^+-independent PK, with the exception of several α proteobacteria and some eukaryotic organisms that contain both PK types. Almost all α proteobacteria that have a Group CIII PK, also have a Group CVI_B PK. All organisms containing the group CX PK possess and additional PK, however, there is no cross-correlation of CX PKs with a specific other PK subgroup.

3.2.2 Pyruvate, water dikinase and pyruvate, phosphate dikinase

PWDK and PPDK use a similar mechanism while catalysing different biochemical reactions. The PPDK catalysed reversible conversion of PYR to PEP is comprised of three Mg^{2+}-dependent partial reactions, involving a catalytic histidine and pyrophospho- and phospho-enzyme intermediate: 1) E-His + ATP <-> E-His-$P_\beta P_\gamma$ + AMP; 2) E-His-$P_\beta P_\gamma$ + Pi <-> E-His-P_β + $P_\gamma P_i$; 3) E-His-P_β + PYR <-> E-His + PEP. In the PWDK catalysed interconversion partial reaction 2 is replaced by a phosphorylysis reaction (E-his-$P_\beta P_\gamma$ <--> E-his-P_β + P_γ) [319, 320, 339].

The tertiary structures of PPDK and PWDK consist of three domains [339, 340]. The nucleotide binding domain is located in the N-terminal region (IPR002192), while the PEP/PYR binding domain is located in the C-terminal part (IPR000121). In-between these two domains the central swivelling domain is located, which harbours the catalytic histidine (E-His) site (His-domain) (IPR008279).

The His-domain and nucleotide binding domain are involved in partial reactions 1 and 2. Site-directed mutagenesis studies and model investigations of the N-terminal nucleotide binding domain of the *C. symbiosum* PPDK have identified the key residues involved in the pyrophosphoryl transfer reaction and the binding of ATP and P_i [340-343]. Gln(Q)335 and Glu (E)323 coordinate the Mg^{2+}-ion which interacts with the β-P of the bound ATP (*C. symbiosum* PPDK AA reference frame, UniprotKB P22983). The pyrophosphoryl transfer (partial reaction 1) is facilitated through the interactions of Lys(K)22 and the ^{101}GMM^{103}-loop with the γ-P of ATP, and the interaction of Arg(R)92 with the α-P and the purine moieties of ATP [343]. Arg(R)337 together with the second Mg^{2+}, coordinated by Asp(D)321, activate Pi to receive the γ-P from the pyrophosphorylated catalytic intermediate (E-His(H)445-$P_\beta P_\gamma$) (partial reaction 2) [339, 343, 344]. A dramatic conformation transition brings the phosphocarrier residue (E-His(H)445-P_β) in range of the PYR/PEP binding pocket located on the C-terminal domain [339, 344], where the β-phosphate group is transferred to pyruvate resulting in the formation of PEP (partial reaction 3). A series of mutagenesis studies on *C. symbiosum* [345, 346] and the PEP bound crystal structures of *C. symbiosum* and *Zea mays* PPDK [339, 347] identified the critical residues involved in the coordination of PEP and the third Mg^{2+}.

PPDK and PWDK sequences belong to COG574. Preliminary alignments revealed selected COG574 sequences also to include sequences missing the C-terminal PEP-binding domain (IPR000121) e.g. glucan, water dikinase or phosphoglucan, water dikinase [348]. A new subset was generated from COG574 sequences encompassing all three 3 catalytic domains as identified for PPDK and PWDK (IPR000121, IPR002192 and IPR008279). Selected sequences were analysed with respect to the indicated catalytic site residues and used for further study.

Sequence alignment allowed a clear distinction between PPDK and PWDK sequences as reported previously [231, 246]. Notably, the critical residues involved in the nucleotide binding of PWDK and PPDK, as identified for *Clostridium symbiosum* PPDK [343], did not align. The residues involved in the coordination of the γ-P of ATP and Pi in *C. symbiosum* (Asp(D)321, Glu(E)323, Gln(Q)335 and Arg(R)337 make up a highly conserved motif (^{321}DxEx$_{11}$QxR^{337}) that is present in all PPDK sequences. In PWDK these residues make up a similar motif, however, the length of the motif varied between PWDK sequences (DxEx$_{10\text{-}16}$QxR) (**Supplementary Table S6.D**).
With respect to the residues involved in the pyrophosphoryl transfer reaction, Lys(K)22 and Arg(R)92 were identified in both PPDK and PWDK sequences. In PWDKs

sequences the arginine (Arg(R)92) is located in the highly conserved (V/L/F)AVRS motif, which strongly resembles the conserved (V/L/F)SVRS motif identified in PPDKs. However, the ^{101}GMM^{103}-loop [343] located 8 residues upstream the (V/L/F)SVRS motif and conserved in all PPDKs could not be identified in PWDK sequences. The catalytic histidine (His(H)445) and residues making up the PEP/PYR binding pocket were highly conserved in all PPDK and PWDK sequences. Investigated PPDK sequences could not be classified into different subgroups, which reflected the high sequence similarity between PPDK sequences from the three domains of life (minimum identity 34% and median identity 54%), which is in line with the findings of Slamovits *et al.* [246]. PWDKs sequences were classified into six groups (**Supplementary Table S6.D**). The PWDK specific IPR006319 ID signature PIRSF000854, is mainly absent in group PWDK_IV members and not present in group PWDK_V and PWDK_VI sequences.

Of the studied genomes 38% contain a PWDK and 43% a PPDK. Overall only 6% of the investigated genomes contain both PPDK and PWDK coding genes, which represents 15% of all PWDK containing genomes and 13% of the genomes having a PPDK indicating that with a few exceptions PPDK and PWDK generally do not co-occur in a single genome. The taxonomic distribution of the PPDK and PWDK subgroups is shown in **Supplementary Figure S6.8** and **Supplementary Table S6.A**. Usually only a single PPDK gene is present in a genome and only 9% of the PWDK containing genomes have one or more additional PWDK genes.

3.2.3 Genotype diversity at the PEP/PYR interconversion node

The taxonomic distribution of the PK, PWDK and PPDK is shown in **Supplementary Figure S6.9**. Investigation of the genotype diversity (**Table 6.5**) revealed that the mutual presences of either PK & PWDK or PK & PPDK are observed most frequently. Furthermore, all organisms that have both PPDK and PWDK always contain a PK coding gene.

3.2.4 The regulation of PK, PWDK and PPDK

Structure analysis of the yeast PK in complex with the allosteric regulator F1,6bP identified the C-terminal allosteric site [327]. The loop ^{431}Ser(S)Thr(T)Ser(S)Gly(G)Thr(T)Thr(T)^{436} (PK AA numbering, *O. cuniculus* M-PK, GenBank: AAB61963.1), identified in yeast, forms the binding pocket of the 6'-phosphate group of F1,6bP, with the serine/threonine at positions 431, 433 and 436 forming hydrogen bounds with the 6'-phosphate [327]. The same loop is shown to be involved in the binding of F2,6bP [332] and has been suggested to be involved in binding the phosphate moiety of the allosteric regulators AMP and R5P [333]. Based

on sequence analysis this allosteric site loop appears to be conserved in most investigated PKs with a general consensus sequence of ^{431}(S/T)X(S/T/G)GX(S/T)^{436}. However, the loop is specifically absent in PK group CVIII members, suggesting a lack of allosteric control for those group members. Indeed, the characterised PKs of the group CVIII members, viz. *T. tenax*, *Aeropyrum pernix* and *Pyrobaculum aerophilum*, have been shown not to be under the control of any effector [323, 335].

Contrary to PKs, which are generally prone to allosteric control, PPDKs and PWDKs usually lack allosteric control. However, in some eukaryotes and bacteria PPDKs and PWDKs are regulated by post-translational modification [321, 337, 349]. The reversible phosphorylation of the regulatory threonine residue of these dikinases is catalysed by a single bi-functional protein (DUF299 gene family). This regulatory protein both acts as a kinase and a phosphatase, with the ADP-dependent phosphorylation resulting in an inactive enzyme, while the Pi-dependent de-phosphorylation reactivates the enzyme. Only the catalytic intermediate form of PPDK/PWDK, i.e. the protein phosphorylated at the catalytic site histidine residue (E-His(455)-P_β) (*C. symbiosum* PPDK AA reference frame, see **section 3.2.2**) located in the central swivelling domain, is a target for phosphorylation/dephosphorylation. The regulatory threonine residue Thr(T)453 is located 2 amino acids downstream the catalytic His(H)455 residue [321, 349]. Phylogenetic analysis revealed the PPDK regulatory proteins (PDRPs) and PWDK regulatory proteins (PSRPs) to form separate clades [321].

To investigate the taxonomic distribution of the DUF299 gene family, the related Interpro ID (Bifunctional kinase-pyrophosphorylase, IPR005177) was used to scan the selected genomes. Sequences with IPR005177 hits were analysed as described in section 2.4. Analysis resulted in the identification of three monophyletic groups. Based on the alignments of Burnell *et al.* [321] group PSRP members were classified as PSRPs, while group PDRP_I members and group PDRP_II were classified as PDRPs.

Table 6.5 The genotype diversity at the PER/PYR node, with at respect to PK, PWDK and PPDK. Percentages are given with respect to the total number of genomes investigated (495).

Only PK & PWDK	**Only PK & PPDK**	**Only PPDK & PWDK**	**PK & PPDK & PWDK**
27	35	0	6

Only PK	**Only PWDK**	**Only PPDK**	**NONE**
23	6	3	1

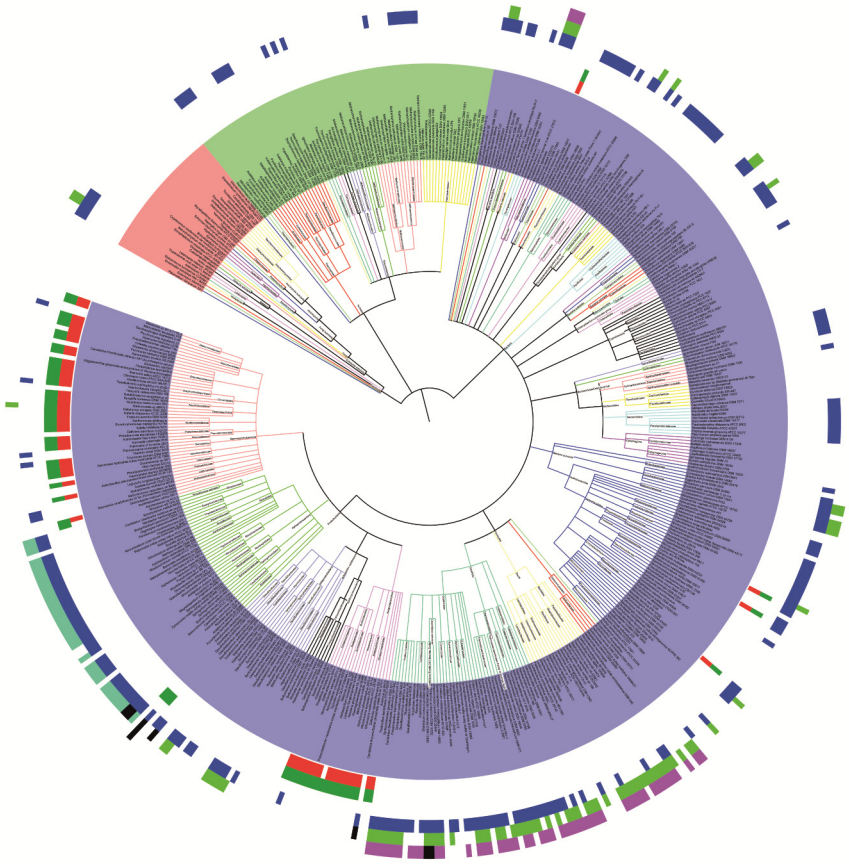


Figure 6.4 Taxonomic distribution the PWDK regulatory protein (2nd circle) in correlation with PWDK subgroup PWDK_II (1st circle) and the PPDK regulatory protein subgroups PDRP_I and PDRP_II (4th circle) in correlation with PPDK (3rd circle) and the ECTS containing members of the PK subgroup PK_IV (5th circle). From inner to outer circle, 1st circle, PWDK_II (**red**); 2nd circle, PWDK regulatory protein (**dark green**); 3rd circle, PPDK (**Blue**); 4th circle, PPDK regulatory protein subgroups, PDRP_ I (**green**), PDRP_II (**turquoise**), unclassified PDRP (**black**). 5th circle, PK subgroup PK_IV with ECTS (**pink**). Additional taxonomic tree labels are given in the capitation of **Figure 6.3**.

Previous analyses showed that PDRP is absent from organisms (e.g. *E. histolytica* and *G. lamblia*) that use PPDK to catalyse the glycolytic conversion of PEP to pyruvate [337]. Likewise, a PSRP is absent in *Thermococcus kodakarensis*, an organism in which PWDK (PWDK_III) plays an essential role in glycolysis [242]. Alternatively, the mutual genomic presence of a specific dikinase and its associated regulator has been

suggested to indicate an exclusive gluconeogenic role of the specific dikinase in that organism [337].

Our findings revealed that the genomic presence of PSRP is correlated only with the group PWDK_II PWDKs (**Figure 6.4**), indicating that this mode of regulation is present in the majority of investigated γ-, β- proteobacteria. The genomic presence of the PPDK regulatory protein subgroup PDRP_II strongly correlates with the PPDK presence in α-proteobacteria. The PDRP_I subgroup appears to correlate with the genomic presence of PPDK in several eukaryotic and bacterial organisms (**Figure 6.4**), however, most investigated Bacilli genomes that have a PDRP_I coding gene generally lack a PPDK coding gene. Moreover, in Firmicutes (which includes the Bacilli class) there is a stronger correlation between PDRP_I and the members of the PK subgroup PK_CIV that have an extra C-terminal domain (ECTS). Ninety percent of the Firmicutes genomes that contain a group PDRP_I PDRP also have a PK_IV with an ECTS; in comparison about 67% of the PDRP_I containing genomes have a PPDK (**Figure 6.4**). The ECTS, which is identified in 73% of the PK subgroup PK_IV (see **section 3.2.1**), contain the central swivelling domain that forms the target for phosphorylation by the PPDK regulatory protein as described above. However, it is not known whether the PPDK regulatory protein (Bifunctional kinase-pyrophosphorylase, IPR005177) can actually phosphorylate this ECTS domain of pyruvate kinases.

3.3 Interconversion of inorganic phosphate and pyrophosphate

Soluble inorganic pyrophosphatases (sPPases) can be divided into the nonhomologous family I and family II sPPases [243]. The hydrolysis of PP_i by sPPases releases the associated free energy as heat, while, membrane-integral pyrophosphates (M-PPases) couple the free energy release to ion translocation or vica versa [243]. All known M-PPases belong to the same protein family [350, 351]. Recently, a novel type of sPPase (Family III), which belongs to the haloacid dehalogenase superfamily (HADSF), has been identified [352, 353]. These latter PPases have not been incorporated in this study since little is known about their physiological function.

3.3.1 Family I soluble inorganic pyrophosphatase

Family I sPPases have been divided into two subtypes based on their sequence and characteristics [354]. The prokaryotic and eukaryotic family I sPPases subtypes, also referred to as type A and B sPPases, respectively [354], share the same structural fold and catalytic site, but generally differ with respect to their quaternary structure [354-361]. Some photosynthetic eukaryotes possess a prokaryotic type family I sPPase homologue in addition to their eukaryotic type sPPases [362].

Family I sPPase sequences are members of COG0221. Identified sequences were investigated with respect to the presence of a set of 14 residues composing the catalytic site of family I sPPase *Saccharomyces cerevisiae* [356] (UniprotKB: P00817). These residues play a crucial role in catalysis [358, 359].

For the majority of the analysed COG0221 sequences the 14 catalytic site residues are completely conserved, however, 33 sequences of bacterial and archaeal origin miss more than 28% (up to 64%) of the catalytic site residues (**Supplementary Table S6.D**). These sequences belonged to 3 different hierarchical clusters, with only one of those clusters forming a monophylogentic group. Phylogenetic analysis placed these sequences between the eukaryotic and prokaryotic family I sPPases. Sequence analysis revealed the absence of a MF_00209 signature (http://hamap.expasy.org/), which is 1 of the 6 signatures associated to the InterPro ID for inorganic pyrophosphatase (IPR008162) and is typically absent in the eukaryotic type sPPases. In addition, sequences from 2 of these 3 hierarchical clusters were also lacking the PANTHER signature for inorganic pyrophosphatase, (PTHR10286, http://www.pantherdb.org/), which was present in all other analysed COG0221 sequences. Though often annotated as a sPPase it can be debated whether these 33 COG0221 members should be regarded as functional sPPases, potentially representing a novel type of family I sPPases, or as non-functional (non-catalytic) sPPases. Based on the nonconformity of their catalytic site they were considered as non-functional sPPases and were not included in our distribution and correlation analysis. The eukaryotic sequences, including the *Saccharomyces cerevisiae* eukaryotic type family I sPPase, that formed a monophyletic group (bootstrap value 90%), were designated as eukaryotic family I sPPases, while the remaining sequences were classified as prokaryotic family I sPPases.

Of the investigated organisms, 67% contain genes coding for the soluble family I PPase (61% prokaryotic family I sPPases, 6% eukaryotic family I sPPases). Only a few eukaryotic species contained both eukaryotic and prokaryotic family I sPPases in their genomes, which is in line with the findings of Gómez-García *et al.* [362]. The taxonomic distribution of the family I sPPases is shown in **Figure 6.5** and **Supplementary Table S6.A**.

3.3.2 Family II soluble inorganic pyrophosphatase

Sequences of the family II sPPases, first discovered in *Bacillus subtilis* [223, 354], occasionally referred to as type C PPases [354], are members of COG1227. This COG also includes cytosolic exopolyphosphatases (PPXs), similar to the PPX1-type PPX identified in *Saccharomyces cerevisiae* [363]. The active site of family II sPPases and PPX1-type PPXs, which both belong to the DHH family of phosphoesterases (IPR001667, IPR004097), are highly similar at the structural level [364]. No specific

InterPro signature exists for PPX1-like exopolyphosphatases or family II sPPases. However, only 10 of the 11 residues making up the catalytic site of the *Streptococcus gordonii* family II sPPase [365] (UniprotKB: P95765) can be identified in the *Saccharomyces cerevisiae* PPX1 [364]. The *S. gordonii* sPPase His(H)9 residue, proposed to be involved in chelation of one of the required two divalent ions [365], is replaced by an Asn(N) in PPX1 (Asn(N)35), whose active site contains only one conserved metal position [364]. Results from the hierarchical clustering and phylogenetic analysis (bootstrap value 98%) showed all sequences with this PPX1 Asn(N)35 residue substitution to cluster together and form a monophyletic subgroup within COG1227 related sequences, thus separating them from the family II sPPases members. These sequences, which were all eukaryotic, were classified as PPXs and were subsequently neglected in the distribution studies since this type of PPX was shown to be inactive on PP_i [363, 364].

The common family II sPPases have a two-domain (DHH and DHHA2) structure, while the family II sPPases subgroup, CBS-PPases, are in general five-domain proteins (DHH-CBS1-DRTGG-CBS2-DHHA2) [366]. For some organisms, like *Moorella thermoacetica*, the DRTGG (IPR010766) domain of the CBS-PPase is lost [367]. Based on the presence of a CBS related signature (Cystathionine beta-synthase, core IPR000644), family II sPPase sequences were classified as either common family II sPPase or CBS-PPase. As expected, all DRTGG domain containing organisms classified as CBS-sPPases. The DRTGG domain (IPR010766) and IPR013785 signatures showed to be correlated with a mutual presence of 73% and 94%, respectively. Furthermore, the MF_00207 signature only gets a hit in common family II sPPases (31%). Both phylogenetic analysis and hierarchical clustering did not reveal CBS-PPases to form a separate cluster, distinguishing them from the common family II PPase. This is in line with findings of Jämsen *et al.* who suggested a common origin for all CBS-PPases as a consequence of a domain insertion occurring in early evolution, subsequently followed by a loss of the CBS and/or DRTGG domain insertions for some organisms [367].

Of the investigated organisms, 23% contain genes coding for the soluble family II PPase (13% common family II sPPase, 11% family II sPPase with an additional CBS-domain) and with the exception of one genome, family II sPPases are absent from the investigated eukaryotic genomes Common family II sPPase and CBS-sPPase do not co-occur in the same organism. The taxonomic distribution of the family II sPPases is displayed in **Figure 6.5** and **Supplementary Table S6.A**.

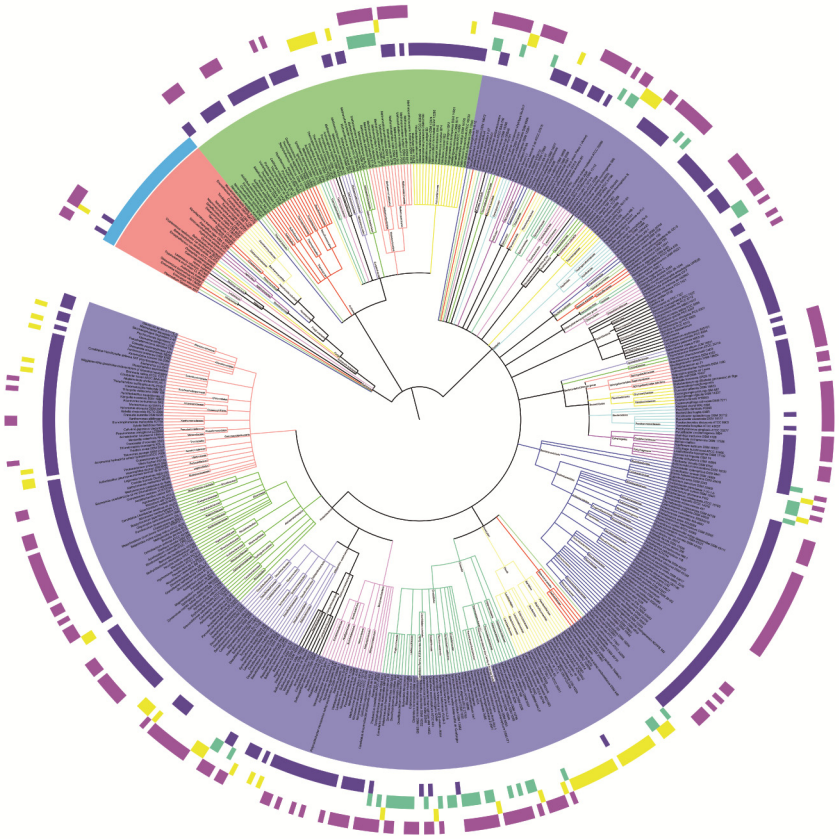


Figure 6.5 Taxonomic distribution of Family I & II soluble PPases (sPPase) and the membrane-integral PPases (M-PPase). From inner to outer circle, 1st circle, Eukaryotic family I sPPase (**light blue**); 2nd circle, prokaryotic family I sPPase (**purple**); 3rd circle, family II CBS-sPPase (**turquoise**); 4th circle, common family II sPPase (**yellow**). 5th circle, M-PPase (**pink**). Additional taxonomic tree labels are given in the capitation of **Figure 6.3**.

3.3.3 Membrane-integral pyrophosphatase

M-PPases can be divided into K^+-dependent and K^+-independent subfamilies [368], while, with respect to the coupling-ion specificity, three subfamilies have been distinguished (H^+, Na^+ or both Na^+ and H^+) [350]. So far, 4 functionally different M-PPases subtypes have been reported: 1) K^+-independent H^+-PPases, 2) K^+-dependent H^+-PPases, 3) K^+-dependent Na^+-PPases and 4) K^+ -dependent Na^+, H^+-PPases. All Na^+-

PPases have an absolute requirement for Na^+ and are additionally activated by K^+ [369], while H^+-PPases either require K^+ or not [370]. The presence of Ala(A) in the conserved M-PPase motif GNXX(**A/K**)AX(G/A/T) correlates with the requirement of a K^+-ion for activity, while the NH_3^+ moiety of Lys(K) substitutes for K^+, leading to K^+-independence [368, 370]. Although the last residue of this motif is not the primary determinant of K^+-dependence both Ala(A) and Gly(G) are characteristically associated with K^+-dependent M-PPase activity while the Thr(T) residue with K^+-independency [368]. The M-PPases catalytic site consists of a hydrolytic site, an "ion-coupling funnel", an ion-gate and a hydrophobic exit channel [243, 371, 372]. Determination of the *Vigna radiate* H^+-PPase crystal structure, binding the PP_i analogue imidodiphosphate, allowed the identification of 9 residues composing the hydrolytic centre [372], which is also conserved in Na^+-translocating M-PPases [371]. Six of the hydrolytic centre residues are distributed over two acidic motifs (DX_3DX_3D) [373, 374] and the PP_i binding motif (DX_3DX_3KXE), which resembles the active site motif $EX_{7\text{-}8}KXE$ of family I sPPases [375]. The exact underlying mechanism for coupling-ion specificity is not known, however, ion-specificity could be predicted based on phylogenetic analysis and, for the majority of K^+-dependent M-PPase, on the position of the pump specificity linked Glu(E) [351]. This Glu(E) residue is a part of the ion-gate associated triplet (D/E/K), as identified in the *T. maritima* K^+-dependent Na^+-PPase [371], however, *Flavobacterium johnsoniae* and *Arabidopsis thaliana* (AVP1)-like K^+-dependent H^+-PPases are characterised by conserved positional shifts of the Glu(E) residue [351].

All M-PPases sequences belong to COG3808. The hydrolytic site residues, K^+-dependence motif, PP_i binding motif, acidic motifs, ion-gate triplet and the correlated Glu(E) residue shift, were used to investigate and classify the COG3808 sequences. Based on phylogenetic and hierarchical clustering analysis, COG3808 sequences were subdivided into 21 different groups (**Supplementary Table S6.D**). Subgroup nomenclature elaborated on the findings of Belogurov *et al.* and Luoto *et al.* [351, 368], 'LX' subgroups designate the K^+-independent (Lys(K)537, *Vigna radiate* H^+-PPase numbering) sequences while 'AX' subgroups designate K^+-dependent (Ala(A)537) sequences. Six of the identified groups did not contain a core set sequence. Of the 21 groups 11 can be specifically identified based on the unique consensus within the investigated sequence motives (**Supplementary Table S6.D**). In addition, for most members of the monophyletic subgroup AE, which included the recently characterized Na^+, H^+-PPase of *Bacteroides vulgatus* [351], the Na^+, H^+-PPases specific (T/S)/F/D/M signature [351] could be identified (strictly conserved 83%), but of these 4 residues only the Asp(D)204 residue equivalent cannot be identified outside the AE subgroup.

Analysed COG3808 sequences, with the exception of sequences belonging to group LY, share an average identity of 48% and ≥30% pairwise amino acid identity. These

findings are in line with Luoto *et al.*, who stated that known M-PPases share 48% average and ≥33% pairwise amino acid identity [351]. In contrast, group LY members share ≤32% pairwise amino acid identities with respect to the other COG3808 sequences, while within the LY group identities are ≥57%. All residues composing the hydrolytic centre and ion gate triplet were conserved in group LY sequences. However, notable differences between group LY members and other M-PPases can be observed in the PP_i binding motif, acid motif I and the K^+-dependence motif (**Supplementary Table S6.D**). Of the 4 IPR004131 ID (Pyrophosphate-energised proton pump) related signatures typically present in all other M-PPases sequences only three could be identified in group LY sequences; specifically signature TIGR01104 was absent. Moreover, some group LY sequences have 18 predicted transmembrane helixes (TMHMM v.2.0, http://www.cbs.dtu.dk/services/TMHMM-2.0/), while known M-PPases generally possess 14-17 transmembrane domains [376]. Based on the observed position of the LY clade in the phylogenetic tree, the typical lysine residue in the GNXX(**A/K**)AX(G/A/T) motif and given the fact that all characterized K^+-independent PPases are H^+-translocating [351], group LY members are proposed to represent a novel type of K^+-independent H^+-PPase.

Table 6.6 The PPase genotype diversity. Correlation between genomic co-occurrence of family I sPPase, family II sPPase and the membrane-integral PPase (top panel) and the PPase associated subtypes (bottom panel). Correlations are given percentages of total genomes investigated (495).

	M-PPase	Family I PPase	Family II PPase
M-PPase	46	27	8
Family I PPase	27	66	2
Family II PPase	8.3	2	23
ONLY a single PPase	10	38	14

K^+-independent, coupling H^+	28	21	2
K^+-dependent, coupling H^+	6	4	1
K^+-dependent, coupling H^+,Na^+	3	1	1
K^+-dependent, coupling Na^+	11	3	5
sPPase I prokaryotic	27	61	2
sPPase I eukaryotic	1	6	<1
sPPase II, common	<1	1	13
sPPase II, with CBS-domain	8	<1	11

An overview of the taxonomic distribution of the identified M-PPase subgroups can be found in **Supplementary Table S6.A**. **Supplementary Figure S6.10** displays the taxonomic distribution of the four, functionally different M-PPases subtypes. Of the investigated organisms, 46% have a M-PPase of which 11% is capable of using both H^+ and Na^+ as coupling ion while 68% of these organisms only use H^+ and 21% only use Na^+ as a coupling ion. Most investigated eukaryotes that possess M-PPase encoding genes have both a K^+-independent H^+-PPase and a K^+-dependent H^+-PPases.

3.3.4 Correlations between the soluble PPases and membrane-integral PPase

For 2% of the investigated organisms no gene encoding a soluble PPase or Membrane-integral PPase could be identified. Only 9 out of the 495 investigated genomes contain both Family I and II PPase, indicating that the family I and II sPPases are generally mutual exclusive (**Figure 6.5**, **Table 6.6**). While 77% of the organisms containing a family II CBS-sPPase additionally have a membrane-integral PPase, only 2% of the genomes with a common family II sPPase coding gene also have a membrane-integral PPase. This indicates that with some minor exceptions M-PPase and common family II sPPase do not co-occur in the same genome (**Figure 6.5**). These findings are in line with the observation by Jamsen *et al.* [367].

3.4 Classical glycolysis/gluconeogenesis versus PP_i-involved glycolysis/gluconeogenesis

An overview of the taxonomic distribution of genomes containing PP_i-PFKs, PPDKs and membrane-integral PPases is given in **Figure 6.6**. Overall 21% of the investigated organisms contain both PP_i-PFK and PPDK coding genes, representing 70% of the PP_i-PFK and 49% of the PPDK coding genomes (**Table 6.7**). Twenty-four percent of the organisms containing both PP_i-PFK- and M-PPase-encoding genes represent 80% of the PP_i-PFK and 53% of the M-PPase encoding genomes, whereas 34% of the investigated organisms containing both PPDK- and M-PPaseencoding genes represent 78% of the PPDK and 74% of the M-PPase encoding genomes. Overall 19% of the investigated organisms have a PPDK, PP_i-PFKs and M-PPase gene.

The gene combination related to classical glycolysis, defined as containing ATP-PFK and PK but no PP_i-PFK or PPDK (**Figure 6.7**), can be identified in 26% of the investigated organisms, while the classical gluconeogenesis combination, defined as containing FBP and PWDK but no PP_i-PFK or PPDK, can be identified in 28% of the investigated organisms. However, in the absence of the gluconeogenic enzyme PWDK, the anaplerotic enzyme phosphoenolpyruvate carboxykinase can replenish PEP from the tricarboxylic acid cycle intermediate oxalocatetate [377], this indicates that the

amount of organisms with a classical gluconeogenesis (genomes lacking PPi-PFK and PPDK) might be higher than 28%.

Alternatively, 52% of the studied genomes contain PP_i-PFK- and/or PPDK-encoding genes, either in the presence or absence of the genes associated with the classical glycolysis/gluconeogenesis. These organisms, which represent the majority of the investigated genomes, have a non-classical glycolysis/gluconeogenesis, tentatively called PP_i-involved glycolysis/gluconeogenesis. There is a low cross correlation between the classical glycolysis/gluconeogenesis and M-PPase, i.e. 10% of the M-PPase-containing organisms have the classical glycolysis enzyme combination and only 7% of the M-PPase-containing organisms have the classical gluconeogenesis enzyme combination.

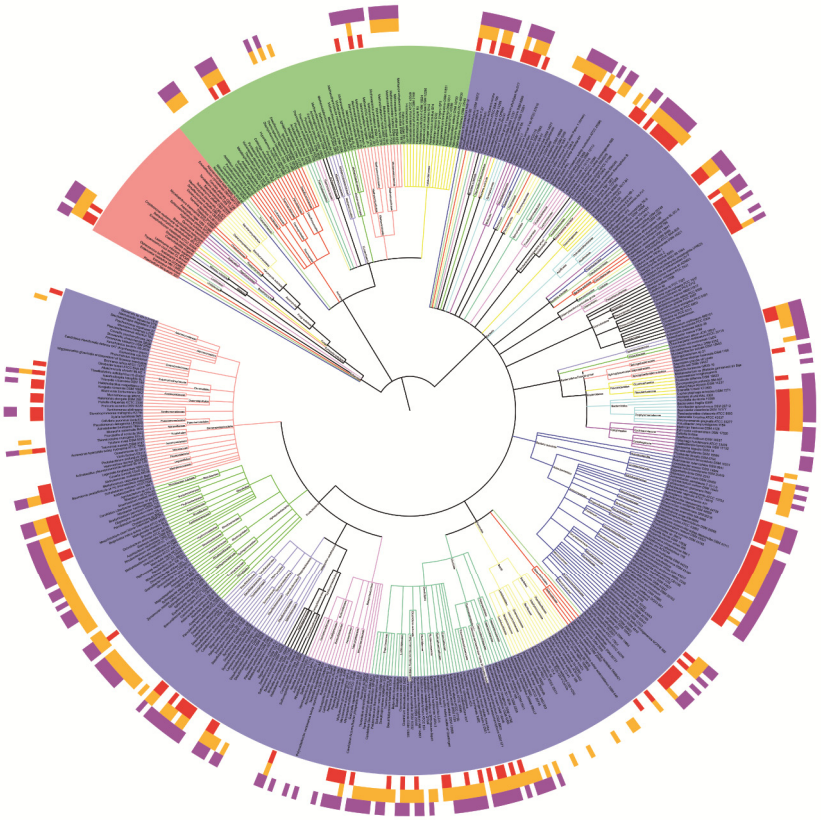


Figure 6.6 Taxonomic distribution of PP_i-PFK (1st circle, red), PPDK (2nd circle, orange) and membrane-integral PPase (3rd circle, pink). Additional taxonomic tree labels are given in the capitation of **Figure 6.3**.

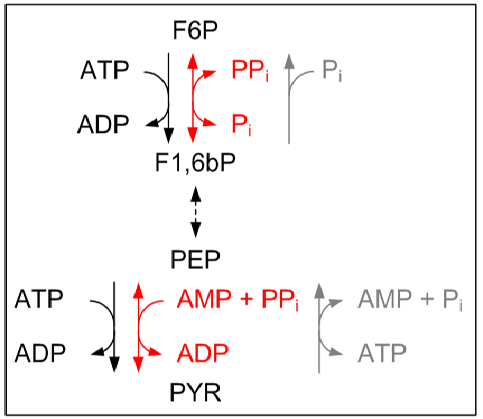
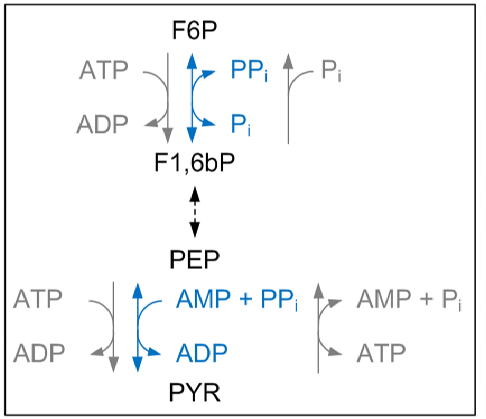
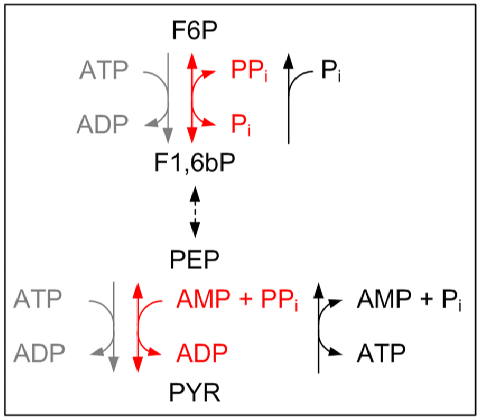
Classic Glycolysis [26%]	**PPi-involved Glycolysis/Gluconeogenesis** [52%]	**Classic Gluconeogenesis** [28%]
Genotype: ATP-PFK and PK present. PP_i-PFK and PPDK absent. Presence or absence of the classic gluconeogenic enzymes. With M-PPase (10%, 16%)	Genotype: PP_i-PFK and/or PPDK present. Presence or absence of the classic glycolytic/gluconeogenic enzymes. With M-PPase (85%, 75%)	Genotype : FBP and PWDK present. PP_i-PFK and PPDK absent. Presence or absence of the classic glycolytic enzymes. With M-PPase (7%, 12%)
		

Figure 6.7 Distinction of different types of glycolysis/gluconeogenesis based on the genomic presence of PP_i-PFK and PPDK. The PP_i-involved glycolysis/gluconeogenesis is more widely distributed (percentage of total investigated genomes between brackets). Most M-Ppase containing organisms have a PP_i-involved glycolysis/gluconeogenesis (between parentheses: 1st value is expressed as percentage of total genomes containing a M-Ppase coding gene, 2nd value is expressed as percentage of total genomes containing the specific glycolysis/gluconeogenesis type).

Table 6.7 Mutual genomic presence of PP_i-PFK, PPDK and M-PPase. Top panel: Mutual genomic presence of PP_i-PFK, PPDK and M-PPase expressed in percentage relative to the number of genomes containing the specific gene (between parentheses). Bottom panel: Mutual genomic presence of PP_i-PFK, PPDK and M-PPase expressed in percentage (between brackets) relative to the total genomes investigated (495).

	PP_i-PFK (149)	PPDK (214)	M-PPase (226)
PP_i-PFK		49	53
PPDK	70		74
M-PPase	80	78	
All three	63	44	42

PP_i-PFK	[30]	[21]	[24]
PPDK	[21]	[43]	[34]
M-PPase	[24]	[34]	[45]

Contrary, there is a high cross correlation between the PP_i-involved glycolysis/gluconeogenesis genotype and the genomic presence of M-PPase. Eighty-five percent of the M-PPase-containing organisms have a PP_i-PFK- and/or PPDK-encoding gene (**Figure 6.7**). None of the investigated genomes contained both PP_i-PFK and PPDK and simultaneously lack the classical glycolytic/gluconeogenic related genes.

4. Discussion

4.1 Enzyme subtype classification

The diversity and taxonomic distribution of the enzymes (enzyme subtypes) involved in the reactions displayed in **Figure 6.1** were investigated using 495 fully sequenced genomes. In general, the assignment of a specific function to a gene is based on the characterisation of its gene product. Supported by comparative genomics, sequence analysis and other bioinformatic tools, known functions can be attributed to uncharacterized genes, primarily based on their associated protein sequences. To be able to distinguish between functionally different proteins within a specific protein family, e.g. substrate specificity of specific allosteric control, proper protein

characterization and classification is required. Within this study reaction associated COG identifiers [247, 248] were used to couple genes, via their protein sequences, to chemical reactions. Subsequently, all investigated proteins were assigned to specific protein family subtypes (subgroups) based on protein sequence analysis. Addition of proteins whose structures have been resolved and/or that were biochemically characterised prior to alignment, allowed the integration of known literature data for proper protein classification.

This COG based analysis provided several advantages over the use of alternative protein associated signatures like Pfam, Superfamily, Hamap, PIR, TIGRFAMS, CATH/Gene3D, PROSITE patterns, PRINTS, Panther, ProDom and SMART, that are all incorporated in the InterPro protein sequence analysis & classification tool [251]. Of these indicated signatures only Pfam signatures [378] can be identified for all the investigated chemical reactions, while the other signatures are often only associated to specific enzyme subtypes. In some cases multiple Pfam signatures are assigned to a single enzyme family, e.g. three Pfam signatures, PF00391, PF00224 and PF02887 representing three different domains are associated to PK, while only one COG ID is related with a single enzyme family. Sometimes a single COG ID or Pfam signature is associated with different chemical reactions, e.g. COG0205 and PF00365 refer to both family A ATP- and PP_i-PFK sequences. In such cases analysis of the substrate binding residues allowed the distinction between the two enzyme subtypes.

Overall, the COG based gene selection in combination with phylogenetic and hierarchical clustering analysis provided a robust method that allowed the identification of all previously reported enzyme subtypes and even led to the identification of novel subtype hitherto not mentioned in literature. Notably, despite revealing all residues composing the hydrolytic centre and ion gate triplet associated to all other characterized M-PPase group LY members have, as far as the author knows, never been incorporated in any phylogeny and characterization studies [351, 368]. This is probably caused by the chosen sequence selection criteria, for example Luoto *et al.* used only the *Rhodospirillum rubrum* H^+-PPase sequence as a query sequence in a BLAST search to form their M-PPase data set and group LY M-PPases could have been missed based on their low sequence similarity/identity with the *R. rubrum* sequence [351].

However, the current resolving resolution sometimes appeared to be too low to distinguish between the subtle differences within the classified protein family subtypes with respect to their characteristics, as is e.g. the case with the family A PFK group E, where several group members are susceptible to a large variety of allosteric regulators while other members lack any known allosteric control. Furthermore, apart from the distinction based on the phylogenetic and hierarchical clustering, many subgroups identified within this study still lack characterization data that may distinguish them from other enzyme subgroups, thus validating the classification into

different subgroups, with respect to functional properties and/or biochemical characteristics.

4.2 Apparent enzyme redundancy

Analysis of the taxonomic distribution of the Family A PFK subgroups revealed an apparent redundancy i.e. multiple distinctly different enzyme subtypes are represented in a single organism. Especially group III PFK is often observed in co-occurrence with one or more other PFK group members, either coding for ATP-dependent and/or PP_i-dependent PFKs (**Figures 6.3** & **Supplementary Figure S6.6**). Also at the F6P/F1,6bP level a seemingly functional redundancy can be observed; 12% of the investigated organisms have a FBPase and possess both ATP-dependent and PP_i-dependent PFKs. To prevent such redundancies it is likely that for these genotypes regulation at the transcriptional level or the enzymatic level is tightly controlled [234, 237], and it seems plausible that different physiological functions can be ascribed to the different enzymes within a single organism. For example, the PP_i-PFK of the bacterial methylotrophic actinomycete *Amycolatopsis methanolica* is exclusively associated with growth on glucose, while the PP_i-PFK is completely replaced by ATP-PFK during growth on one carbon compounds, thus indicating the different physiological roles of the PFK types for this specific organism [234]. Furthermore cellular localisation appears to be an important feature with respect to enzyme physiological function, e.g. in the glycosome of the eukaryote *Trypanosome cruzi* PPDK is responsible for the conversion of PEP to PYR, additionally replacing the function of the soluble inorganic pyrophosphatase (sPPase), while at the same time PK is responsible for the conversion of PEP to PYR in the cytosol (**Acosta 2004 E58**).

4.3 The involvement of PP_i as phosphoryl-donor and its role in energy metabolism

With respect to the participation of PP_i-PFK and PPDK in glycolysis and the use of PP_i as phosphoryl-donor several reports have been made [231, 234, 237, 244-246]. The involvement of PP_i-PFK in glycolysis would save the investment of an ATP at the level of F6P and the participation of PPDK, which couples the phosphorylation of AMP into ATP to PP_i consumption, makes an additional investment of an ATP to phosphorylate AMP to form 2 ADP (via adenylate kinase, EC 2.7.4.3) needless. If electron transport phosphorylation would be the main source of ATP generation the relative increase in ATP yields is minimal. In contrast, when glycolysis is the sole source of ATP, as in fermentative metabolism, the participation of the PP_i-PFK and PPDK would enhance the overall ATP yield of glycolysis significantly.

The required PP_i can originate from biosynthesis, where PP_i is a by-product of the synthesis of proteins, polysaccharides and polynucleotides (DNA and RNA) [68].

However, Zhou *et al.* proposed that biosynthesis alone cannot provide enough PP_i to sustain glycolytic fluxes i.e. biosynthetic reactions only provide 20% of the PP_i required to sustain the total glycolytic flux via PP_i-PFK [379]. Using their model output values this corresponds to 10% of the total glycolytic flux when only PPDK is involved and 7% if both PP_i-PFK and PPDK participate in a strictly PP_i-dependent glycolysis. Although these values could increase when protein and mRNA turnover rate are taken into account (higher turnover rates would result in higher PP_i formation rates), they clearly indicate that an additional PP_i source is required to completely sustain the flux through the PP_i-dependent glycolysis steps. Albeit likely representing a smaller portion of the total glycolytic flux the recruitment of PP_i-PFK or PPDK for PP_i scavenging, i.e. running in parallel with the classical glycolytic enzymes, can still results in an energetic advantage under low energy conditions in comparison to the situation where PP_i is simply hydrolysed by sPPases. Furthermore, for some organisms a particular PP_i-dependent glycolytic conversions is the only known alternative.

Several alternative sources for PP_i have been proposed [379]. 1) the PP_i-generating metabolic reactions: PP_i-dependent acetate kinase and PP_i-dependent PEP carboxylase [320]. 2) ADP-glucose synthase activity, as a part of the glycogen cycling pathway, generate PP_i at the expense of ATP. 3) ATP-pyrophosphatase activity, which also requires the investment of ATP to form PP_i. 4) PP_i could be generated in the M-PPase catalysed reaction driven by a H^+-or Na^+-electrochemical gradient. For the first three alternatives ATP is simply exchanged for PP_i in an 1:1 ratio, which does not lead to a net energetic advantage. While with respect to the latter, PP_i formation from two inorganic phosphates requires less energy (-21.92 KJ/mol) compared to the formation of ATP from ADP and P_i (-31.8 KJ/mol) [4], PP_i synthesis via an electrochemical gradient requires a lower gradient compared to ATP synthesis [68] and the coupling stoichiometry for M-PPase (2 H^+ per PP_i) is lower compared to the H^+/ATP stoichiometry typically associated with F-type H^+-ATPases (3.6 H^+ per ATP) [380]. So investment of the membrane gradient to form PP_i, which is subsequently used as a phosphoryl-donor in glycolysis, could lead to an energetic advantage.

In this study 35% of the investigated genomes contain both M-PPases and sPPase encoding genes (**section 3.3.4**), suggesting that the genomic co-occurrence of soluble and membrane bound PPase is widely distributed among different taxa. The presence of both a sPPase and a M-PPase in the same cellular compartment could lead to an undesirable situation where the energy stored in the trans-membrane H^+ or Na^+ electrochemical gradient is simply lost as heat, via the intermediate PP_i. This indicates the requirement of a tight regulation of these inorganic PPases, either at the enzyme level or transcriptional level.

Unlike the common family II PPases the family II PPase subgroup CBS-PPases are subjected to regulation by adenine nucleotides. Specifically, CBS-PPase activity of

Moorella thermoacetica is inhibited at low energy charge, with AMP and ADP acting as strong inhibitors while ATP stimulates activity [367]. The difference in regulatory properties of family II PPases might form the basis for the high correlation between the genomic presence of both the CBS-PPase and a M-PPase. Seventy-seven percent of family II PPase subgroup CBS-PPases containing organisms also have a M-PPase encoding gene, while with a few exceptions common family II sPPase and M-PPase do not co-occur in the same genome (**Table 6.6**, **Figure 6.5**). Aerobically grown *Rhodospirillum rubrum* cells revealed the presence of high sPPase transcript levels and enzyme activity (family I sPPase, prokaryotic subtype) but no M-PPase transcription [381]. However, acute salt stress induced expression of the M-PPase, indicating a tight regulation of the *R. rubrum* M-PPase [381]. Moreover, the up-regulation of M-PPase transcripts in response to stress has also been reported in several plant species [382]. These observations might indicate that transcriptional regulation provides the important mode of regulation of M-PPase activity in organisms containing a family I PPase.

Moreover, a concomitant ATP-, PP_i-PFK and sPPase activity might results in a futile cycle i.e. the energy stored in ATP is lost via PP_i and PPase activity without any net glycolytic conversion. To effectively apply PP_i as phosphoryl-donor, regulation of PPase activity is required. Beschastnyi *et al.* observed high PP_i-PFK activity (group P and B2 PFKs) accompanied by low sPPase activity and an absence of ATP-PFK activity in several aerobic methylothropic bacteria. While high sPPase activities were detected in those methylothropic bacteria displaying high ATP-PFK activity, thus these data demonstrate an interplay between PP_i-, ATP-PFK and sPPase activity [383].

Under gluconeogenic conditions the PP_i generated by PP_i-PFK or PPDK activity can be directly hydrolysed by a sPPase, overall yielding the same net reaction as their classical gluconeogenic counterparts FBPase and PWDK, respectively. Alternatively, the generated PP_i could be utilized by organisms containing a M-PPase to establish an electrochemical gradient thus exploiting the available energy stored in the phosphate bond.

Overall the genomic presence of M-PPase has been proposed to contribute to cell survival under low energy conditions and abiotic stress. This hypothesis is supported by the observation that the introduction of M-PPase in eukaryotes and prokaryotes, which lack a M-PPase coding gene, confers a higher tolerance toward abiotic stress [382, 384]. A switch from a soluble PPase to the M-PPase in a two PPase system may allow the cell to handle low energy conditions more efficiently than cells containing only a sPPase i.e. by utilizing the energy of PP_i to establish a trans-membrane electrochemical gradient that can be subsequently used for ATP synthesis or active transport of solutes. However, on the basis of the genomic presence alone it is not possible to assign a specific role for M-PPase. Likewise, PP_i-PFK and PPDK could specifically fulfil a role in either glycolysis or gluconeogenesis or, alternatively, these

enzymes could have both a glycolytic and gluconeogenic function, depending on the environmental/growth conditions. While utilization of PP_i as a phosphoryl donor and/or energy source could provide an energetic advantage, a general disadvantage of PP_i-mediated glycolysis/gluconeogenesis is the strong coupling of biosynthesis and catabolism, where the hampering of either system as a result of increased PP_i levels could lead to the cessation of growth. In addition, involvement of PP_i-dependent parallel pathways could annul the allosteric control of glycolysis/gluconeogenesis at the level of ATP-PFK and PK. Furthermore, although the regulatory properties of PP_i with respect to glycolytic/gluconeogenic enzymes has been marginally investigated several reports on the inhibitory effect of PP_i on ATP-PFK [234, 284] and PK activity [107, 229] have been made. And sPPase overexpression studies in *Toxoplasma gondii*, a parasitic organism with a PP_i-dependent glycolysis, revealed PP_i to have a significant regulatory role in glycolysis for this organism [385].

5. Conclusion

In this study the diversity and taxonomic distribution of the enzymes associated with the glycolytic/gluconeogenic interconversion steps of F6P/F1,6bP and PEP/PYR alongside the soluble and integral membrane pyrophosphatases has been investigated (**Figure 6.1** and **Table 6.1**) in 495 organisms, representing the three domains of life. Within the framework of this survey several observations were made:

- The COG based approach to correlate gene and catalytic function allowed the identification of novel enzyme subtypes in addition to all previously reported enzyme subtypes. Novel subtypes included a PP_i-PFK subtype (group Y, **section 3.1.1**), several PK subtypes with non-canonical domain architectures (**section 3.2.1**) and a new M-PPase subgroup (group LY) (**section 3.3.3**).
- Overall nine family A PFK subgroups were distinguished. Most organisms possess only a single PFK subgroup, however 25% of the investigated genomes contain two or more genes encoding PFKs belonging to different subgroups (**Table 6.2**). The pairwise cross correlations between ATP-PFK groups E, B1 and X is low, likewise, the simultaneous genomic presence of PP_i-PFK subgroups LONG, SHORT, Y, P and B2, which is hardly observed. Although some patterns with respect to the taxonomic distribution of the family A PFK associated genotypes can be observed (**Figure 6.3**), there appears to be no strict conservation of genotype at the Class level of the investigated organism (**section 3.1.1**).
- Of all investigated organisms 40% contain only an ATP- and no PP_i-dependent PFK (37% also contain a FBPase), while 19% have both an ATP- and PP_i-dependent PFK (12% also contain a FBPase) and 12% of the organisms only

have a PP_i-PFK (7% also contain a FBPase) (**Table 6.4**). These data reveal the wide distribution of the PP_i-dependent PFK (30%) (**Supplementary Figure S6.6**).

- The genomic presence of ADP-PFK does not correlate with family A PFKs, while all ADP-PFK containing organisms (3%) possess either a class I, II or V FBPase encoding gene (**section 3.1.2**).
- Eighty-three percent of the investigated genomes have one or more FBPase encoding genes. Within the Class I, II and V FBPases there is no correlation between the identified subtypes (no subtypes distinguished for Class III FBPases), however, 16% of the FBPase-containing organism possess two or more different FBPase classes, leading to an apparent functional redundancy at the FBPase level for these organisms (**Supplementary Figure S6.6**).
- Of the investigated catalytic reactions PK appeared to be the most widely distributed (90% of all genomes contained a PK).

- In general PWDK containing organisms (38%) additionally possess PK, likewise the genomic presence of PPDK (43%) frequently co-occurs with the presence of a PK encoding gene. Furthermore all organisms containing both PWDK and PPDK, making up 6% of all investigated organism, also have a PK (**Table 6.5**).
- Investigation of the taxonomic distribution of the PWDK and PPDK regulatory proteins, PSRP and PDRP, respectively, revealed that the genomic presence of PSRP is only correlated with the PWDK subgroup PWDK_II (**Figure 6.4**). PPDK regulatory protein subgroup PDRP_II showed a strong correlation with the PPDK present in α-proteobacteria. Although PDRP_I subgroup seems to correlate with the genomic presence of PPDK in several eukaryotic and bacterial organisms, there appeared to be a stronger genomic correlation between PDRP_I and members of the PK subgroup PK_CIV that have an extra C-terminal domain (**Figure 6.4**) which includes the central swivelling domain that generally forms the target for phosphorylation by the PPDK regulatory protein.
- Of the investigated genomes 88% contain a soluble PPase and 46% a M-PPase. Family I and II sPPases are generally mutual exclusive. Most family II CBS-sPPase containing organisms additionally have a M-PPase, while usually common family II sPPase and M-PPase do not co-occur in the same genome (**Table 6.6**, **Figure 6.5**).
- About 52% of the studied genomes contain PPi-PFK (30%) and/or PPDK (43%) encoding genes and there is a high cross correlation between the genomic presence of a PP_i-involved glycolysis/gluconeogenesis and M-PPase (**Figure 6.7**). 85% of the M-PPase containing organisms have PP_i-PFK and/or

PPDK encoding genes either in the presence or absence of the genes associated with the classical glycolysis/gluconeogenesis. While only 10% of the M-PPase containing organisms have the classical glycolysis enzyme combination (ATP-PFK/PK) and only 7% of the M-PPase containing organisms have the classical gluconeogenesis enzyme combination (FBPase/PWDK).

Based on the genomic presence of PP_i-PFK and/or PPDK a distinction between the classical glycolysis/gluconeogenesis and PP_i-involved glycolysis/gluconeogenesis can be made. The genotype identified with the PP_i-involved glycolysis/gluconeogenesis appeared to be present in the majority of the investigated organisms revealing the widespread distribution of both PP_i-PFK and PPDK across the three domains of life. However, on the basis of the genomic presence alone it is not possible to assign a specific role for PP_i-PFK or PPDK in glycolysis and/or gluconeogenesis, alternatively these enzymes might catalyse both glycolytic and/or gluconeogenic reactions albeit under different growth conditions thus providing functional flexibility. In general, the involvement of PP_i as a phosphoryl-donor in glycolysis or as an energy source to establish an electrochemical gradient could provide an energetic advantage over the alternative use of ATP and could contribute to the stress survival under low energetic conditions. Given the genomic distributions of the PP_i-PFK, PPDK and M-PPase encoding genes and their high level of genomic co-occurrence, PP_i metabolism appears to be an important feature of the glycolysis/gluconeogenesis and energy metabolism.

Supplementary materials

Supplementary Figures S6.1-S6.10
Supplementary Tables S6.A-S6.D

Supplementary materials are available upon requist

Chapter 7

Thesis summary and general discussion

Thesis Summary

Hydrogen gas (H_2) is an important chemical commodity used in many industrial processes. Alternatively, H_2 can be used directly as a fuel in combustion engines or fuel cells. Presently more than 90% of the global H_2 production is based on the use of fossil fuels. Considering the non-renewable nature of fossil fuels and the negative environmental impact associated to their use (e.g. greenhouse gas emissions) these H_2 production processes are presumed to be non-sustainable for the long term. This warrants the development of sustainable H_2 production processes, based on renewable resources. Microbial H_2 formation via dark fermentation has been opted as one of the potential alternative H_2 production processes. During this process the conversion of renewable resources, i.e. biomass-derived carbohydrates from plant biomass or industrial waste streams, results in the formation of H_2.

The research presented in this thesis aims to contribute to the emerging field of biological H_2 production via dark fermentation. The described studies centre around the molecular processes occurring within the H_2-forming thermophilic anaerobic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima*, mainly focussing on their metabolic pathways and their H_2 production capabilities. These organisms were selected based on their capacity to produce H_2 with yields close to the theoretical maximum, on their ability to convert a wide variety of bio-mass derived carbohydrates, and on the availability of their sequenced and annotated genomes (**chapter 1**).

Crude glycerol is considered a chemical waste, as it is formed as an inevitable by-product during biodiesel formation. Given the highly reduced state of carbon in glycerol this low cost substrate is of special interest for sustainable biofuel production including biohydrogen. The data presented in **chapter 2** for the first time reveal *T. maritima* to sustain growth on glycerol in both batch and chemostat cultivation setups, where glycerol is primarily fermented to acetate, CO_2 and H_2. The observed H_2 yields nearly reach the theoretical maximum of 3 H_2 per glycerol, which is 3 times the yield generally observed for mesophilic conversions, and suggests that the reductant derived from the oxidation of glycerol-3-phosphate is also channelled to H_2. Based on the annotated genome of *T. maritima* a glycerol degradation pathway is proposed. In addition, a comparative genomics study shows that the ability to grow on glycerol can be considered a general trait of *Thermotoga* species.

Elevated H_2 levels are known to inhibit H_2 formation during dark fermentation. In **chapter 3** the strategy of *C. saccharolyticus* to deal with elevated H_2 levels is investigated in glucose-limited chemostats. Presented transcription data reveal the up-regulation of genes involved in the disposal of reducing equivalents under elevated H_2 levels, like lactate dehydrogenase and alcohol dehydrogenase as well as the NADH-dependent and ferredoxin-dependent hydrogenases. These findings are in line with

the observed shift in the fermentation profile from mainly acetate and H_2 to acetate, lactate, ethanol and H_2 under elevated levels of H_2. In addition, differential transcription was observed for genes involved in carbon metabolism, fatty acid biosynthesis and several transport systems. The presented data provides evidence for the involvement of the redox sensing Rex protein, the iron uptake regulator Fur and the fatty acid biosynthesis regulator FapR in gene regulation under cultivation conditions at elevated H_2 concentrations. Different presented cultivation setups were applied to differentiate between the effects of stirring speed, type of reactor flushing and type of flushing gas on the H_2 forming capacity of *C. saccharolyticus*. It was demonstrated that under similar cultivation conditions, when using a high stirring speed, sparging with H_2 gas is almost as efficient as sparging with nitrogen (N_2) gas. This confirmed that the dissolved H_2 concentration is the determining factor in such cultivations, not the H_2 concentration in the gas phase.

Chapter 4 further investigates the effect of increased H_2 levels on the fermentation profile of *C. saccharolyticus* with respect to (i) growth on ammonium (NH_4^+) deficient media and to (ii) low/high glucose loads. Furthermore, the rhamnose fermentation pathway and its associated mechanisms of reductant disposal are discussed. The increase in glucose load during chemostat cultivations under elevated H_2 levels resulted in a clear shift to lactate formation whereas ethanol concentrations were unaffected, thus demonstrating that lactate formation is the main mechanism to alleviate redox stress in *C. saccharolyticus* (**chapter 4**). The omission of NH_4^+ from the cultivation medium resulted in incomplete substrate conversion during chemostat cultivation, when H_2 was allowed to accumulate in the liquid phase. Fermentation profiles of *C. saccharolyticus* grown on rhamnose in the presence of CO, a competitive inhibitor of hydrogenases, suggested electron exchange between reduced ferredoxin and $NAD(P)^+$. However, no gene coding for an enzyme capable of catalysing such reaction could be identified in the genome of *C. saccharolyticus*. The ability of a *Caldicellulosiruptor* species to ferment rhamnose correlates with the genomic presence of two specific gene clusters. Moreover, **chapter 4** provides an extensive review regarding the hydrolytic potential, the sugar metabolism, the H_2 forming capacity, and the mechanism involved in mixed acid fermentation in *C. saccharolyticus*.

A study on the role of pyrophosphate (PP_i) as both an energy carrier and inhibitor of the central metabolism of *C. saccharolyticus* is presented in **chapter 5**. In agreement with the annotated genome sequence, cell extracts were shown to exhibit pyruvate kinase, pyruvate phosphate dikinase, ATP- and PP_i-dependent phosphofructokinase activity. Also a membrane-bound pyrophosphatase activity was observed, while no significant soluble pyrophosphatase activity was detected. The presence of the PP_i-dependent activity of pyruvate phosphate dikinase and PP_i-dependent phosphofructokinase in glucose-grown cultures suggest a catabolic role for these enzymes, which represent parallel pathways with respect to the classical glycolytic

alternatives: pyruvate kinase and ATP-dependent phosphofructokinase, respectively. The clustering of the pyruvate phosphate dikinase gene together with the genes coding for all C3-branch EMP-pathway enzymes (except pyruvate kinase) appears to be an unique feature of the *Caldicellulosiruptor* genus. Investigation of the PP_i dynamics during growth on glucose revealed a relatively high PP_i concentration during exponential growth compared to the PP_i concentration observed at the transition from the exponential to the stationary growth phase. Moreover, the activity of pyruvate kinase, which gene clusters with the gene encoding the ATP-dependent phosphofructokinase, was shown to be inhibited by PP_i. Altogether, these findings support an important role for PP_i in the central carbon and energy metabolism of *C. saccharolyticus*.

The potential involvement of PP_i as phosphoryl-donor in glycolysis can lead to an energetic advantage during cell growth. In literature several reports have appeared describing the involvement of PP_i in glycolysis and gluconeogenesis, i.e at the level of PP_i-dependent phosphofructokinase and pyruvate phosphate dikinase. The comparative genomics study presented in **chapter 6** provides an overview of the genomic distribution of these enzymes and their related enzyme-subtypes over the fully sequenced genomes of 70 archaea, 30 eukarya and 395 bacteria. Additionally, correlations (genomic co-occurrence) of PP_i-dependent phosphofructokinase and pyruvate phosphate dikinase are investigated with respect to (i) each other, (ii) the genomic presence of soluble and membrane-bound pyrophosphatases and (iii) the genomic presence of the classical glycolytic/gluconeogenic enzymes (ATP-dependent phosphofructokinase, fructose 1,6-bisphosphatsse, pyruvate kinase and pyruvate water dikinase). The presented, COG-based, *ab initio* enzyme-subtypes classification, that incorporates characterized protein features (e.g. catalytic site residues and allosteric regulatory site residues), identified several novel enzyme-subtypes alongside all previously reported enzyme-subtypes. Halve of the investigated genomes contain genes coding for the non-classical PP_i-dependent phosphofructokinase and/or pyruvate phosphate dikinase. The observed distribution patterns suggest a correlation between the genomic co-occurrence of the non-classical PP_i-dependent enzymes and the membrane-bound pyrophosphatase. Overall, the data indicate that the involvement of PP_i in glycolysis/gluconeogenesis is a wide spread phenomenon present in all three domains of life.

General discussion

1. Bio-hydrogen by dark fermentation: An integrated production process

Bio-H_2 formation via dark fermentation can be based on renewable resources, thus potentially providing a more sustainable H_2 production process compared to the currently applied processes in which non-renewable fossil resources are used. By itself H_2 formation via dark fermentation is not considered to be a profitable process [386], however, dark fermentation can play a key role in a larger integrated H_2 production process. A simplified overview of a two stage biological H_2 production process involving dark fermentation is presented in **Figure 7.1**.

Sugar-rich biomass substrates, required for fermentation, can be derived from agricultural crops specifically grown for this purpose, like switch grass [43], or from agricultural/industrial waste streams like potato steam peels [47], carrot pulp [45] or molasses, a waste stream from sugar product-plants [186]. If H_2 production facilities are located close to biomass production sites, the transport/storage associated costs are minimized, a concept that is often referred to as decentralized H_2-production [210].

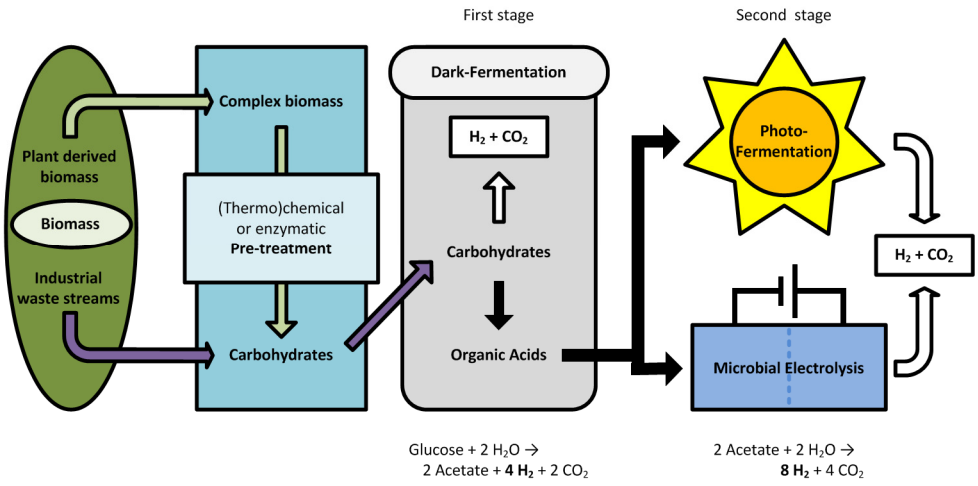


Figure 7.1 Simplified schematic overview of a two stage biohydrogen production system. In the first stage biomass-derived carbohydrates are converted to hydrogen (H_2), CO_2 and organic acids by dark fermentation, in a subsequent second stage, produced organic acids can be converted to H_2 and CO_2 either by photofermentation or microbial electrolysis. Theoretically the complete conversion of glucose to H_2 and CO_2 via a two-stage system can lead to the formation of 12 H_2 per glucose.

Dark fermentation processes primarily use simple soluble carbohydrates, such as di- or mono-saccharides. Generally, when using complex biomass some form of pre-treatment is required that results in the decomposition of the biomass into readily fermentable sugars [145]. This can be achieved by several (thermo)chemical, physical or enzymatic pre-treatment methods. Preferably, biomass hydrolysis is achieved in the same step as fermentation, either in a simultaneous saccharification and fermentation process or a consolidated bioprocess [144, 146]. In the latter process both biomass hydrolysis and sugar fermentation is done by the fermentative microorganism.

During bio-H_2 formation via dark fermentation sugars are ideally converted to H_2, CO_2 and the organic acid acetate. However, depending on the organism, specific metabolic restrictions or process conditions, other more reduced end-products like lactate or ethanol can be formed which reduces the overall H_2 yield of the process. The H_2 yield of dark fermentation is limited to a maximum of 4 H_2 per consumed hexose (e.g. glucose, fructose, or galactose) and is coupled to the concomitant formation of two acetates. The produced organic acid end-products can serve as substrates in a second fermentation process, however, for a complete conversion of the organic acids to H_2 and CO_2 an additional input of energy is required. During photofermentation [211] sunlight provides this energy, while in electrochemically-assisted microbial H_2 production processes [11], the supply of a small electrical potential helps to drive the reaction towards H_2 formation. Overall such a two-stage fermentation process setup, i.e. coupling dark fermentation with either photofermentation or microbial electrolysis, can increase the overall H_2 yield from hexose fermentation to the theoretical limit of 12 H_2 per hexose.

2. Thermophilic bioH_2 production: Are we there yet?

A large European project named HYVOLUTION [210] has been directed at the development of a decentralized H_2 production process, covering the whole production chain from biomass to hydrogen. This project included the integration of biomass logistics and pre-treatment, a thermophilic dark fermentation stage coupled to a photofermentation stage, gas upgrading, as well as cost prediction and social integration studies. One of the HYVOLUTION studies investigated bio-H_2 formation from beet-molasses and reported on currently achievable conversion efficiencies in a realistic two-stage process setup. The thermophilic *Caldicellulosiruptor saccharolyticus* was used for the first stage (dark fermentation) and a *Rhodobacter capsulatus* (hup$^-$) mutant, lacking an uptake hydrogenase gene, for the second stage (photofermentation) [186].

The efficiency of the dark fermentation stage, performed in a controlled batch setup with N_2 sparging, was 53% (2.1 H_2/hexose) with a maximal volumetric

productivity of 7.1 mmol H_2/(L*h), while the efficiency of the photofermentation stage reached 58% (equivalent to 1.7 H_2/acetate), with a maximal volumetric productivity of 1.4 mmol H_2/(L*h). Using this two-stage process setup, H_2 was generated with an overall efficiency of 57% (6.9 H_2/hexose), which is one of the highest reported efficiencies to date [211]. As an alternative to photofermentation, a microbial electrolysis-based process could be used in the second stage. For such bio-electrochemically assisted microbial system overall yields of 2.9 H_2/acetate have been reported, with an energy cost equivalent (energy required for augmentation) of 0.6 H_2/acetate [11].

Table 7.1: H_2 yields and volumetric H_2 production rates from mesophilic, thermophilic, extremely-thermophilc and hyperthermophilic pure cultures grown on glucose (10g/L). ^Abbreviations used for cultivation methods, B: batch; CB: controlled batch.

Organism	Temp. grown (°C)	Cultivation method^	H_2 yield (mol/mol)	Volumetric H_2 production rate (mmol/L/h)	Reference
Mesophiles					
Clostridium tyrobutyricum FYa102	35	CB	1.1	17.4	[388]
Ethanoligenens harbinense YUAN-3	35	CB	1.9	27.5	[389]
Clostridium beijerinckii AM21B	36	B	1.5	26.6	[390]
Klebsiella pneumoniae ECU-15	37	CB	1.7	16.9	[391]
Enterobacter aerogenes strain HO-39	38	B	1.0	12.1	[392]
Clostridium butyricum W5	39	B	0.8	7.6	[393]
Thermophiles					
Thermoanaerobacterium thermosaccharolyticum W16	60	B	2.4	12.9	[394]
Extreme thermophiles					
Caldicellulosiruptor saccharolyticus DSM 8903	72	CB	3.4	12.0	[47]
Thermoanaerobacter mathranii A3N	70	B	2.6	4.5	[395]
Hyperthermophiles					
Thermotoga neapolitana DSM 4359	80	CB	3.5	10.9	[45]

The estimated costs for H_2 production by the two-stage HYVOLUTION process was calculated to be circa € 55- 60 /kg H_2. The major cost factor (80% of the total cost) was the photofermentation, or more specifically the tubular photo-bioreactor [210]. In a techno-economic comparison study between a second generation bio-ethanol production process and a two-stage bio-H_2 production process (dark and photofermentation) with barley straw as feedstock, it was shown that the bio-H_2 process related costs (421.7 €/GJ, equivalent to 51.0 €/kg H_2) were 20 times higher compared to the bio-ethanol production costs (19.5 €/GJ) [197]. Moreover, based on the estimated oil prices in 2020, the cost goal for 'CO_2-free' bio-H_2 in 2020 is estimated to be €1–2 per kg H_2 for merchant (non-fuel) H_2, while the target costs for H_2 as a fuel for fuel cells or internal combustion engines is estimated to be €2.6-5.5 per kg H_2 and €0.6-3.2 per kg H_2, respectively [387].

These data indicate that current two-stage bioH_2 production technologies still lead to a relatively high cost price for H_2 and that improvements are required before such a process becomes economically feasible. Although thermophilic organisms have the potential to ferment sugar-substrates with H_2 yields close to the theoretical maximum [147], in practice high dissolved H_2 concentrations result in lower overall H_2 yields [112, 175]. Moreover, an increase in H_2 productivity is required as well [197, 198, 386]. Some strategies to improve H_2 productivity, while maintaining maximal H_2 yields, for dark fermentation are discussed in the following section.

3. Thermophilic dark fermentation: Room for improvement

Dark fermentation can be performed by pure cultures or by designed microbial consortia [208, 396, 397]. The implementation of consortia would allow for the use of positive features, like synergistic hydrolytic capabilities of different (thermophilic) organisms [34, 35]. Recently, a selection of desirable properties were suggested that fit the ideal H_2-producing organism or consortium member [398]. Such an organisms should, i) be thermophilic, ii) be genetically accessible, iii) possess Fd-dependent hydrogenases (H_2 formation from Fd_{red} is energetically more favourable than H_2 formation from NAD(P)H), iv) be prototrophic for all amino acids, v) be able to degrade a wide range of biomass types, vi) lack any form of carbon catabolite repression (be able to metabolize different sugars simultaneously), vii) possess high tolerance towards osmotic stress (tolerate high substrate loads and soluble end-product concentrations) and viii) be aerotolerant. Among the currently known thermophilic organisms, the species belonging to the genus *Caldicellulosiruptor* (properties i-vi) and *Thermotoga* (properties i-v) come closest to being the ideal H_2-producer [398].

However, relatively low H_2 productivities and end-product inhibition (e.g. by H_2 and organic acids) signify the main challenges to overcome before the commercial

application of thermophilic dark fermentation in a bio-H_2 production process becomes economically feasible [386, 398]. Interestingly, compared to mesophilic H_2 producers, thermophiles typically reveal lower H_2 productivity rates. An overview of reported volumetric productivity values of glucose grown (10 g/L) pure cultures is given in **Table 7.1**. The underlying mechanism of the difference in productivity is unclear; it could be based on differences in H_2 formation mechanisms involved in mesophilic and thermophilic dark fermentation, e.g. hydrogenase sensitivity towards H_2.

Although, H_2 yields reaching the theoretical limit of 4 H_2/hexose have been reported for several thermophiles including *T. maritima* and *C. saccharolyticus*, these yields were mainly achieved under ideal cultivation conditions. Often these conditions included relatively low substrate loads (5g/L hexose or less), while an increase in substrate load generally leads to lower H_2 yields [44, 45, 47, 188]. This decrease in H_2 yield coincides with a switch in fermentation leading to other, more reduced end-products like lactate or ethanol. This phenomenon is proposed to be triggered by the increase in the dissolved H_2 concentration, which occurs when the H_2 productivity rate surpasses the H_2 liquid to gas mass-transfer rate [112, 175]. To overcome this problem and secure high H_2 productivity while maintaining high H_2 yields, the dissolved H_2 concentration should be kept below the threshold value that leads to the fermentation switch. This can be achieved by maximizing the liquid-gas interface area or promoting H_2 gas bubble formation, which are primarily matters of reactor design or operation settings. For continuous stirred tank reactors (CSTR), parameters to be optimized could include reactor dimensions, gas sparging rate, and stirring speed. Induced bubble formation could be achieved by applying a vacuum or via a catalytic surface, e.g. by coating the reactor surface or adding bubble formation promoting particles [204-207]. Alternatively, reactors might be made from gas permeable materials. For continuous systems adjusting the feed inflow can prevent H_2 build-up in the liquid phase.

Higher productivities can also be reached by i) increasing the amount of biomass (specific hydrogen productivity (mmol H_2/($g_{biomass}$*h)) or ii) using higher substrate loads. Although increased biomass or substrate load only leads to an efficient increase in H_2 productivity when low dissolved H_2 concentrations are secured, thus maintaining optimal H_2-yields. These considerations indicate that measuring the dissolved H_2 concentration should become a standard for process monitoring [399]. Contemporary CSTRs (batch and continuous) do not allow for biomass accumulation. Other reactor types might be employed to promote biomass accumulation, leading to a prolonged biomass retention time, however, some of these cultivation methods require the formation of biofilms. Cultivations in an UASB reactor (up-flow anaerobic sludge bed) [400], a trickle bed reactor [196] or membrane bioreactor [401, 402] have been shown to lead to increased biomass retention times while supporting H_2 production by dark fermentation. The use of high substrate loads is inevitably linked

with an increase in osmotic strength of the media which negatively effects growth, potentially leading to the complete inhibition growth [47, 50, 112]. When substrates are converted into organic acids, base addition is required to maintain the desired pH, which is likely to cause a further increase in osmotic strength. In this respect marine organisms like *Thermotoga* species are preferred since they have an intrinsic higher tolerance towards elevated salt concentrations (osmotic strength). During continuous cultivation setups the salt concentration of the influent could be adjusted to compensate for the base addition required for pH control.

Moreover, deviating from the optimum carbon to nitrogen ratio negatively affects the H_2 production and growth [186, 403, 404], indicating that feedstock-specific adjustments of the medium composition is required [186]. Although, cultivation of *C. saccharolyticus* on a defined medium with xylose (4.5 g/L) under N-limiting conditions (with or without C-limitation) resulted in a higher specific H_2 production rate [404], N-limitation might make the system less robust with respect to increased dissolved H_2 concentrations [405].

4. Metabolic engineering: Tweaking metabolic fluxes or introducing non-native pathways

Currently genetic systems are available for the following thermophilic H_2-producers; *Clostridium thermocellum* [406], *Thermoanaerobacterium saccharolyticum* [17, 177], *Thermoanaerobacter ethanolicus* [407], *Thermococcus kodakaraensis* [408, 409], *Pyrococcus furiosus* [410] *Thermoanaerobacter tengcongensis* [411], *Caldicellulosiruptor bescii* [213]. Although, no *C. saccharolyticus* specific system is presently available, a similar approach as used during the development of the system for *C. bescii* [213] is likely to result in a genetic system for *C. saccharolyticus*. Though the transformation of *T. maritima* is possible a specific genetic system is not available [412].

Available genetic systems can serve as tools for the investigation of the physiological function of a single protein or entire (regulatory) networks [17, 242, 413-415], moreover, they can be used to optimize specific metabolic traits, improve the robustness of the system or be used to introduce new metabolic pathways [398, 416, 417]. Improvements of H_2 yield and productivity through metabolic engineering can be separated into two categories based on their focus, (i) engineering of the native pathways, and (ii) introducing non-native pathways leading to H_2. Native pathway engineering focuses mainly on redirecting metabolic fluxes by blocking pathways that compete with H_2 formation, e.g. by inactivating pathways leading to alternative reduced end-products [413, 416, 417]. For example, a lactate dehydrogenase knockout in *C. bescii* resulted in increased acetate and H_2 formation [413].

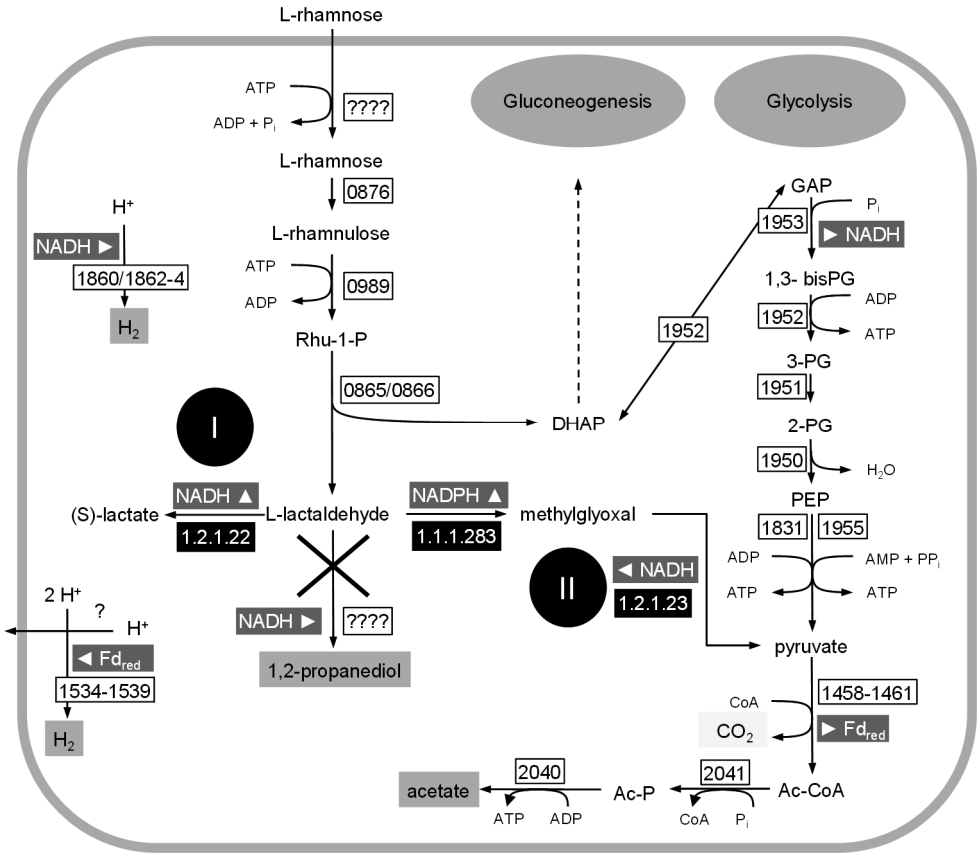


Figure 7.2 Two proposed pathways for improving H_2 yields on rhamnose in *Caldicellulosiruptor saccharolyticus*. A 1,2-propanediol dehydrogenase knockout in combination with (I) an insertion of a L-lactaldehyde dehydrogenase (EC 1.2.1.22) or (II) the insertion of a methylglyoxal reductase (EC 1.1.1.283) and a methylglyoxal oxidase (EC 1.2.1.23) could result in H_2 yields of 3 or 5 H_2 per rhamnose, respectively. Locus tag numbers of the *C. saccharolyticus* genes encoding the corresponding enzymes of the specific catalytic steps are given in square boxes.

However, these changes cannot elevate yields above the network's potential. This can only be done by introducing non-native pathways. In the following section two strategies to improve H_2 yields on rhamnose, based on the *C. saccharolyticus* model organism, are discussed in more detail. A comparable route for rhamnose catabolism is also present in *T. maritima* [418].

Fermentation of rhamnose by *C. saccharolyticus* has been shown to result in the production of 1,2-propanediol, acetate, H_2 and CO_2 in a 1:1:1:1 ratio (**Eq. 7.1**) [40]. During rhamnose fermentation half of the carbon flux enters glycolysis via DHAP (**Figure 7.2**), which is ultimately converted to acetate yielding two reducing

equivalents (NADH and Fd_{red}^{2-}). Half of the generated reduced electron carriers are consumed during the reduction of L-Lactaldehyde to 1,2-propanediol, while the other half is used for H_2 formation, indicating that the H_2 yield is limited to 1 mole H_2 per mole rhamnose consumed [40]. Via metabolic engineering of *C. saccharolyticus* the overall H_2 yield on rhamnose might be enhanced 3 to 5 fold.

The introduction of a L-lactaldehyde dehydrogenase encoding gene (EC 1.1.2.22) [419] combined with a L-lactaldehyde reductase knockout, would result in the formation of lactate from lactaldehyde rather than 1,2-propanediol (**Figure 7.2**, pathway I). Under standard conditions the L-lactaldehyde dehydrogenase-catalysed conversion of L-lactaldehyde to (S)-lactate is associated with a negative Gibbs reaction energy ($\Delta_r G^{\circ\prime}$= -42.81 kJ/reaction, http://www.metacyc.org/). In the suggested pathway no reducing equivalents are required for 1,2-propanediol formation, while the conversion of L-lactaldehyde is accompanied with the additional formation of NADH. Overall this route of rhamnose conversion to acetate and lactate would theoretically lead to a maximal H_2 yield of 3 mole H_2 per mole of rhamnose consumed (**Eq. 7.2**). The iron-containing alcohol dehydrogenase encoded by the gene with locus tag Csac_0407 has been proposed to catalyse the reduction of L-lactaldehyde [40]. Homologues of this enzyme are, however, missing in most *Caldicellulosiruptor* species reported to ferment rhamnose (**Figure 7.3**). This suggests that Csac_0407 might not code for a dedicated L-lactaldehyde reductase, but rather has a more general function. So a knockout of Csac_0407 might not suffice to generate a L-lactaldehyde reductase deficient strain. Analysis of the iron- and zinc-type alcohol dehydrogenases identified in all sequenced *Caldicellulosiruptor* species suggests that the zinc-type alcohol dehydrogenase encoded by Csac_0869 is a more likely candidate for the reduction of L-lactaldehyde (**Figure 7.3**). Homologs of this protein are only present in rhamnose fermenting *Caldicellulosiruptor* species, which can be expected if a dedicated pathway for rhamnose fermentation is present.

(**Eq. 7.1**) $\text{Rhamnose} + 2\ H_2O \rightarrow \text{1,2propanediol} + \text{acetate}^- + HCO_3^- + 2\ H^+ + H_2$
-145.60 kJ/reaction($\Delta_r G^{\circ\prime}$)

(**Eq. 7.2**) $\text{Rhamnose} + 3\ H_2O \rightarrow \text{lactate}^- + \text{acetate}^- + HCO_3^- + 3\ H^+ + 3\ H_2$
-130.5 kJ/reaction($\Delta_r G^{\circ\prime}$)

(**Eq. 7.3**) $\text{Rhamnose} + 5\ H_2O \rightarrow 2\ \text{acetate}^- + 2\ HCO_3^- + 4\ H^+ + 5\ H_2$
-133.1 kJ/reaction($\Delta_r G^{\circ\prime}$)

An alternative pathway for rhamnose fermentation leading to the formation of two acetate could lead to a theoretical yield of 5 H_2/rhamnose (**Figure 7.2**, pathway II, **Eq. 7.3**). This pathway involves the introduction of a methylglyoxal reductase (EC

1.1.1.283, [420]) and a methylglyoxal oxidase (1.2.1.23, [421]) in combination with a knockout of the L-lactaldehyde reductase. The overall conversion reaction is associated with a negative Gibbs energy (**Eq. 7.3**), also the oxidation of methylglyoxal to pyruvate is an exergonic reaction ($\Delta_r G^{\circ\prime}$= -41.88 kJ/reaction, http://www.metacyc.org/). However, under standard conditions, the oxidation of L-lactaldehyde to methylglyoxal is associated with a positive Gibbs energy ($\Delta_r G^{\circ\prime}$= 20.04 kJ/reaction, http://www.metacyc.org/), indicating that the a methylglyoxal reductase-catalysed reaction might form a thermodynamic barrier. However, if carbon fluxes toward acetate and the removal of NADPH are secured, this conversion might be pulled towards methylglyoxal formation. The latter might be problematic since the hydrogenases of *C. saccharolyticus* are not able to use NADPH for H_2 formation. This problem might be solved by the additional insertion of a cytosolic NADPH-dependent hydrogenase, like the cytosolic [NiFe]-hydrogenase identified in archaeal species [408, 422]. Both suggested pathways to enhance H_2 yields during rhamnose fermentation significantly affect the redox pool, resulting in different ratios of reduced and oxidized electron carriers compared to the default situation. Whether the organism is able to handle the suggested changes in metabolism remains to be investigated.

Another strategy to improve H_2 yields could be based on decreasing the overall energy yield (ATP yield) of the carbon metabolism. Both ATP- and H_2-formation are coupled to the carbon metabolism. During carbon metabolism a part of the carbon flux ends up in biomass, the remaining carbon flux is used for ATP synthesis, which is required as an energy input for the biosynthesis reactions. Carbon fluxes towards biomass reduce the maximal achievable H_2-yield. Theoretically, if glucose is completely converted to 2 acetate, 4 H_2 and 2 CO_2, 3 ATP are produced (ATP investment during transmembrane glucose transport included). However, when assuming that ten percent of the total carbon consumed ends up in the biomass the maximal achievable H_2 yield would be roughly 3.6 H_2 per glucose. When ATP generation would be made inefficient, e.g. by lowering the ATP yields from carbon metabolism, a relative larger part of the carbon flux would be required for ATP generation, thus changing the ratio between carbon flux for biomass and carbon flux for ATP generation consequently leading to increased H_2 yields. Moreover, when it is possible for the organism to maintain its growth rate, such a strategy also contributes to an increase in the H_2 production rate.

Some archaeal species possess an alternative NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN [238]), that catalyses the phosphate-independent, single-step oxidation of glyceraldehyde-3-phosphate to 3-phosphoglycerate. The introduction of GAPN combined with a knock-out of both the native GAPDH and phosphoglycerate kinase, would reduce the ATP yield while maintaining NADH as electron carrier.

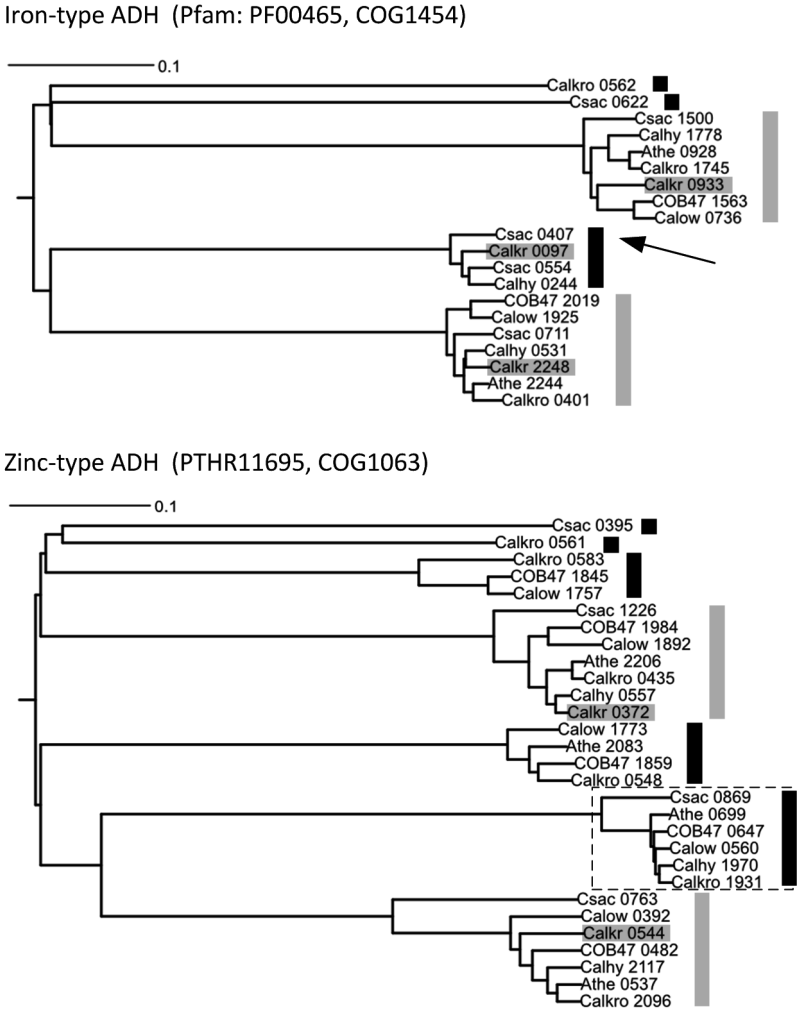


Figure 7.3: Phylogenetic trees of the iron-type (**top panel**) and zinc-type (**bottom panel**) alcohol dehydrogenases (ADH) protein sequences identified in the fully sequenced genomes of seven *Caldicellulosiruptor* species. Phylogenetic trees are arbitrarily rooted. Clusters of homologs containing sequences from all seven *Caldicellulosiruptor* species are indicated by grey bars. Clusters of homologs not containing sequences from all seven *Caldicellulosiruptor* species are indicated by black bars. Within clusters, sequences share a sequence identity >73%, sequence identities between sequences from different clusters are <40%. Zinc-type alcohol dehydrogenase genes present in the putative 'rhamnose gene cluster' are boxed. The locus tags of the ADHs of *Caldicellulosiruptor kristjanssonii*, which is unable to ferment rhamnose, are highlighted in grey. The proposed lactaldehyde dehydrogenase gene (Csac_0407) is indicated by the arrow.

This strategy does not alter the NADH:FD_{red} ratio associated with an unaltered metabolism in for example in *C. saccharolyticus* or *T. maritima*. Moreover, the introduction of an archaeal glyceraldehyde-3-phosphate: ferredoxin oxidoreductase (GAPOR [238]) would not only result in a decrease in overall ATP yield but also exchange NADH for Fd_{red} as electron carrier, which is an even better electron donor for H_2 formation.

Alternatively, when the glucose carbon flux is sent through the pentose phosphate pathway (PPP) and all formed fructose 6-phosphate is forced to re-enter PPP again via glucose 6-phosphate, 1 mol of glucose could theoretically be converted to 1 mol of acetate, 4 mol of CO_2, and H_2 with a yield of 8 H_2/glucose (**Eq. 7.4**) [45]. This would also require the introduction of an NADPH-NADH transhydrogenase or an NADPH:ferredoxin oxidoreductase, allowing the formation of H_2 from the NADPH generated during the oxidative part of the PPP. Also the ATP yield per glucose would be lower (1 ATP/glucose) compared to fermentation via the EMP pathway (3 ATP/glucose). An *in silico* metabolic model of *T. maritima* predicted the introduction of such a pathway to be feasible [423], however, this strategy still awaits experimental confirmation.

(**Eq. 7.4**) $$\text{Glucose} + 8\ H_2O \rightarrow\ \text{acetate}^- + 4\ HCO_3^- + 5\ H^+ + 8\ H_2$$

Also the glucose conversion to acetate via the Entner–Doudoroff pathway is less energy efficient compared to the Embden-Meyerhof-Parnas pathway. While both pathways generate the same amount of electron carriers (1 NADPH, 1NADH and 2 Fd_{red} for the ED pathway and 2 NADH and 2 Fd_{red} for the EMP pathway), the ED-pathway generates 1 ATP/glucose versus 3 ATP/glucose for the EMP pathway. Redirecting fluxes primarily through the ED- pathway could potentially lead to an increased H_2 yield.

Since both *C. saccharolyticus* and *T. maritima* have desirable hydrolytic capabilities [34, 35] and a relative simple reductant metabolism [40, 53], these organisms could also provide a good starting point for metabolic engineering focused on the formation of reduced end-products other than H_2. Disruption of the acetate forming pathway is likely to result in a well performing lactate producing organism, while the introduction of a bifunctional acetaldehyde-CoA/alcohol dehydrogenase [173], in combination with a [FeFe] hydrogenase knockout may provide a dedicated ethanol forming pathway.

To adequately design a novel metabolic route to a desired end-product, a detailed comprehension of the metabolic processes is of utmost importance. More specifically, the information of required enzyme co-factors, e.g. specificity of electron carriers or phosphoryl donors, and the identification of potential thermodynamic barriers associated with specific catalytic conversions are some key factors in successful

engineering an organism. To be able to exploit the introduction of non-native pathways one must safeguard the organism's energy supply, carbon flow and reductant recycling mechanisms.

5. On the physiological role of the *C. saccharolyticus* and *T. maritima* hydrogenases

Hydrogenases are key enzymes in bio-H_2 formation via dark fermentation. These enzymes couple the re-oxidation of reduced electron carriers, generated throughout metabolism, with proton reduction, subsequently leading to H_2 formation. Hydrogenases can be classified into three distinct groups based on the metal composition of their catalytic site [131]. The [NiFe] hydrogenases and the [FeFe] hydrogenases typically contain one or more iron-sulfur clusters ([2Fe2S] and/or [4Fe4S]) [131, 424, 425]. The [Fe] hydrogenases, which are only found in methanogens do not harbour iron-sulfur clusters, but alternatively incorporate an iron-guanylylpyridinol cofactor in the active site [131, 426]. Further subdivisions of the [NiFe] and [FeFe] hydrogenases can be made on the basis of domain organisation and subunit composition and characteristics [424, 425, 427].

The *C. saccharolyticus* genome contains genes coding for a Hyd-type [FeFe] and a Ech-type [NiFe] hydrogenase. These hydrogenases have not been biochemically characterized, however, the genomic organisation of the genes encoding the Hyd and Ech hydrogenase subunits in *C. saccharolyticus* resembles those of *Caldanaerobacter subterraneus* subsp. *tengcongensis* [40]. Moreover, the domain organisation of the hydrogenase subunits are identical and their associated amino acid sequences share a high level of identity, suggesting a similar function.

The *C. subterraneus* multimeric Ech hydrogenase, isolated from the membrane fraction of glucose grown cells, was demonstrated to be able to catalyse both H_2 evolution with Fd_{red} as electron donor and its reverse reaction i.e. reduction of ferredoxin by H_2 [22]. The six subunit membrane-bound Ech hydrogenase from the methanogenic archaeon *Methanosarcina barkeri* [428], which resembles the *C. subterraneus* Ech hydrogenase [22], is described to function as a ferredoxin-linked H_2 uptake hydrogenase during acetoclastic methanogenesis and as ferredoxin-linked H_2 forming hydrogenase under autotrophic growth conditions, where the unfavourable reduction of Fd_{red} by H_2 is proposed to be driven by reversed electron transport. These findings were supported by Ech hydrogenase knockout studies, highlighting the dual function of this [NiFe] hydrogenase in *M. barkeri* [428, 429].

The tetrameric Hyd hydrogenase from *C. subterraneus*, isolated from the soluble fraction of glucose grown cells, displayed NADH-dependent H_2 formation. No activity was detected on NADP(H) [22]. Alternatively, Schuchmann *et al.* demonstrated that the tetrameric Hyd [FeFe] hydrogenase of *Acetobacterium woodii* couples the

endergonic reduction of ferredoxin (E_o' =∼500mV) and the exergonic reduction of NAD^+ (E_o' = -320mV) with H_2 as electron donor (E_o' = -414mV) [430] in a process called electron-bifurcation (the 4 electrons derived from H_2 end up on NADH and Fd_{red}^{2-} in a 1:1 ratio, see **Figure 7.4 C**). Reduction of either NAD^+ or Fd_{ox} could not be detected in the absence of the other electron carrier, thus signifying the strict nature of this observed electron-bifurcation. The genomic and domain organisation of the *A. woodii* Hyd hydrogenase highly resembles the genomic and domain organisation of the Hyd hydrogenases from *C. subterraneus* and *C. saccharolyticus* (**Figure 7.4 A and B**).

The HydD subunit forms a part of the complex [22, 430], however, this subunit does not contain any catalytic sites involved in electron transport and is proposed to have a role in assembly or stability of the HydABCD complex. The multimeric Hyd hydrogenase from *A. woodii* is present in fructose grown cells, but protein levels showed a 3-fold up-regulation in H_2/CO_2-grown cells, indicating an important role for this enzyme in H_2 oxidation (uptake) during CO_2 fixation via the Wood-Ljungdahl pathway [431]. The H_2-evolving capability of cell lysates from carbohydrate grown *T. maritima* cells were ascribed to the activity of a single hydrogenase [30]. This trimeric [FeFe] hydrogenase (HydABC, **Figure 7.4**) is able to utilize both Fd_{red} or NADH during the catalysis of H_2 formation and is proposed to exploit the exergonic oxidation of Fd_{red}^{2-} (E_m =-453mV) to drive the endergonic oxidation of NADH (E_o' = -320mV) forming H_2 (E_o' = -420mV) [30], thus acting as an electron-confurcating hydrogenase (the electrons donated by Fd_{red} and NADH end up on H_2, using a 1:1 ratio of Fd_{red} and NADH **Figure 7.4 D**). The [FeFe] Hyd hydrogenase from *Moorella thermoacetica*, which has the same trimeric subunit structure as the *T. maritima* Hyd hydrogenase (**Figure 7.4 A and B**), is characterized to be able to catalyse both the electron-bifurcating as the electron confurcatinghydrogenase reaction [432]. Furthermore, *M. thermoacetica* HydABC activity is observed in extracts from both glucose and H_2/CO_2-grown cells, supporting the reversibility of H_2 formation via this hydrogenase [433]. The presented models for electron-bifurcation/confurcation via Hyd hydrogenase activity from *A. woodii* and *T. maritima*, respectively, mainly differ with respect to their proposed ferredoxin binding site (**Figure 7.4 C and D**).

Based on these biochemical characteristics it can be argued that the tetrameric Hyd hydrogenase from *C. saccharolyticus* and *C. subterraneus* may also function as electron confurcating/bifurcation hydrogenase. Although, the absolute requirement of electron-confurcation as a mechanism for H_2 formation by these H_2 producers, including *T. maritima*, might depend on the physiological condition. In nature H_2 consumption by, for example, methanogens can keep H_2 concentration relatively low [434], such low H_2 levels could even make proton reduction by NADH an exergonic process [147].

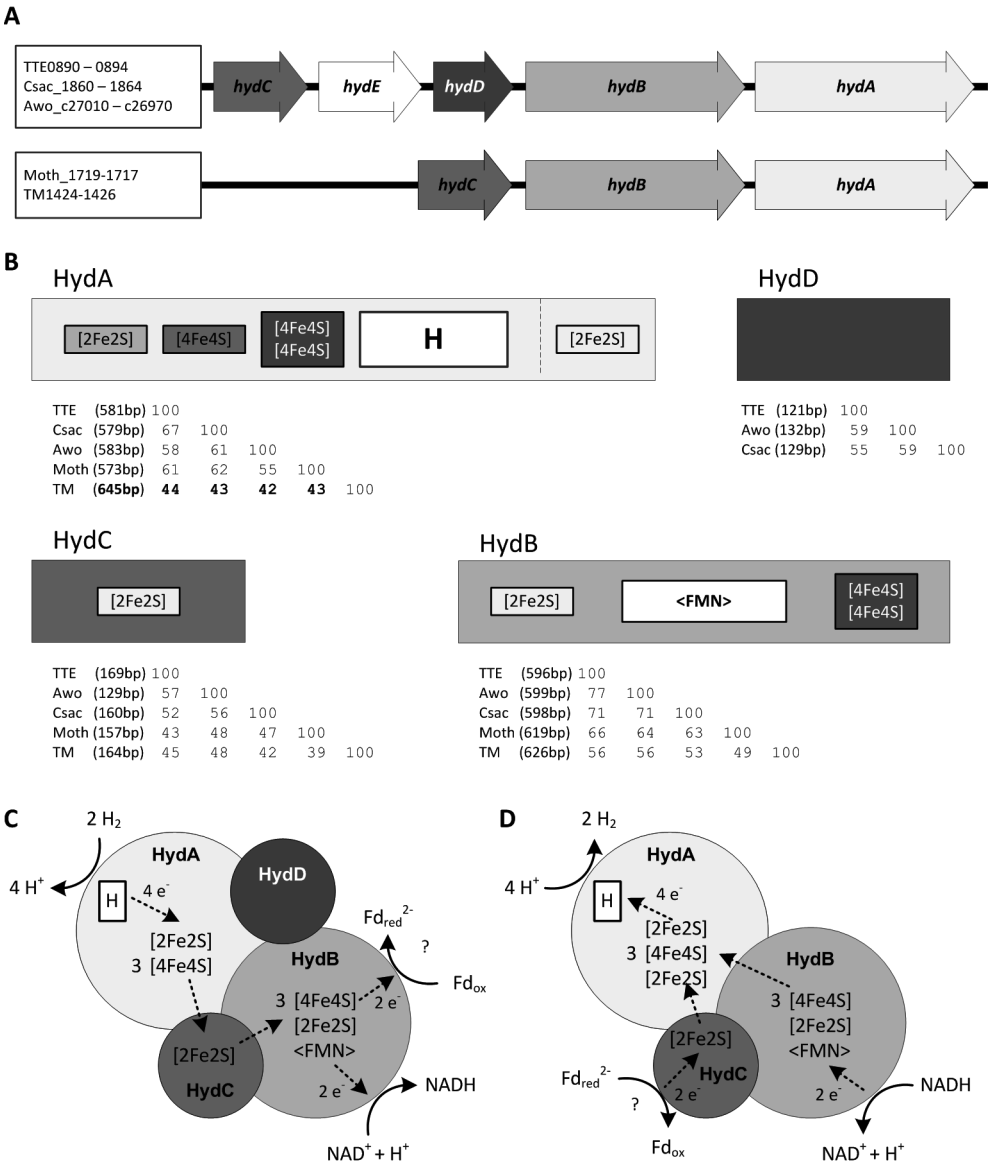


Figure 7.4: **A**) Genomic organisation of the multimeric [FeFe]-hydrogenases (hyd) from *Caldanaerobacter subterraneus* subsp. *tengcongensis* sp. MB4 (TTE), *Caldicellulosiruptor saccharolyticus* DSM 8903 (Csac), *Acetobacterium woodii* DSM 1030 (Awo), *Moorella thermoacetica* DSM 521 (Moth) and *Thermotoga maritima* MSB8 (TM); gene associated locus tags are given. **B**) Domain organisation of the multimeric [FeFe]-hydrogenase subunits of the indicated organisms. [2Fe2S], [2Fe2S] cluster binding sites; [4Fe4S], [4Fe4S] cluster binding sites; H, H-cluster; <FMN>, NAD-, FMN-, and 2Fe-2S cluster binding site; adapted from Soboh *et al.* [22]. The HydA C-terminal [2Fe2S] domain is only present in *T. maritima* HydA subunit. Subunits size and sequence identities are given (sequence alignment were performed in ClustalX 2.1). **C**)

Proposed model of the electron-bifurcating hydrogenase from *A. woodii*, The electrons donated by H_2 are finally accepted by NAD^+ and oxidized ferredoxin (Fd_{ox}); adapted from Schuchmann *et al.* [430]. **D**) Proposed model of the electron-confurcating hydrogenase activity from *T. maritima*, The electrons donated by NADH and reduced ferredoxin (Fd_{red}) are finally accepted by protons thus evolving H_2; adapted from Schut *et al.* [30].

The proposed non-mandatory electron-confurcating nature of the Hyd hydrogenase activity can be supported by two observations, (i) a sole NADH dependency observed for the Hyd hydrogenase of *C. subterraneus* [22], and (ii) the *T. maritima* Hyd hydrogenase displayed moderate H_2 evolving activity with only Fd_{red} as electron donor [30]. However, electron-confurcation is probably the preferred mechanism, clearly giving an energetic advantage under high H_2 levels.

The HydABC hydrogenase is the only candidate in *T. maritima* to be involved in the recycling of both NADH and Fd_{red} generated during heterotrophic growth [30]. Alternatively, the Ech hydrogenase of *C. saccharolyticus* and *C. subterraneus* could be involved in Fd_{red} recycling [22, 40]. Fermentation experiments on *C. subterraneus* grown in a sparged cultivation system (no H_2 build up) and a closed system (no sparging, H_2 was allowed to accumulate) showed that growth in a closed system was accompanied by a fermentation switch from mainly acetate to a mix of acetate and ethanol. Interestingly, the specific activity of the Ech hydrogenase was not influenced by the growth conditions, while the specific activity of Hyd hydrogenase in the closed system was fourfold lower compared to the sparged system [22]. Additionally, for *C. saccharolyticus* an increase in transcript level was observed under increased glycolytic flux (glucose and/or xylose versus rhamnose) for of all the genes associated with the C-3 of glycolysis, including the pyruvate:ferredoxin oxidoreductase which is responsible for Fd_{red} generation, and the Hyd hydrogenase, while no differential expression of the Ech-hydrogenase was observed [40]. So for both *C. subterraneus* and *C. saccharolyticus* the involvement of Ech hydrogenase for ferredoxin recycling generated during glycolysis can be debated. Together the two examples seem to indicate that for these organisms Ech-hydrogenase expression/activity is not coupled to the carbon flux to acetate and might not be involved in the regeneration of the catabolism derived ferredoxin.

Interestingly, carbon sources with an oxidation state that differs from that of glucose (e.g. glycerol) may result in NADH/ferredoxin ratio's other than 1. If all reductant is channelled to H_2 and electron-confurcation is the sole mechanism for hydrogen formation such a deviant NADH/Fd_{red} ratio must require additional reactions to sustain H_2 formation. For *T. maritima* the involvement of the ion-translocating Fd:NADH oxidoreductase (Rnf [106]) is opted to uphold the 1:1 ratio of NADH and Fd_{red} [30, 435]. So far no Fd:NADH oxidoreductase encoding genes were identified for *C. saccharolyticus*, leaving the exact mechanism of electron transfer between NADH and ferredoxin an open question. Moreover, growth on rhamnose also

drastically deranges the 1:1 ratio of NADH and Fd_{red}. Half of the generated reductant pool is consumed during the NADH-dependent formation of 1,2 propanediol from lactaldehyde leaving mainly Fd_{red} to be recycled as H_2. Alternatively, a putative alpha-keto acid dehydrogenase complex, whose genes are part of a rhamnose associated gene cluster [405], may catalyse the conversion of pyruvate to acetyl-CoA while generating NADH. To uphold sufficient reductive power for the lactaldehyde reductase activity and a 1:1 ratio of NADH and Fd_{red} for the electron-confurcating hydrogenase activity, half of the carbon flux from pyruvate to acetyl-CoA must go via the pyruvate dehydrogenase dependent route while the other half should go via the pyruvate:ferredoxin oxidoreductase complex.

To confirm the involvement of electron-confurcation as a mechanism for H_2 formation in *C. saccharolyticus* its Hyd hydrogenase should be biochemically characterized. Furthermore, the involvement of Ech-hydrogenase activity during heterotrophic growth and the participation of the putative alpha-keto acid dehydrogenase in rhamnose metabolism should be investigated by gene knockout studies.

6. Energy metabolism of *C. saccharolyticus* and *T. maritima*

The EMP pathway is the main route of glycolysis in both *C. saccharolyticus* and *T. maritima* [23, 40, 48, 53]. During the conversion of glucose to pyruvate 1 NADH is generated per pyruvate (**Figure 7.5**). The subsequent conversion of pyruvate to acetate is associated with the formation 1 Fd_{red} (2 electrons) per acetate. To maintain catabolic fluxes through this pathway all reduced electron carriers need to be recycled. This reaction is catalysed by the cytosolic [FeFe] electron-confurcating hydrogenase, consuming NADH and Fd_{red} in a 1:1 ratio as discussed above. Some of the carbon is lost for biosynthesis, e.g. from pyruvate to valine or isoleucine, most likely leading to a lower Fd_{red} yield compared to the NADH yield. To restore the required ratio for hydrogenase activity electrons have to be transferred from NADH to Fd_{red}, which is an energy consuming process. In *T. maritima* this reaction can be catalysed by an ion-translocating Fd:NADH oxidoreductase at the expense of a Na^+ gradient (Rnf-cluster). In *C. saccharolyticus* the membrane bound [NiFe] hydrogenase could, at the expense of a proton gradient, catalyse the reduction of ferredoxin with H_2 as electron donor.

In *T. maritima* small fluxes through the pentose phosphate pathway (PPP) and/or the ED pathway can provide the NADPH required for biosynthesis (**Figure 7.5**). *C. saccharolyticus* lacks the possibility to produce NADPH via these pathways, only the non-oxidative part of the PPP is present in the *C. saccharolyticus* genome. The origin of NADPH in *C. saccharolyticus* is still unknown. It could be provided by the citric acid cycle (oxidative direction) via an NADPH-producing isocitrate dehydrogenase as

proposed by van de Werken *et al.* [40]. Alternatively, NADPH could be generated via an NADH-dependent ferredoxin:$NADP^+$ oxidoreductase (NfnAB) reaction. However, the genes coding for such enzyme can be identified in all genomes of the sequenced *Caldicellulosiruptor* species except in *C. saccharolyticus*.

T. maritima is able to grow and ferment glycerol while achieving high H_2 yields. Glycerol is expected to enter glycolysis via DHAP via the intermediate glycerol 3P, as proposed by Maru *et al.* [435] (**Figure 7.5**). The observed yield of 2.5-3.0 H_2/acetate suggests that the electrons derived from the oxidation of glycerol-3P end up in H_2. However, the exact mechanism of electron transfer from glycerol 3-P to H_2 is still unclear. NADH is proposed to serve as an intermediate electron carrier [435], which requires the uphill reduction of ferredoxin by NADH to reach the 1:1 NADH/ferredoxin ratio required for H_2 formation. Based on comparative genomics, it was expected that *C. saccharolyticus* was also able to catabolise glycerol. Growth experiments on glycerol and co-fermentation with glucose could not support this prediction, since no glycerol consumption was observed in either experiments. The reason for this is unclear. *T. maritima* contains two glycerol kinases; one gene (TM1430) is located in the glycerol 'operon' alongside the three subunit glycerol-3P dehydrogenase encoding genes while the other kinase gene (TM0952) is located elsewhere in the genome. *C. saccharolyticus* contains only one glycerol kinase gene which is not located next to the three subunit glycerol-3P dehydrogenase genes. Disruption of transcriptional regulation of the glycerol degradation pathway associated genes might be the reason for the inability of *C. saccharolyticus* to grow on glycerol.

The conversion of fructose 6P to fructose 1,6bisP (**Figure 7.5**) in *C. saccharolyticus* can be catalysed by group B1 ATP-dependent phospofructokinase (ATP-PFK) or a group B2 pyrophosphate-dependent PFK (PP_i-PFK). The latter enzyme is reversible and may also have a gluconeogenic function since no known fructose bisphosphatase (FBPase), i.e. class I-V FBPase, is encoded in the genome. During growth on rhamnose the formation of fructose 6P from fructose-1,6-bisphosphate is required for biosynthesis, for example the important precursor 5-phosphoribosyl diphosphate, required for *de novo* synthesis of nucleic acids, is formed via the non-oxidative part of the pentose phosphate pathway from both fructose 6P and GAP. Since the *C. saccharolyticus* genome does not have a FBPase, a gluconeogenic function for the PP_i-dependent PFK is likely. *T. maritima* also has a group B1 ATP-PFK while its PP_i-PFK belongs to the SHORT group of PFKs. Although *T. maritima* does not have a dedicated FBPase, i.e. FBPase belonging to class I,II, III an V, its genome does contain a class IV FBPase. TM1415 (annotated as 'putative IMPase') exhibited high FBPase activity, however, it also displays IMPase activity. Moreover, it also showed activity on p-nitrophenyl phosphate and glucose-1-phosphate [436]. So one could argue whether TM1415 actually functions as a FBPase under physiological conditions.

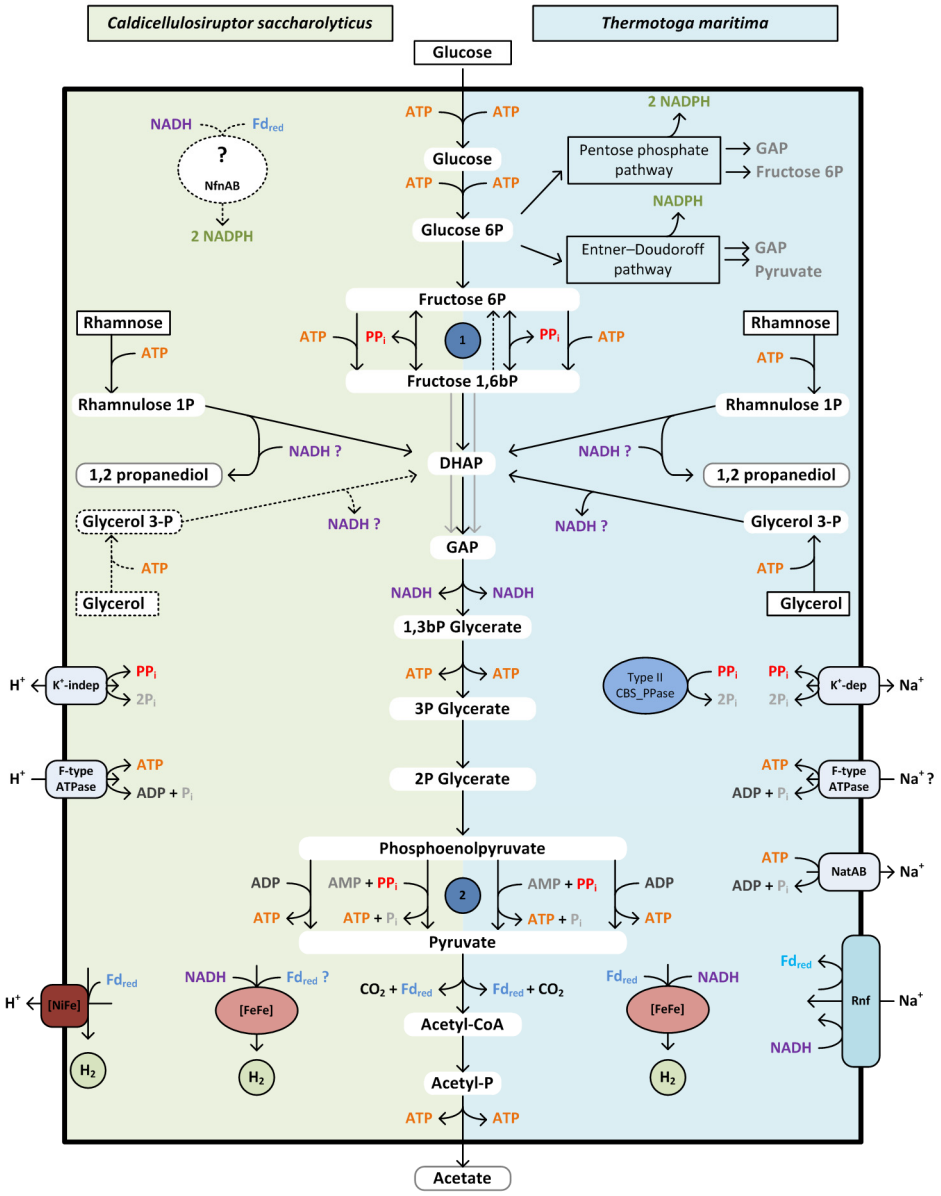


Figure 7.5 Schematic representation of the main catabolic route of glucose conversion to acetate and CO_2 in *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima*. Parallel pathways are present at the (**1**) Fructose-6P to Fructose-1,6bP and (**2**) phosphoenolpyruvate to pyruvate interconversion nodes. *Caldicellulosiruptor* species only have the non-oxidative part of the pentose phosphate pathway. The ED pathway is absent in *Caldicellulosiruptor* species. Differences with respect to reductant recycling and transmembrane transport associated energy coupling are given. Both organisms have a similar pathway for rhamnose catabolism. Glycerol catabolism has not been observed for *C. saccharolyticus* although genes

encoding the potential conversion pathway are present. A NADH dependent ferredoxin:NADP+ oxidoreductase (NfnAB) has been identified in all sequenced *Caldicellulosiruptor* species except for *C. saccharolyticus*.

If so, this would suggest an apparent redundancy with respect to the PP_i-PFK; if not, PP_i-PFK could support the gluconeogenic reaction. An involvement of *T. maritima* PP_i-PFK in gluconeogenesis, is supported by the observation that the ATP-PFK of *T. maritima* is inhibited by PP_i [284]. This feature is also observed for the group X ATP-PFK of *Amycolatopsis methanolica* [234]. Pyrophosphate resembles the phosphate moiety of ATP and ADP, which could be the reason for a form of competitive inhibition. Whether ATP-PFK inhibition by PP_i is a general feature of all ATP-dependent PFKs belonging to PFK family A is not known, since inhibition by PP_i is generally not tested in ATP-PFK characterisation studies.

The conversion of phosphoenolpyruvate to pyruvate in *C. saccharolyticus* and *T. maritima* can be by both catalysed by pyruvate kinase (PK) and/or pyruvate, phosphate dikinase (PPDK) (**Figure 7.5**). The PPDK encoding gene in *C. saccharolyticus* is located in a gene cluster that includes all enzymes involved in the C3 part of glycolysis, excluding PK. The PK gene is located next to the ATP-PFK encoding gene. PPDK transcript levels were high under increased glycolytic flux (glucose versus rhamnose) [40]. In *T. maritima* the PK gene is also located alongside the ATP-PFK, while the PPDK encoding gene is clustered together with fructose-bisphosphate aldolase and acetate kinase. Together these findings support the potential involvement of PPDK in glycolysis in both these organisms. The involvement of PPDK in glycolysis would lead to an energetic advantage, since the energy from PP_i is used to phosphorylate AMP, while the investment of ATP is required to phosphorylate AMP to ADP in an adenylate kinase catalysed reaction. Although PP_i generated from biosynthesis alone is not sufficient to completely support the carbon fluxes through PP_i-PFK and/or PPDK, only supporting a part of the flux would already lead to an energetic advantage. Knockout studies can be useful to investigate the physiological function of both PP_i-PFK or PPDK. Such studies should include low energy conditions, e.g. during lactate formation, because under those conditions the potential energetic advantage associated with the use of PP_i as a phosphoryl donor is more pronounced.

Many biosynthesis reactions e.g. RNA, DNA and protein formation, lead to the formation of PP_i. The removal of PP_i is required for these biosynthesis reactions to proceed. *C. saccharolyticus* lacks a soluble pyrophosphatase (PPase) (**Figure 7.5**), but has a gene coding for a H^+-translocating membrane-integral pyrophosphatase (M-PPases). The M-PPases generate an electrochemical gradient upon hydrolysis of PP_i, thus allowing for the conservation of the energy stored in the phosphate bond of PP_i. Alternatively, PP_i could be used as a phosphoryl donor in PP_i-dependent conversions in the glycolysis as discussed above. One way or the other, the energy from PP_i is

conserved and does not dissipate as heat, what happens when PP_i is hydrolysed by a soluble pyrophosphatase. The *T. maritima* genome codes for a Na^+-translocating (M-PPases) and a soluble type II CBS_PPase. When both PPases would be active simultaneously, the coupling of their reaction could cause a dissipation of the energy stored in the Na^+-gradient, which would be energetically unfavourable. Activity of the type II soluble CBS-PPase of *Moorella thermoacetica* is inhibited at low energy charge, with AMP and ADP acting as strong inhibitors while ATP stimulates activity [367]. This mode of regulation might also be present in *T. maritima*, thereby preventing the futile cycle associated with its soluble PPase and M-PPase. Moreover, it suggests that M-PPase activity is associated with a low energy state of the organism. Thus, PP_i is used as energy source for the generation of a Na^+-gradient under low energy conditions, while PP_i is simply hydrolysed without energy conservation during a high energy state. In both *C. saccharolyticus* and *T. maritima*, the ion translocating M-PPase may work in concert with ATP-synthase to scavenge energy from biosynthetic waste (PP_i), especially under energy-limiting conditions. Experimental studies focused on the specific inhibition of the M-PPase during low energy cultivation conditions, could provide useful insights with respect to the involvement of this enzyme in energy conservation.

The genomic presence of M-PPase has been proposed to contribute to cell survival under low energy conditions and abiotic stress [382, 384]. This hypothesis is supported by the observation that the introduction of M-PPase in eukaryotes and prokaryotes, which lack a M-PPase encoding gene, confers a higher tolerance toward abiotic stress [382, 384]. The presence of a 'PP_i -dependent' glycolysis was originally proposed to be a specific feature of some anaerobic organisms [181, 437], however, the identification of PP_i-PFKs and PPDKs in aerobic organisms debunked this proposition. Moreover, the genomic distribution of these two PP_i-dependent enzymes (Chapter 6) showed that the presence of such enzymes are not linked to a specific subset of organisms but is widely distributed over all domains of life. Moreover, there appears to be a pattern with respect to the mutual genomic presence of M-PPase and PP_i-PFK/PPDK, suggesting a linkage between both enzymes, possibly reflecting an energy-efficient type of metabolism. These findings suggest that the involvement of PP_i in energy metabolism should be seen as a commonality instead of an organism specific singular phenomenon. In addition, the investigation of the regulatory role of PP_i in glycolysis, reported for several organisms [107, 229, 234, 284, 385], is still underexposed.

7. Genome scale models

Since annotated genomes are available for several thermophilic H_2-producers, the road to reconstructing biochemical networks is open [438]. When combining genome

scale models with high throughput data sets derived from proteome, transcriptome and metabolome analysis [439], a deeper understanding of the system as a whole can be achieved (*Systems Biology*). Flux balance analysis, although limited in regulatory and kinetic details, provides a useful tool in analysing steady state fluxes through metabolic networks [440, 441]. The solution space of a Flux balance analysis model, defined by the investigated biochemical network and constrained by the stoichiometry of the network's reactions, represents the feasible network fluxes. Additional physiochemical constraints further restrict these possible fluxes, thus defining the organisms network capability. By optimizing the steady-state performance, with respect to a defined objective function, like maximizing H_2 yield, the total system's behaviour can be investigated. In general, constructed genome scale models could be used as drawing boards for the design of new non-native pathways in existing H_2 producers or for developing newly constructed systems based on genetically accessible organisms (*Synthetic Biology*). Additionally, computer models allow *in silico* testing prior to experimental work and can be used to reveal possible bottlenecks in the central metabolism leading to hydrogen.

Genome scale models of both *C. saccharolyticus* [442] and *T. maritima* [423] have been made. Based on the construction of the *C. saccharolyticus* model, the lack of amino acid auxotrophy was predicted, which could be experimentally confirmed [194]. Furthermore, the predicted flux distributions by the model were comparable with the experimentally-measured fluxes in a glucose grown chemostat [442]. The *T. maritima* model was used to i) 'visualize' the H_2 metabolism in this organism, ii) identify suitable carbon sources for enhancing H_2 production, and iii) to design knockout strains, which increased the *in silico* H_2 production up to 20% [423]. Unfortunately, the absence of a genetic system for *T. maritima* does not allow an experimental verification of the proposed improvements.

Though a genome scale model can be useful in predicting both general and specific characteristics of a system, they are defined by their creator. When their construction is based on an automated genome annotation flaws in the network might be introduced, like enzyme electron-carrier specificity or required phosphoryl donor. Without experimental verification, model prediction should be looked upon with suspicion. For example, flux balance analysis performed on the *E.coli* iJR904 model with an introduced non-native NAD(P)H:H_2 pathway predicted increased H_2 yields, surpassing the enteric-type fermentation limit of 2 moles H_2/glucose [443]. However, when an almost identical NAD(P)H pathway was introduced in *E.coli* the H_2 yields were substantially lower [444].

8. Bio-H_2: an *in vitro* systems approach

Enzymes from thermophiles display extreme (thermo)stability and a wide range of these enzymes have been examined to explore their potential for various biotechnological processes [445, 446]. One of their potential applications can be in designed *in vitro* systems for H_2 formation, which could increase the total energy extraction (H_2 yield) of the overall substrate conversion. This involves building up an entire H_2 producing pathway in a reaction vessel, while making use of the large thermophilic enzyme arsenal available. A simple two-component system demonstrated the proof of principle of *in vitro* H_2 production [447]. By combining the *Thermoplasma acidophilum* glucose dehydrogenase and a NADPH-dependent hydrogenase from *Pyrococcus furiosus*, glucose could be converted to glucuronic acid and H_2, concomitant with the recycling of the co-factor. This research was expanded to include a broader range of substrates including different renewable resources [448]. *In vitro* system research was taken to a next level when the enzymes of the oxidative pentose phosphate cycle were combined with the NADPH- dependent sulfhydrogenase of *P. furiosus* [448]. In this system each mole of glucose-6-phosphate could be converted into 11.6 moles of H_2. The *in vitro* H_2 production from cellobiose, by combining 13 enzyme components, resulted in an overall yield of 11.2 moles H_2 (93.1% of the theoretical yield) [449]. The high cost and slow reaction rates of *in vitro* systems limit applications in the near future. Securing stability of involved enzymes, cofactors and generated intermediates, but also the up-scaling form major challenges. The usage of (hyper)thermophilic enzymes is suggested to increase conversion rates and decrease cost due to their stability and activity at elevated temperatures.

9. Concluding remarks

During the last two decades the efforts of many researchers has proven that both *C. saccharolyticus* and *T. maritima* are excellent candidates for biohydrogen production via dark fermentation. These organisms have favourable genomic backgrounds, their catabolic pathways lead only to a limited number of possible fermentation end-products, and moreover, the pathway generating the highest ATP yield (acetate formation), also leads to the highest H_2 yield. In other words, these organisms want to produce H_2!

It has been shown that growth under ideal conditions, i.e. low substrate concentration and low dissolved H_2 concentration, allows them to ferment sugar substrates to H_2 with yields close to the maximum theoretical yield. This feature is a very desirable starting point with respect to their biotechnological application as H_2-producers. Moreover, the ability of *C. saccharolyticus* and *T. maritima* species to hydrolyze and ferment complex sugars might allow them to be used in a consolidated

bioprocess, where no additional pre-treatment is required, either as a pure culture or in co-culture with other species.

The dissolved H_2 concentration was shown to be a dominant process determinant in causing the fermentation profile to shift away from maximal H_2 yields. To be able to uphold desirable features for a H_2 production process under high sugar load conditions, in terms of high H_2 yields and productivities, the dissolved hydrogen concentrations should be kept below the fermentation switch threshold. This is only achievable via proper reactor design and certainly required for the efficient up-scaling of this bioH_2 production process.

For the future application of bioH_2 production by *C. saccharolyticus* and *T. maritima,* research should focus on i) the factors limiting complex sugar degradation, ii) the specific nutritional requirements to sustain growth during high sugar loads, and iii) specific mechanisms to make the organism more robust against stresses like elevated dissolved H_2 concentrations or osmotic stress. Moreover, for the incorporation of these desirable traits the development of genetic systems is required.

Appendices

References

[1] **Cavendish H.** Three papers, containing experiments on factitious air, by the Hon. Henry Cavendish, F. R. S. *Philosophical Transactions (1683-1775)* (**1766**), 56:141-84.

[2] **Bennaceur K, *et al.*** Hydrogen: A future energy carrier? *Oilfield Review* (**2005**), 17:30-41.

[3] **Winter CJ.** Hydrogen energy - Abundant, efficient, clean: A debate over the energy-system-of-change. *Int. J. Hydrogen. Energy* (**2009**), 34:S1-S52.

[4] **Thauer RK, *et al.*** Energy-conservation in chemotropic anaerobic bacteria. *Bacteriol. Rev.* (**1977**), 41:100-80.

[5] **Holladay JD, *et al.*** An overview of hydrogen production technologies. *Catalysis Today* (**2009**), 139:244-60.

[6] **Dau H, *et al.*** The mechanism of water oxidation: From electrolysis via homogeneous to biological catalysis. *Chemcatchem* (**2010**), 2:724-61.

[7] **Li ZS, *et al.*** Photoelectrochemical cells for solar hydrogen production: current state of promising photoelectrodes, methods to improve their properties, and outlook. *Energy & Environmental Science* (**2013**), 6:347-70.

[8] **Corbo P, *et al.*** 2011. Hydrogen as future energy carrier. *In* P. Corbo, F. Migliardini,O. Veneri (ed.), Hydrogen Fuel Cells for Road Vehicles. Springer-Verlag, London.

[9] **Das D, *et al.*** Hydrogen production by biological processes: a survey of literature. *Int. J. Hydrogen. Energy* (**2001**), 26:13-28.

[10] **Hallenbeck PC, *et al.*** Advances in fermentative biohydrogen production: the way forward? *Trends Biotechnol.* (**2009**), 27:287-97.

[11] **Liu H, *et al.*** Electrochemically assisted microbial production of hydrogen from acetate. *Environmental Science & Technology* (**2005**), 39:4317-20.

[12] **Basak N, *et al.*** The prospect of purple non-sulfur (PNS) photosynthetic bacteria for hydrogen production: The present state of the art. *World Journal of Microbiology & Biotechnology* (**2007**), 23:31-42.

[13] **Kim BH, *et al.*** Control of carbon and electron flow in *Clostridium acetobutylicum* fermentations: Utilization of carbon-monoxide to inhibit hydrogen-production and to enhance butanol yields. *Appl. Environ. Microbiol.* (**1984**), 48:764-70.

[14] **Rachman MA, *et al.*** Hydrogen production with high yield and high evolution rate by self-flocculated cells of *Enterobacter aerogenes* in a packed-bed reactor. *Appl. Microbiol. Biotechnol.* (**1998**), 49:450-4.

[15] **Minnan L, *et al.*** Isolation and characterization of a high H_2-producing strain *Klebsiella oxytoca* HP1 from a hot spring. *Res. Microbiol.* (**2005**), 156:76-81.

[16] **Vancanneyt M, *et al.*** Ethanol production in batch and continuous culture from some carbohydrates with *Clostridium thermosaccharolyticum* LMG 6564. *Syst. Appl. Microbiol.* (**1990**), 13:382-7.

[17] **Shaw AJ, *et al.*** Identification of the FeFe -hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium saccharolyticum* and demonstration of increased ethanol yield via hydrogenase knockout. *J. Bacteriol.* (**2009**), 191:6457-64.

[18] **Magnusson L, *et al.*** Continuous hydrogen production during fermentation of alpha-cellulose by the thermophillic bacterium *Clostridium thermocellum*. *Biotechnol. Bioeng.* (**2009**), 102:759-66.

[19] **O-Thong S, *et al.*** Thermophilic fermentative hydrogen production by the newly isolated *Thermoanaerobacterium thermosaccharolyticum* PSU-2. *Int. J. Hydrogen. Energy* (**2008**), 33:1204-14.

[20] **de Vrije T, *et al.*** Pretreatment of Miscanthus for hydrogen production by *Thermotoga elfii*. *Int. J. Hydrogen. Energy* (**2002**), 27:1381-90.

[21] **van Niel EWJ, *et al.*** Distinctive properties of high hydrogen producing extreme thermophiles, *Caldicellulosiruptor saccharolyticus* and *Thermotoga elfii*. *Int. J. Hydrogen. Energy* (**2002**), 27:1391-8.

[22] **Soboh B, *et al.*** A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis*. *Microbiology* (**2004**), 150:2451-63.

[23] **Schröder C, *et al.*** Glucose fermentation to acetate, CO_2 and H_2 in the anaerobic hyperthermophilic eubacterium *Thermotoga maritima*: involvement of the Embden-Meyerhof pathway. *Arch. Microbiol.* (**1994**), 161:460-70.

[24] **Munro SA, *et al.*** The fermentation stoichiometry of *Thermotoga neapolitana* and influence of temperature, oxygen, and pH on hydrogen production. *Biotechnol. Prog.* (**2009**), 25:1035-42.

[25] **Kanai T, *et al.*** Continuous hydrogen production by the hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1. *J. Biotechnol.* (**2005**), 116:271-82.

[26] **Kengen SWM, *et al.*** Growth and energy-conservation in batch cultures of *Pyrococcus furiosus*. *FEMS Microbiol. Lett.* (**1994**), 117:305-9.

[27] **Kengen SWM, *et al.*** 2009. Biological hydrogen production by anaerobic microorganisms, p. 197-221. *In* W. Soetaert,E. J. Verdamme (ed.), Biofuels. John Wiley & Sons, Chichester.

[28] **Amend JP, *et al.*** Carbohydrates in thermophile metabolism: Calculation of the standard molal thermodynamic properties of aqueous pentoses and hexoses at elevated temperatures and pressures. *Geochimica Et Cosmochimica Acta* (**2001**), 65:3901-17.

[29] **Amend JP, *et al.*** Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* (**2001**), 25:175-243.

[30] **Schut GJ, *et al.*** The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: A new perspective on anaerobic hydrogen production. *J. Bacteriol.* (**2009**), 191:4451-7.

[31] **Burton K.** Enthalpy change for reduction of nicotinamide-adenine dinucleotide. *Biochem. J.* (**1974**), 143:365-8.

[32] **Watt GD*, et al.*** Thermochemical characterization of sodium dithionite, flavin mononucleotide, flavin-adenine dinucleotide and methyl and benzyl viologens as low-potential reductants for biological-systems. *Biochem. J.* (**1975**), 152:33-7.

[33] **Bergquist PL*, et al.*** Molecular diversity of thermophilic cellulolytic and hemicellulolytic bacteria. *Fems Microbiology Ecology* (**1999**), 28:99-110.

[34] **Blumer-Schuette SE*, et al.*** Extremely thermophilic microorganisms for biomass conversion: Status and prospects. *Curr. Opin. Biotechnol.* (**2008**), 19:210-7.

[35] **VanFossen AL*, et al.*** Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. *Incredible Anaerobes: From Physiology to Genomics to Fuels* (**2008**), 1125:322-37.

[36] **Sissons CH*, et al.*** Isolation of cellulolytic anaerobic extreme thermophiles from New Zealand thermal sites. *Appl. Environ. Microbiol.* (**1987**), 53:832-8.

[37] **Rainey FA*, et al.*** Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: An obligately anaerobic, extremely thermophilic, cellulolytic bacterium. *FEMS Microbiol. Lett.* (**1994**), 120:263-6.

[38] **Donnison AM*, et al.*** The degradation of lignocellulosics by extremely thermophilic microorganisms. *Biotechnol. Bioeng.* (**1989**), 33:1495-9.

[39] **Reynolds PHS*, et al.*** Comparison of cellulolytic activities in *Clostridium thermocellum* and three thermophilic, cellulolytic anaerobes. *Appl. Environ. Microbiol.* (**1986**), 51:12-7.

[40] **van de Werken HJG*, et al.*** Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl. Environ. Microbiol.* (**2008**), 74:6720-9.

[41] **VanFossen AL*, et al.*** Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. *Appl. Environ. Microbiol.* (**2009**), 75:7718-24.

[42] **Blumer-Schuette SE*, et al.*** *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. *J. Bacteriol.* (**2012**), 194:4015-28.

[43] **VanFossen AL*, et al.*** Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Biotechnol. Bioeng.* (**2011**), 108:1559-69.

[44] **de Vrije T*, et al.*** Efficient hydrogen production from the lignocellulosic energy crop *Miscanthus* by the extreme thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. *Biotechnol Biofuels* (**2009**), 2:12.

[45] **de Vrije T*, et al.*** Hydrogen production from carrot pulp by the extreme thermophiles *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. *Int. J. Hydrogen. Energy* (**2010**), 35:13206-13.

[46] **Herbel Z*, et al.*** Exploitation of the extremely thermophilic *Caldicellulosiruptor saccharolyticus* in hydrogen and biogas production from biomasses. *Environmental Technology* (**2010**), 31:1017-24.

[47] **Mars AE*, et al.*** Biohydrogen production from untreated and hydrolyzed potato steam peels by the extreme thermophiles *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. *Int. J. Hydrogen. Energy* (**2010**), 35:7730-7.

[48] **de Vrije T*, et al.*** Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. *Appl. Microbiol. Biotechnol.* (**2007**), 74:1358-67.

[49] **Adams MWW*, et al.*** Iron-sulfur clusters of hydrogenase-I and hydrogenase-II of *Clostridium pasteurianum*. *Proc. Natl. Acad. Sci. U. S. A.* (**1989**), 86:4932-6.

[50] **van Niel EWJ*, et al.*** Substrate and product inhibition of hydrogen production by the extreme thermophile, *Caldicellulosiruptor saccharolyticus*. *Biotechnol. Bioeng.* (**2003**), 81:255-62.

[51] **Willquist K*, et al.*** Lactate formation in *Caldicellulosiruptor saccharolyticus* is regulated by the energy carriers pyrophosphate and ATP. *Metabolic Engineering* (**2010**), 12:282-90.

[52] **Huber R*, et al.*** *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90^{o}C. *Arch. Microbiol.* (**1986**), 144:324-33.

[53] **Nelson KE*, et al.*** Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* (**1999**), 399:323-9.

[54] **Frock AD*, et al.*** Hyperthermophilic *Thermotoga* species differ with respect to specific carbohydrate transporters and glycoside hydrolases. *Appl. Environ. Microbiol.* (**2012**), 78:1978-86.

[55] **Chhabra SR*, et al.*** Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Biol. Chem.* (**2003**), 278:7540-52.

[56] **Bredholt S*, et al.*** *Caldicellulosiruptor kristjanssonii* sp. nov., a cellulolytic extremely thermophilic, anaerobic bacterium. *International Journal of Systematic Bacteriology* (**1999**), 49:991-6.

[57] **Elkins JG*, et al.*** Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47T. *J. Bacteriol.* (**2010**), 192:6099-100.

[58] **Huang CY*, et al.*** *Caldicellulosiruptor owensensis* sp. nov., an anaerobic, extremely thermophilic, xylanolytic bacterium. *International Journal of Systematic Bacteriology* (**1998**), 48:91-7.

[59] **Miroshnichenko ML*, et al.*** *Caldicellulosiruptor kronotskyensis* sp. nov. and *Caldicellulosiruptor hydrothermalis* sp. nov., two extremely thermophilic, cellulolytic, anaerobic bacteria from Kamchatka thermal springs. *International Journal of Systematic and Evolutionary Microbiology* (**2008**), 58:1492-6.

[60] **Mladenovska Z*, et al.*** Isolation and characterization of *Caldicellulosiruptor lactoaceticus* sp. nov., an extremely thermophilic, cellulolytic, anaerobic bacterium. *Arch. Microbiol.* (**1995**), 163:223-30.

References

[61] **Nielsen P, *et al.*** *Thermoanaerobium acetigenum* spec. nov., a new anaerobic, extremely thermophilic, xylanolytic non-spore-forming bacterium isolated from an Icelandic hot spring. *Arch. Microbiol.* (**1993**), 159:460-4.

[62] **Svetlichnyi VA, *et al.*** *Anaerocellum thermophilum* gen. nov. sp. nov.: An extremely thermophilic cellulolytic eubacterium isolated from hot-springs in the Valley of Geysers. *Microbiology* (**1990**), 59:598-604.

[63] **Balk M, *et al.*** *Thermotoga lettingae* sp nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. *International Journal of Systematic and Evolutionary Microbiology* (**2002**), 52:1361-8.

[64] **Jannasch HW, *et al.*** *Thermotoga neapolitana* sp. Nov. of the extremely thermophilic, eubacterial genus *Thermotoga*. *Arch. Microbiol.* (**1988**), 150:103-4.

[65] **Takahata Y, *et al.*** *Thermotoga petrophila* sp nov and *Thermotoga naphthophila* sp nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. *International Journal of Systematic and Evolutionary Microbiology* (**2001**), 51:1901-9.

[66] **Windberger E, *et al.*** *Thermotoga thermarum* sp nov and *Thermotoga neapolitana* occurring in African continental solfataric springs. *Arch. Microbiol.* (**1989**), 151:506-12.

[67] **Zhaxybayeva O, *et al.*** On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl. Acad. Sci. U. S. A.* (**2009**), 106:5865-70.

[68] **Heinonen JK.** 2001. Biological production of PP_i, p. 264. *In* J. K. Heinonen (ed.), Biological Role of Inorganic Pyrophosphate. Kluwer Academic Publishers, Boston / Dordrecht / London.

[69] **Chen J, *et al.*** Pyrophosphatase is essential for growth of *Escherichia coli*. *J. Bacteriol.* (**1990**), 172:5686-9.

[70] **Fields S.** Making the best of biomass - Hydrogen for fuel cells. *Environ. Health Perspect.* (**2003**), 111:A38-A41.

[71] **Das D, *et al.*** Advances in biological hydrogen production processes. *Int. J. Hydrogen. Energy* (**2008**), 33:6046-57.

[72] **Logan BE.** Extracting hydrogen electricity from renewable resources. *Environmental Science & Technology* (**2004**), 38:160A-7A.

[73] **Veneri O.** 2011. Hydrogen as future energy carrier, p. 33-70. *In* P. Corbo, F. Migliardini,O. Veneri (ed.), Hydrogen Fuel Cells for Road Vehicles. Springer London.

[74] **Ren NQ, *et al.*** Bioconversion of lignocellulosic biomass to hydrogen: Potential and challenges. *Biotechnology Advances* (**2009**), 27:1051-60.

[75] **Cappelletti M, *et al.*** Biohydrogen production from glucose, molasses and cheese whey by suspended and attached cells of four hyperthermophilic *Thermotoga* strains. *J. Chem. Technol. Biotechnol.* (**2012**), 87:1291-301.

[76] **da Silva GP, *et al.*** Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnology Advances* (**2009**), 27:30-9.

[77] **Guo XM, *et al.*** Hydrogen production from agricultural waste by dark fermentation: A review. *Int. J. Hydrogen. Energy* (**2010**), 35:10660-73.

[78] **Sarkar N, *et al.*** Bioethanol production from agricultural wastes: An overview. *Renewable Energy* (**2012**), 37:19-27.

[79] **Maru BT, *et al.*** Biohydrogen production by dark fermentation of glycerol using *Enterobacter* and *Citrobacter* Sp. *Biotechnol. Prog.* (**2013**), 29:31-8.

[80] **Yazdani SS, *et al.*** Anaerobic fermentation of glycerol: A path to economic viability for the biofuels industry. *Curr. Opin. Biotechnol.* (**2007**), 18:213-9.

[81] **Santibanez C, *et al.*** Residual glycerol from biodiesel manufacturing, waste or potential source of bioenergy: A review. *Chilean Journal of Agricultural Research* (**2011**), 71:469-75.

[82] **Clomburg JM, *et al.*** Metabolic engineering of *Escherichia coli* for the production of 1,2-propanediol from glycerol. *Biotechnol. Bioeng.* (**2011**), 108:867-79.

[83] **Ito T, *et al.*** Hydrogen and ethanol production from glycerol-containing wastes discharged after biodiesel manufacturing process. *J Biosci Bioeng* (**2005**), 100:260-5.

[84] **Temudo MF, *et al.*** Glycerol fermentation by (open) mixed cultures: A chemostat study. *Biotechnol. Bioeng.* (**2008**), 100:1088-98.

[85] **Mathews J, *et al.*** Metabolic pathway engineering for enhanced biohydrogen production. *Int. J. Hydrogen. Energy* (**2009**), 34:7404-16.

[86] **de Vrije T, *et al.*** 2003. Dark hydrogen fermentations, p. 103-23. *In* J. H. Reith, R. H. Wijffels,H. Barten (ed.), Bio-methane & Bio-hydrogen. Smiet Offset, The Hague.

[87] **Wiegel J, *et al.*** The importance of thermophilic bacteria in biotechnology. *Crc Critical Reviews in Biotechnology* (**1986**), 3:39-108.

[88] **Van Ooteghem SA, *et al.*** 2002.

[89] **Ngo TA, *et al.*** High-yield biohydrogen production from biodiesel manufacturing waste by *Thermotoga neapolitana*. *Int. J. Hydrogen. Energy* (**2011**), 36:5836-42.

[90] **Eriksen NT, *et al.*** H_2 synthesis from pentoses and biomass in *Thermotoga* spp. *Biotechnology Letters* (**2011**), 33:293-300.

[91] **Maru BT, *et al.*** Biohydrogen production from glycerol using *Thermotoga* spp. *Energy Procedia* (**2012**), 29:300-7.

[92] **Nguyen TAD, *et al.*** Optimization of hydrogen production by hyperthermophilic eubacteria, *Thermotoga maritima* and *Thermotoga neapolitana* in batch fermentation. *Int. J. Hydrogen. Energy* (**2008**), 33:1483-8.

[93] **Mu Y, *et al.*** A kinetic approach to anaerobic hydrogen-producing process. *Water Research* (**2007**), 41:1152-60.

[94] **Heijnen JJ.** 1999. Bioenergetics of microbial growth, p. 267–91, Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation. John Wiley & Sons, Inc.

[95] **Gonzalez R, *et al.*** A new model for the anaerobic fermentation of glycerol in enteric bacteria: Trunk and auxiliary pathways in *Escherichia coli*. *Metabolic Engineering* (**2008**), 10:234-45.
[96] **Biebl H.** Fermentation of glycerol by *Clostridium pasteurianum* - batch and continuous culture studies. *J. Ind. Microbiol. Biotechnol.* (**2001**), 27:18-26.
[97] **Kumar N, *et al.*** Redirection of biochemical pathways for the enhancement of H_2 production by *Enterobacter cloacae*. *Biotechnology Letters* (**2001**), 23:537-41.
[98] **Ngo TA, *et al.*** Dark fermentation of hydrogen from waste glycerol using hyperthermophilic eubacterium *Thermotoga neapolitana*. *Environmental Progress & Sustainable Energy* (**2012**), 31:466-73.
[99] **Cronan JE, *et al.*** Mutants of *Escherichia coli* defective in membrane phospholipid synthesis: Mapping of structural gene for L-glycerol 3-phosphate dehydrogenase. *J. Bacteriol.* (**1974**), 118:598-605.
[100] **Austin D, *et al.*** Nucleotide sequence of the *glpD* gene encoding aerobic sn-glycerol 3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* (**1991**), 173:101-7.
[101] **Yeh JI, *et al.*** Structure of glycerol-3-phosphate dehydrogenase, an essential monotopic membrane enzyme involved in respiration and metabolism. *Proc. Natl. Acad. Sci. U. S. A.* (**2008**), 105:3280-5.
[102] **Cole ST, *et al.*** Nucleotide sequence and gene-polypeptide relationships of the glpABC operon encoding the anaerobic sn-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* (**1988**), 170:2448-56.
[103] **Unden G, *et al.*** Alternative respiratory pathways of *Escherichia coli*: Energetics and transcriptional regulation in response to electron acceptors. *Biochim. Biophys. Acta-Bioenergetics* (**1997**), 1320:217-34.
[104] **Selig M, *et al.*** Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. *Arch. Microbiol.* (**1997**), 167:217-32.
[105] **Blamey JM, *et al.*** Characterization of an ancestral type of pyruvate ferredoxin oxidoreductase from the hyperthermophilic bacterium, *Thermotoga maritima Biochem. (Mosc.)* (**1994**), 33:1000-7.
[106] **Biegel E, *et al.*** Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell. Mol. Life Sci.* (**2011**), 68:613-34.
[107] **Bielen AAM, *et al.*** Pyrophosphate as a central energy carrier in the hydrogen-producing extremely thermophilic *Caldicellulosiruptor saccharolyticus*. *FEMS Microbiol. Lett.* (**2010**), 307:48-54.
[108] **Schink B, *et al.*** Energetics of syntrophic fatty-acid oxidation. *FEMS Microbiol. Rev.* (**1994**), 15:85-94.
[109] **Hallenbeck PC.** Fermentative hydrogen production: Principles, progress, and prognosis. *Int. J. Hydrogen. Energy* (**2009**), 34:7379-89.
[110] **Kadar Z, *et al.*** Yields from glucose, xylose, and paper sludge hydrolysate during hydrogen production by the extreme thermophile *Caldicellulosiruptor saccharolyticus*. *Appl. Biochem. Biotechnol.* (**2004**), 113-16:497-508.
[111] **Willquist K, *et al.*** Evaluation of the influence of CO_2 on hydrogen production by *Caldicellulosiruptor saccharolyticus*. *Int. J. Hydrogen. Energy* (**2009**), 34:4718-26.
[112] **Ljunggren M, *et al.*** A kinetic model for quantitative evaluation of the effect of hydrogen and osmolarity on hydrogen production by *Caldicellulosiruptor saccharolyticus*. *Biotechnol Biofuels* (**2011**), 4:31.
[113] **Willquist K, *et al.*** Reassessment of hydrogen tolerance in *Caldicellulosiruptor saccharolyticus*. *Microbial Cell Factories* (**2011**), 10:111.
[114] **Novichkov PS, *et al.*** RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* (**2010**), 38:W299-W307.
[115] **Blumer-Schuette SE, *et al.*** Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*. *J. Bacteriol.* (**2011**), 193:1483-4.
[116] **Kataeva IA, *et al.*** Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. *J. Bacteriol.* (**2009**), 191:3760-1.
[117] **Hebbeln P, *et al.*** Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. *Proc. Natl. Acad. Sci. U. S. A.* (**2007**), 104:2909-14.
[118] **Eisenberg MA, *et al.*** Purification and properties of the biotin repressor - a bifunctional protein. *J. Biol. Chem.* (**1982**), 257:5167-73.
[119] **Siche S, *et al.*** A bipartite S unit of an ECF-type cobalt transporter. *Res. Microbiol.* (**2010**), 161:824-9.
[120] **Mai XH, *et al.*** Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*: a new enzyme involved in peptide fermentation. *J. Biol. Chem.* (**1994**), 269:16726-32.
[121] **McBride MJ.** Bacterial gliding motility: Multiple mechanisms for cell movement over surfaces. *Annu. Rev. Microbiol.* (**2001**), 55:49-75.
[122] **McLaughlin KJ, *et al.*** Structural basis for NADH/NAD(+) redox sensing by a Rex family repressor. *Mol. Cell* (**2010**), 38:563-75.
[123] **Pei JJ, *et al.*** The mechanism for regulating ethanol fermentation by redox levels in *Thermoanaerobacter ethanolicus*. *Metabolic Engineering* (**2011**), 13:186-93.
[124] **Wietzke M, *et al.*** The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* (**2012**):1-13.
[125] **Ravcheev DA, *et al.*** Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex. *J. Bacteriol.* (**2012**), 194:1145-57.
[126] **Kammler M, *et al.*** Characterization of the ferrous iron uptake system of *Escherichia coli*. *J. Bacteriol.* (**1993**), 175:6212-9.

References

[127] **Schujman GE, *et al.*** FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Developmental Cell* (**2003**), 4:663-72.

[128] **Pauss A, *et al.*** Liquid-to-gas mass transfer in anaerobic processes: Inevitable transfer limitations of methane and hydrogen in the biomethanation process. *Appl. Environ. Microbiol.* (**1990**), 56:1636-44.

[129] **Wilhelm E, *et al.*** Low-pressure solubility of gases in liquid water. *Chem. Rev.* (**1977**), 77:219-62.

[130] **Ellis LD, *et al.*** Closing the carbon balance for fermentation by *Clostridium thermocellum* (ATCC 27405). *Bioresource Technol.* (**2012**), 103:293-9.

[131] **Heinekey DM.** Hydrogenase enzymes: Recent structural studies and active site models. *J Organomet Chem* (**2009**), 694:2671-80.

[132] **Burdette D, *et al.*** Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase (2^{o} adh) as a bifunctional alcohol-dehydrogenase acetyl-coA reductive thioesterase. *Biochem. J.* (**1994**), 302:163-70.

[133] **Burdette DS, *et al.*** Physiological function of alcohol dehydrogenases and long-chain (C-30) fatty acids in alcohol tolerance of *Thermoanaerobacter ethanolicus*. *Appl. Environ. Microbiol.* (**2002**), 68:1914-8.

[134] **Ma K, *et al.*** Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. *Proc. Natl. Acad. Sci. U. S. A.* (**1997**), 94:9608-13.

[135] **Wang SN, *et al.*** NADP(+) reduction with reduced ferredoxin and NADP(+) reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*. *J. Bacteriol.* (**2010**), 192:5115-23.

[136] **Schujman GE, *et al.*** Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. *EMBO J.* (**2006**), 25:4074-83.

[137] **Payot S, *et al.*** Induction of lactate production associated with a decrease in NADH cell content enables growth resumption of *Clostridium cellulolyticum* in batch cultures on cellobiose. *Res. Microbiol.* (**1999**), 150:465-73.

[138] **Brekasis D, *et al.*** A novel sensor of NADH/NAD(+) redox poise in *Streptomyces coelicolor* A3(2). *EMBO J.* (**2003**), 22:4856-65.

[139] **Pagels M, *et al.*** Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in *Staphylococcus aureus*. *Mol. Microbiol.* (**2010**), 76:1142-61.

[140] **Wang E, *et al.*** Structure and functional properties of the *Bacillus subtilis* transcriptional repressor Rex. *Mol. Microbiol.* (**2008**), 69:466-78.

[141] **Bitoun JP, *et al.*** Transcriptional repressor Rex is involved in regulation of oxidative stress response and biofilm formation by *Streptococcus* mutans. *FEMS Microbiol. Lett.* (**2011**), 320:110-7.

[142] **Novichkov PS, *et al.*** RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.* (**2010**), 38:D111-D8.

[143] **Lynd LR, *et al.*** Biocommodity engineering. *Biotechnol. Prog.* (**1999**), 15:777-93.

[144] **Lynd LR, *et al.*** Consolidated bioprocessing of cellulosic biomass: An update. *Curr. Opin. Biotechnol.* (**2005**), 16:577-83.

[145] **Agbor VB, *et al.*** Biomass pretreatment: Fundamentals toward application. *Biotechnology Advances* (**2011**), 29:675-85.

[146] **Olson DG, *et al.*** Recent progress in consolidated bioprocessing. *Curr. Opin. Biotechnol.* (**2012**), 23:396-405.

[147] **Verhaart MRA, *et al.*** Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: Mechanisms for reductant disposal. *Environmental Technology* (**2010**), 31:993-1003.

[148] **Willquist K, *et al.*** Physiological characteristics of the extreme thermophile *Caldicellulosiruptor saccharolyticus*: An efficient hydrogen cell factory. *Microbial Cell Factories* (**2010**), 9:89.

[149] **Dwivedi PP, *et al.*** Cloning, sequencing and overexpression in *Escherichia coli* of a xylanase gene, xynA from the thermophilic bacterium Rt8B.4 genus *Caldicellulosiruptor*. *Appl. Microbiol. Biotechnol.* (**1996**), 45:86-93.

[150] **Gibbs MD, *et al.*** Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile *Caldicellulosiruptor* isolate Tok7B.1. *Curr. Microbiol.* (**2000**), 40:333-40.

[151] **Hamilton-Brehm SD, *et al.*** *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian pool, Yellowstone national park. *Appl. Environ. Microbiol.* (**2010**), 76:1014-20.

[152] **Morris DD, *et al.*** Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp. strain Rt69B.1. *Extremophiles* (**1999**), 3:103-11.

[153] **Onyenwoke RU, *et al.*** Reclassification of "*Thermoanaerobium acetigenum*" as *Caldicellulosiruptor acetigenus* comb. nov and emendation of the genus description. *International Journal of Systematic and Evolutionary Microbiology* (**2006**), 56:1391-5.

[154] **Yang SJ, *et al.*** Reclassification of "*Anaerocellum thermophilum*" as *Caldicellulosiruptor bescii* strain DSM 6725T sp. nov. *Int J Syst Evol Microbiol* (**2009**), 60:2011-5.

[155] **Lamed R, *et al.*** The cellulosome of *Clostridium thermocellum*. *Adv. Appl. Microbiol.* (**1988**), 33:1-46.

[156] **Blumer-Schuette SE, *et al.*** Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. *Appl. Environ. Microbiol.* (**2010**), 76:8084-92.

[157] **Bergquist PL, *et al.*** Genetics and potential biotechnological applications of thermophilic and extremely thermophilic microorganisms. *Biotechnol. Genet. Eng. Rev.* (**1987**), 5:199-244.

[158] **Love DR, *et al.*** Molecular cloning of a beta-glucosidase gene from an extremely thermophilic anaerobe in *Escherichia coli* and *Bacillus subtilis*. *Bio-Technology* (**1987**), 5:384-7.

[159] **Hudson RC, *et al.*** Purification and properties of an aryl beta-xylosidase from a cellulolytic extreme thermophile expressed in *Escherichia coli*. *Biochem. J.* (**1991**), 273:645-50.

[160] **Schofield LR, *et al.*** Purification and properties of a beta-1,4-xylanase from a cellulolytic extreme thermophile expressed in *Escherichia coli*. *Int. J. Biochem.* (**1993**), 25:609-17.

[161] **Albertson GD, *et al.*** Cloning and sequence of a type I pullulanase from an extremely thermophilic anaerobic bacterium, *Caldicellulosiruptor saccharolyticus*. *Biochim. Biophys. Acta-Gene Structure and Expression* (**1997**), 1354:35-9.

[162] **Luthi E, *et al.*** Xylanase from the extremely thermophilic bacterium "*Caldocellum saccharolyticum*": Overexpression of the gene in *Escherichia coli* and characterization of the gene product. *Appl. Environ. Microbiol.* (**1990**), 56:2677-83.

[163] **Luthi E, *et al.*** A beta-D-xylosidase from the thermophile "*Caldocellum saccharolyticum*" expressed in *Escherichia coli*. *FEMS Microbiol. Lett.* (**1990**), 67:291-4.

[164] **Luthi E, *et al.*** Overproduction of an acetylxylan esterase from the extreme thermophile "*Caldocellum saccharolyticum*" in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* (**1990**), 34:214-9.

[165] **Te'o VSJ, *et al.*** Cela, another gene coding for a multidomain cellulase from the extreme thermophile "*Caldocellum saccharolyticum*". *Appl. Microbiol. Biotechnol.* (**1995**), 43:291-6.

[166] **Park CS, *et al.*** Characterization of a recombinant cellobiose 2-epimerase from *Caldicellulosiruptor saccharolyticus* and its application in the production of mannose from glucose. *Appl. Microbiol. Biotechnol.* (**2011**), 92:1187-96.

[167] **Saul DJ, *et al.*** Celb, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum*". *Appl. Environ. Microbiol.* (**1990**), 56:3117-24.

[168] **Morris DD, *et al.*** Correction of the beta-mannanase domain of the *Celc* pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene product on Kraft pulp. *Appl. Environ. Microbiol.* (**1995**), 61:2262-9.

[169] **Luthi E, *et al.*** Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding for a beta-mannanase from the extremely thermophilic bacterium "*Caldocellum saccharolyticum*". *Appl. Environ. Microbiol.* (**1991**), 57:694-700.

[170] **Andrews G, *et al.*** Part I: Characterization of the extracellular proteome of the extreme thermophile *Caldicellulosiruptor saccharolyticus* by GeLC-MS2. *Anal. Bioanal. Chem.* (**2010**), 398:1837.

[171] **Ozdemir I, *et al.*** S-Layer homology domain proteins Csac_0678 and Csac_2722 are implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl. Environ. Microbiol.* (**2012**), 78:768-77.

[172] **Ivanova G, *et al.*** Hydrogen production from biopolymers by *Caldicellulosiruptor saccharolyticus* and stabilization of the system by immobilization. *Int. J. Hydrogen. Energy* (**2008**), 33:6953-61.

[173] **Brown SD, *et al.*** Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U. S. A.* (**2011**), 108:13752-7.

[174] **Peng H, *et al.*** The aldehyde/alcohol dehydrogenase (AdhE) in relation to the ethanol formation in *Thermoanaerobacter ethanolicus* JW200. *Anaerobe* (**2008**), 14:125-7.

[175] **Bielen AAM, *et al.*** A thermophile under pressure: Transcriptional analysis of the response of *Caldicellulosiruptor saccharolyticus* to different H_2 partial pressures. *Int. J. Hydrogen. Energy* (**2013**), 38:1837-49.

[176] **Kannan V, *et al.*** Ethanol fermentation characteristics of *Thermoanaerobacter ethanolicus*. *Enzyme Microb. Technol.* (**1985**), 7:87-9.

[177] **Desai SG, *et al.*** Cloning of L-lactate dehydrogenase and elimination of lactic acid production via gene knockout in *Thermoanaerobacterium saccharolyticum* JW/SL-YS485. *Appl. Microbiol. Biotechnol.* (**2004**), 65:600-5.

[178] **Ma KS, *et al.*** An unusual oxygen-sensitive, iron- and zinc-containing alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* (**1999**), 181:1163-70.

[179] **DeLacey AL, *et al.*** IR spectroelectrochemical study of the binding of carbon monoxide to the active site of *Desulfovibrio fructosovorans* Ni-Fe hydrogenase. *Journal of Biological Inorganic Chemistry* (**2002**), 7:318-26.

[180] **Lemon BJ, *et al.*** Binding of exogenously added carbon monoxide at the active site of the iron-only hydrogenase (CpI) from *Clostridium pasteurianum*. *Biochem. (Mosc.)* (**1999**), 38:12969-73.

[181] **Mertens E.** Pyrophosphate-dependent phosphofructokinase, an anaerobic glycolytic enzyme? . *FEBS Lett.* (**1991**), 285:1-5.

[182] **Kraemer JT, *et al.*** Supersaturation of dissolved H_2 and CO_2 during fermentative hydrogen production with N_2 sparging. *Biotechnology Letters* (**2006**), 28:1485-91.

[183] **Panagiotopoulos IA, *et al.*** Integration of first and second generation biofuels: Fermentative hydrogen production from wheat grain and straw. *Bioresource Technol.* (**2013**), 128:345-50.

[184] **Panagiotopoulos IA, *et al.*** Dilute-acid pretreatment of barley straw for biological hydrogen production using *Caldicellulosiruptor saccharolyticus*. *Int. J. Hydrogen. Energy* (**2012**), 37:11727-34.

[185] **Panagiotopoulos I, *et al.*** Exploring critical factors for fermentative hydrogen production from various types of lignocellulosic biomass. *Nihon Enerugi Gakkaishi/Journal of the Japan Institute of Energy* (**2011**), 90:363-8.

[186] **Özgür E, *et al.*** Biohydrogen production from beet molasses by sequential dark and photofermentation. *Int. J. Hydrogen. Energy* (**2010**), 35:511-7.

[187] **Panagiotopoulos JA, *et al.*** Prospects of utilization of sugar beet carbohydrates for biological hydrogen production in the EU. *Journal of Cleaner Production* (**2010**), 18:S9-S14.

[188] **Panagiotopoulos IA, *et al.*** Pretreatment of sweet sorghum bagasse for hydrogen production by *Caldicellulosiruptor saccharolyticus*. *Int. J. Hydrogen. Energy* (**2010**), 35:7738-47.

[189] **Yang SJ, *et al.*** Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. *Appl. Environ. Microbiol.* (**2009**), 75:4762-9.

References

[190] **Panagiotopoulos IA, *et al.*** Fermentative hydrogen production from pretreated biomass: A comparative study. *Bioresource Technol.* (**2009**), 100:6331-8.

[191] **Ivanova G, *et al.*** Thermophilic biohydrogen production from energy plants by *Caldicellulosiruptor saccharolyticus* and comparison with related studies. *Int. J. Hydrogen. Energy* (**2009**), 34:3659-70.

[192] **Bagi Z, *et al.*** Biotechnological intensification of biogas production. *Appl. Microbiol. Biotechnol.* (**2007**), 76:473-82.

[193] **Kadar Z, *et al.*** Hydrogen production from paper sludge hydrolysate. *Appl. Biochem. Biotechnol.* (**2003**), 105 -108:557-66.

[194] **Willquist K, *et al.*** Growth and hydrogen production characteristics of *Caldicellulosiruptor saccharolyticus* on chemically defined minimal media. *Int. J. Hydrogen. Energy* (**2012**), 37:4925-9.

[195] **Muddiman D, *et al.*** Part II: defining and quantifying individual and co-cultured intracellular proteomes of two thermophilic microorganisms by GeLC-MS(2) and spectral counting. *Anal. Bioanal. Chem.* (**2010**), 398:391-404.

[196] **van Groenestijn JW, *et al.*** Performance and population analysis of a non-sterile trickle bed reactor inoculated with *Caldicellulosiruptor saccharolyticus*, a thermophilic hydrogen producer. *Biotechnol. Bioeng.* (**2009**), 102:1361-7.

[197] **Ljunggren M, *et al.*** Techno-economic comparison of a biological hydrogen process and a 2nd generation ethanol process using barley straw as feedstock. *Bioresource Technol.* (**2011**), 102:9524-31.

[198] **Ljunggren M, *et al.*** Techno-economic evaluation of a two-step biological process for hydrogen production. *Biotechnol. Prog.* (**2010**), 26:496-504.

[199] **Nath K, *et al.*** Improvement of fermentative hydrogen production: Various approaches. *Appl. Microbiol. Biotechnol.* (**2004**), 65:520-9.

[200] **Kim DH, *et al.*** Effect of gas sparging on continuous fermentative hydrogen production. *Int. J. Hydrogen. Energy* (**2006**), 31:2158-69.

[201] **Kraemer JT, *et al.*** Optimisation and design of nitrogen-sparged fermentative hydrogen production bioreactors. *Int. J. Hydrogen. Energy* (**2008**), 33:6558-65.

[202] **Clark IC, *et al.*** The effect of low pressure and mixing on biological hydrogen production via anaerobic fermentation. *Int. J. Hydrogen. Energy* (**2012**), 37:11504-13.

[203] **Lamed RJ, *et al.*** Effects of stirring and hydrogen on fermentation products of *Clostridium thermocellum*. *Appl. Environ. Microbiol.* (**1988**), 54:1216-21.

[204] **Junghare M, *et al.*** Improvement of hydrogen production under decreased partial pressure by newly isolated alkaline tolerant anaerobe, *Clostridium butyricum* TM-9A: Optimization of process parameters. *Int. J. Hydrogen. Energy* (**2012**), 37:3160-8.

[205] **Mandal B, *et al.*** Improvement of biohydrogen production under decreased partial pressure of H_2 by *Enterobacter cloacae*. *Biotechnology Letters* (**2006**), 28:831-5.

[206] **Sonnleitner A, *et al.*** Process investigations of extreme thermophilic fermentations for hydrogen production: Effect of bubble induction and reduced pressure. *Bioresource Technol.* (**2012**), 118:170-6.

[207] **Fritsch M, *et al.*** Enhancing hydrogen production of *Clostridium butyricum* using a column reactor with square-structured ceramic fittings. *Int. J. Hydrogen. Energy* (**2008**), 33:6549-57.

[208] **Zeidan AA, *et al.*** Developing a thermophilic hydrogen-producing co-culture for efficient utilization of mixed sugars. *Int. J. Hydrogen. Energy* (**2009**), 34:4524-8.

[209] **Zeidan AA, *et al.*** A quantitative analysis of hydrogen production efficiency of the extreme thermophile *Caldicellulosiruptor owensensis* OL(T). *Int. J. Hydrogen. Energy* (**2010**), 35:1128-37.

[210] **Claassen PAM, T. de Vrije.** Non-thermal production of pure hydrogen from biomass: HYVOLUTION. *Int. J. Hydrogen. Energy* (**2006**), 31:1416 – 23.

[211] **Özgür E, *et al.*** Potential use of thermophilic dark fermentation effluents in photofermentative hydrogen production by *Rhodobacter capsulatus*. *Journal of Cleaner Production* (**2010**), 18:S23-S8.

[212] **Özkan E, *et al.*** Photofermentative hydrogen production using dark fermentation effluent of sugar beet thick juice in outdoor conditions. *Int. J. Hydrogen. Energy* (**2012**), 37:2044-9.

[213] **Chung D, *et al.*** Methylation by a unique α-class N4-Cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. *PLoS ONE* (**2012**), 7.

[214] **Bradford MM.** Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* (**1976**), 72:248-54.

[215] **Jeanmougin F, *et al.*** Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* (**1998**), 23:403-5.

[216] **Baykov AA, *et al.*** A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay *Anal. Biochem.* (**1988**), 171:266-70.

[217] **Cole HA, *et al.*** ATP pool in *Escherichia coli* .I. Measurement of pool using a modified luciferase assay. *Biochim. Biophys. Acta* (**1967**), 143:445-&.

[218] **Meyer CL, *et al.*** Increased levels of ATPand NADH are associated with increased solvent production in continuous cultures of *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* (**1989**), 30:450-9.

[219] **Drake HL, *et al.*** New, convenient method for the rapid analysis of inorganic pyrophosphate. *Anal. Biochem.* (**1979**), 94:117-20.

[220] **Bapteste E, *et al.*** Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase. *Gene* (**2003**), 318:185-91.

[221] **Chi A, *et al.*** The primordial high energy compound: ATP or inorganic pyrophosphate? *J. Biol. Chem.* (**2000**), 275:35677-9.

[222] **Moore SA*, et al.*** The structure of a pyrophosphate-dependent phosphofructokinase from the Lyme disease spirochete *Borrelia burgdorferi*. *Structure* (**2002**), 10:659-71.
[223] **Shintani T, *et al.*** Cloning and expression of a unique inorganic pyrophosphatase from *Bacillus subtilis*: evidence for a new family of enzymes. *FEBS Lett.* (**1998**), 439:263-6.
[224] **Serrano A, *et al.*** Proton-pumping inorganic pyrophosphatases in some archaea and other extremophilic prokaryotes. *J. Bioenerg. Biomembr.* (**2004**), 36:127-33.
[225] **Trotsenko YA, *et al.*** Studies on phosphate-metabolism in obligate methanotrophs. *FEMS Microbiol. Rev.* (**1990**), 87:267-71.
[226] **Heinonen JK, *et al.*** Comparative assessment of inorganic pyrophosphate and pyrophosphatase levels of *Escherichia coli*, *Clostridium pasteurianum*, and *Clostridium thermoaceticum*. *FEMS Microbiol. Lett*. (**1988**), 52:205-8.
[227] **Belouski E, *et al.*** Cloning, sequence, and expression of the phosphofructokinase gene of *Clostridium acetobutylicum* ATCC 824 in *Escherichia coli*. *Curr. Microbiol.* (**1998**), 37:17-22.
[228] **Llanos RM, *et al.*** Identification of a novel operon in *Lactococcus lactis* encoding 3 enzymes for lactic-acid synthesis - phosphofructokinase, pyruvate-kinase, and lactate-dehydrogenase. *J. Bacteriol.* (**1993**), 175:2541-51.
[229] **Acosta H, *et al.*** Pyruvate phosphate dikinase and pyrophosphate metabolism in the glycosome of *Trypanosoma cruzi* epimastigotes. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* (**2004**), 138:347-56.
[230] **Feng XM, *et al.*** The catalyzing role of PPDK in *Giardia lamblia*. *Biochem. Biophys. Res. Commun.* (**2008**), 367:394-8.
[231] **Tjaden B, *et al.*** Phosphoenolpyruvate synthetase and pyruvate, phosphate dikinase of *Thermoproteus tenax*: key pieces in the puzzle of archaeal carbohydrate metabolism. *Mol. Microbiol.* (**2006**), 60:287-98.
[232] **Chen RZ, *et al.*** Comparative studies of *Escherichia coli* strains using different glucose uptake systems: Metabolism and energetics. *Biotechnol. Bioeng.* (**1997**), 56:583-90.
[233] **McIntosh MT, *et al.*** Vacuolar type H^+ pumping pyrophosphatases of parasitic protozoa. *Int. J. Parasitol.* (**2002**), 32:1-14.
[234] **Alves A, *et al.*** Different physiological roles of ATP- and PP_i-Dependent phosphofructokinase isoenzymes in the methylotrophic actinomycete *Amycolatopsis methanolica*. *J. Bacteriol.* (**2001**), 183:7231-40.
[235] **Desantis D, *et al.*** Metabolism of mollicutes: the Embden-Meyerhof-Parnas pathway and the hexose monophosphate shunt. *J. Gen. Microbiol.* (**1989**), 135:683-91.
[236] **Reeves RE, *et al.*** Pyrophosphate-D-fructose 6-6hosphate 1-phosphotransferase - New enzyme with glycolytic function of 6-phosphofructokinase. *J. Biol. Chem.* (**1974**), 249:7737-41.
[237] **Saavedra E, *et al.*** Glycolysis in *Entamoeba histolytica* - Biochemical characterization of recombinant glycolytic enzymes and flux control analysis. *Febs Journal* (**2005**), 272:1767-83.
[238] **Verhees CH, *et al.*** The unique features of glycolytic pathways in Archaea. *Biochem. J.* (**2003**), 375:231-46.
[239] **Munoz ME, *et al.*** Pyruvate kinase: current status of regulatory and functional properties. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* (**2003**), 135:197-218.
[240] **Ronimus RS, *et al.*** The biochemical properties and phylogenies of phosphofructokinases from extremophiles. *Extremophiles* (**2001**), 5:357-73.
[241] **Sato T, *et al.*** Genetic evidence identifying the true gluconeogenic fructose-1,6- bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.* (**2004**), 186:5799-807.
[242] **Imanaka H, *et al.*** Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.* (**2006**), 61:898-909.
[243] **Kajander T, *et al.*** Inorganic pyrophosphatases: One substrate, three mechanisms. *FEBS Lett.* (**2013**).
[244] **Reeves RE.** Pyruvate, phosphate dikinase from *Bacteroides symbiosus*. *Biochem. J.* (**1971**), 125:531-&.
[245] **Siebers B, *et al.*** Glucose catabolism of the hyperthermophilic archaeum *Thermoproteus tenax*. *FEMS Microbiol. Lett.* (**1993**), 111:1-8.
[246] **Slamovits CH, *et al.*** Pyruvate-phosphate dikinase of oxymonads and parabasalia and the evolution of pyrophosphate-dependent glycolysis in anaerobic eukaryotes. *Eukaryotic Cell* (**2006**), 5:148-54.
[247] **Tatusov RL, *et al.*** The COG database: an updated version includes eukaryotes. *Bmc Bioinformatics* (**2003**), 4.
[248] **Tatusov RL, *et al.*** A genomic perspective on protein families. *Science* (**1997**), 278:631-7.
[249] **Kanehisa M, *et al.*** KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* (**2000**), 28:27-30.
[250] **Markowitz VM, *et al.*** IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* (**2012**), 40:D115-D22.
[251] **Hunter S, *et al.*** InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* (**2012**), 40:D306-D12.
[252] **Kengen SWM, *et al.*** Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. *J. Biol. Chem.* (**1994**), 269:17537-41.
[253] **Kengen SWM, *et al.*** Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic Archaeon Pyrococcus furiosus. *J. Biol. Chem.* (**1995**), 270:30453-7.
[254] **Ito S, *et al.*** Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Structure* (**2001**), 9:205-14.
[255] **Donahue JL, *et al.*** Purification and characterization of glpX-encoded fructose 1,6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*. *J. Bacteriol.* (**2000**), 182:5624-7.
[256] **Fujita Y, *et al.*** Identification and expression of the *Bacillus subtilis* fructose-1,6-bisphosphatase gene (fbp). *J. Bacteriol.* (**1998**), 180:4309-13.

References

[257] **Verhees CH*, et al.*** Biochemical adaptations of two sugar kinases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochem. J.* (**2002**), 366:121-7.

[258] **Zhang YP*, et al.*** Structural similarities between fructose-1,6-bisphosphatase and inositol monophosphatase. *Biochem. Biophys. Res. Commun.* (**1993**), 190:1080-3.

[259] **Verhees CH*, et al.*** Molecular and biochemical characterization of a distinct type of fructose-1,6-bisphosphatase from *Pyrococcus furiosus*. *J. Bacteriol.* (**2002**), 184:3401-5.

[260] **Banaszak K*, et al.*** The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle. *J. Mol. Biol.* (**2011**), 407:284-97.

[261] **Evans PR*, et al.*** Phosphofructokinase - structure and control. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* (**1981**), 293:53-62.

[262] **Mcnae IW*, et al.*** The crystal structure of ATP-bound phosphofructokinase from *Trypanosoma brucei* reveals conformational transitions different from those of other phosphofructokinases. *J. Mol. Biol.* (**2009**), 385:1519-33.

[263] **Shirakihara Y*, et al.*** Crystal-structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction-products. *J. Mol. Biol.* (**1988**), 204:973-94.

[264] **Xu J*, et al.*** Identification of basic residues involved in substrate-binding and catalysis by pyrophosphate-dependent phosphofructokinase from *Propionibacterium freudenreichii*. *J. Biol. Chem.* (**1994**), 269:15553-7.

[265] **Claustre S*, et al.*** Exploring the active site of *Trypanosoma brucei* phosphofructokinase by inhibition studies: Specific irreversible inhibition. *Biochem. (Mosc.)* (**2002**), 41:10183-93.

[266] **Lopez C*, et al.*** *Leishmania donovani* phosphofructokinase - Gene characterization, biochemical properties and structure-modelling studies. *Eur. J. Biochem.* (**2002**), 269:3978-89.

[267] **Schirmer T*, et al.*** Structural basis of the allosteric behaviour of phosphofructokinase. *Nature* (**1990**), 343:140-5.

[268] **Schoneberg T*, et al.*** Structure and allosteric regulation of eukaryotic 6-phosphofructokinases. *Biol. Chem.* (**2013**), 394:977-93.

[269] **Strater N*, et al.*** Molecular architecture and structural basis of allosteric regulation of eukaryotic phosphofructokinases. *FASEB J.* (**2011**), 25:89-98.

[270] **Martínez-Costa OH*, et al.*** Purification and properties of phosphofructokinase from *Dictyostelium discoideum*. *Eur. J. Biochem.* (**1994**), 226:1007-17.

[271] **Mertens E*, et al.*** Pyrophosphate-dependent phosphofructokinase from the amoeba *Naegleria fowleri*, an AMP-sensitive enzyme. *Biochem. J.* (**1993**), 292:797-803.

[272] **Vanschaftingen E*, et al.*** A kinetic-study of pyrophosphate - fructose-6-phosphate phosphotransferase from potato-tubers - application to a micro-assay of fructose 2,6-bisphosphate. *Eur. J. Biochem.* (**1982**), 129:191-5.

[273] **Carlisle SM*, et al.*** Pyrophosphate-dependent phosphofructokinase. Conservation of protein-sequence between the alpha- and beta-subunit and with the ATP-dependent phosphofructokinase. *J. Biol. Chem.* (**1990**), 265:18366-71.

[274] **Blangy D*, et al.*** Kinetics of allosteric interactions of phosphofructokinase from *Escherichia coli*. *J. Mol. Biol.* (**1968**), 31:13-&.

[275] **Xu J*, et al.*** Tetramer-dimer conversion of phosphofructokinase from *Thermus thermophilus* induced by its allosteric effectors. *J. Mol. Biol.* (**1990**), 215:597-606.

[276] **Hansen T*, et al.*** ATP-dependent 6-phosphofructokinase from the hyperthermophilic bacterium *Thermotoga maritima*: characterization of an extremely thermophilic, allosterically regulated enzyme. *Arch. Microbiol.* (**2002**), 177:401-9.

[277] **Kirchberger J*, et al.*** 6-phosphofructokinase from *Pichia pastoris*: purification, kinetic and molecular characterization of the enzyme. *Yeast* (**2002**), 19:933-47.

[278] **Su JGJ*, et al.*** Purification, kinetics and inhibition by antimonials of recombinant phosphofructokinase from *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* (**1996**), 81:171-8.

[279] **Alves AMCR*, et al.*** Characterization and phylogeny of the *pfp* gene of *Amycolatopsis methanolica* encoding PP_i-dependent phosphofructokinase. *J. Bacteriol.* (**1996**), 178:149-55.

[280] **Alves AMCR*, et al.*** Identification of ATP-dependent phosphofructokinase as a regulatory step in the glycolytic pathway of the actinomycete *Streptomyces coelicolor* A 3(2). *Appl. Environ. Microbiol.* (**1997**), 63:956-61.

[281] **Siebers B*, et al.*** PP_i-dependent phosphofructokinase from *Thermoproteus tenax*, an archaeal descendant of an ancient line in phosphofructokinase evolution. *J. Bacteriol.* (**1998**), 180:2137-43.

[282] **Rozova ON*, et al.*** Characterization of recombinant pyrophosphate-dependent 6-phosphofructokinase from halotolerant methanotroph *Methylomicrobium alcaliphilum* 20Z. *Res. Microbiol.* (**2010**), 161:861-8.

[283] **Rozova ON*, et al.*** Characterization of recombinant PP_i-dependent 6-phosphofructokinases from *Methylosinus trichosporium* OB3b and *Methylobacterium nodulans* ORS 2060. *Biochem. (Mosc.)* (**2012**), 77:288-95.

[284] **Ding YHR*, et al.*** *Thermotoga maritima* phosphofructokinases: Expression and characterization of two unique enzymes. *J. Bacteriol.* (**2001**), 183:791-4.

[285] **Mertens E*, et al.*** Presence of a fructose-2,6-bisphosphate-insensitive pyrophosphate - fructose-6-phosphate phosphotransferase in the anaerobic protozoa *Tritrichomonas foetus*, *Trichomonas vaginalis* and *Isotricha prostoma*. *Mol. Biochem. Parasitol.* (**1989**), 37:183-90.

[286] **Li ZJ*, et al.*** Pyrophosphate-dependent phosphofructokinase from *Giardia lamblia* - purification and characterization. *Protein Expr. Purif.* (**1995**), 6:319-28.

[287] **Ronimus RS*, et al.*** Phosphofructokinase activities within the order Spirochaetales and the characterisation of the pyrophosphate-dependent phosphofructokinase from *Spirochaeta thermophila*. *Arch. Microbiol.* (**1999**), 172:401-6.

[288] **Cronin CN*, et al.*** Purification and regulatory properties of phosphofructokinase from *Trypanosoma* (Trypanozoon) *brucei-brucei*. *Biochem. J.* (**1985**), 227:113-24.

[289] **Pollack JD*, et al.*** PP_i-dependent phosphofructotransferase (phosphofructokinase) activity in the mollicutes (mycoplasma) *Acholeplasma laidlawii*. *J. Bacteriol.* (**1986**), 165:53-60.

[290] **Reshetnikov AS*, et al.*** Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus capsulatus* Bath. *FEMS Microbiol. Lett.* (**2008**), 288:202-10.

[291] **Guixe V, *et al.*** The ADP-dependent sugar kinase family: Kinetic and evolutionary aspects. *Iubmb Life* (**2009**), 61:753-61.

[292] **Sakuraba H, *et al.*** ADP-dependent glucokinase/phosphofructokinase, a novel bifunctional enzyme from the hyperthermophilic Archaeon *Methanococcus jannaschii*. *J. Biol. Chem.* (**2002**), 277:12495-8.

[293] **Ronimus RS*, et al.*** Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochem. Biophys. Res. Commun.* (**2004**), 315:652-8.

[294] **Currie MA, *et al.*** ADP-dependent 6-phosphofructokinase from *Pyrococcus horikoshii* OT3 structure determination and biochemical characterization of ph1645. *J. Biol. Chem.* (**2009**), 284:22664-71.

[295] **Ito S, *et al.*** Crystal structure of an ADP-dependent glucokinase from *Pyrococcus furiosus*: Implications for a sugar-induced conformational change in ADP-dependent kinase. *J. Mol. Biol.* (**2003**), 331:871-83.

[296] **Jiang YH, *et al.*** The independent prokaryotic origins of eukaryotic fructose-1, 6-bisphosphatase and sedoheptulose-1, 7-bisphosphatase and the implications of their origins for the evolution of eukaryotic Calvin cycle. *BMC Evolutionary Biology* (**2012**), 12.

[297] **Raines CA, *et al.*** cDNA and gene-sequences of wheat chloroplast sedoheptulose-1,7-bisphosphatase reveal homology with fructose-1,6-bisphosphatases. *Eur. J. Biochem.* (**1992**), 205:1053-9.

[298] **Yoo JG, *et al.*** Analysis of the cbbf genes from *Alcaligenes eutrophus* that encode fructose-1,6-/sedoheptulose-1,7-bisphosphatase. *Curr. Microbiol.* (**1995**), 31:55-61.

[299] **Hines JK, *et al.*** Structures of mammalian and bacterial fructose-1,6-bisphosphatase reveal the basis for synergism in AMP/fructose 2,6-bisphosphate inhibition. *J. Biol. Chem.* (**2007**), 282:36121-31.

[300] **Pilkis SJ, *et al.*** Inhibition of fructose-1,6-bisphosphatase by fructose 2,6-bisphosphate. *J. Biol. Chem.* (**1981**), 256:3619-22.

[301] **Ke HM, *et al.*** Crystal-structure of fructose-1,6-bisphosphatase complexed with fructose 6-phosphate, AMP and magnesium. *Proc. Natl. Acad. Sci. U. S. A.* (**1990**), 87:5243-7.

[302] **Hines JK, *et al.*** Structure of inhibited fructose-1,6-bisphosphatase from *Escherichia coli* - Distinct allosteric inhibition sites for AMP and glucose 6-phosphate and the characterization of a gluconeogenic switch. *J. Biol. Chem.* (**2007**), 282:24697-706.

[303] **Hines JK, *et al.*** Novel allosteric activation site in *Escherichia coli* fructose-1,6-bisphosphatase. *J. Biol. Chem.* (**2006**), 281:18386-93.

[304] **Hines JK, *et al.*** Structures of activated fructose-1,6-bisphosphatase from *Escherichia coli* - Coordinate regulation of bacterial metabolism and the conservation of the R-state. *J. Biol. Chem.* (**2007**), 282:11696-704.

[305] **Choe JY, *et al.*** Crystal structures of fructose 1,6-bisphosphatase: Mechanism of catalysis and allosteric inhibition revealed in product complexes. *Biochem. (Mosc.)* (**2000**), 39:8565-74.

[306] **Brown G, *et al.*** Structural and biochemical characterization of the type II fructose-1,6-bisphosphatase GlpX from *Escherichia coli*. *J. Biol. Chem.* (**2009**), 284:3784-92.

[307] **Tamoi M, *et al.*** Acquisition of a new type of fructose-1,6-bisphosphatase with resistance to hydrogen peroxide in cyanobacteria: molecular characterization of the enzyme from *Synechocystis* PCC 6803. *Biochim. Biophys. Acta-Protein Structure and Molecular Enzymology* (**1998**), 1383:232-44.

[308] **Yan C, *et al.*** Bifunctional enzyme FBPase/SBPase is essential for photoautotrophic growth in cyanobacterium *Synechocystis* sp PCC 6803. *Progress in Natural Science* (**2008**), 18:149-53.

[309] **Rittmann D, *et al.*** Fructose-1,6-bisphosphatase from *Corynebacterium glutamicum*: expression and deletion of the *fbp* gene and biochemical characterization of the enzyme. *Arch. Microbiol.* (**2003**), 180:285-92.

[310] **Movahedzadeh F, *et al.*** The *Mycobacterium tuberculosis* Rv1099c gene encodes a GlpX-like class II fructose 1,6-bisphosphatase. *Microbiology* (**2004**), 150:3499-505.

[311] **Jules M, *et al.*** The *Bacillus subtilis ywjI* (glpX) gene encodes a class II fructose-1,6-bisphosphatase, functionally equivalent to the class III FBP enzyme. *J. Bacteriol.* (**2009**), 191:3168-71.

[312] **Fujita Y, *et al.*** Purification and properties of fructose-1,6-bisphosphatase of *Bacillus subtilis*. *J. Biol. Chem.* (**1979**), 254:5340-9.

[313] **Say RF, *et al.*** Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* (**2010**), 464:1077-U156.

[314] **Lee YG, *et al.*** Characterization of hyperthermostable fructose-1,6-bisphosphatase from *Thermococcus onnurineus* NA1. *Journal of Microbiology* (**2010**), 48:803-7.

[315] **Rashid N, *et al.*** A novel candidate for the true fructose-1,6-bisphosphatase in Archaea. *J. Biol. Chem.* (**2002**), 277:30649-55.

[316] **Du J, *et al.*** Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase. *Nature* (**2011**), 478:534-U134.

[317] **Fushinobu S, *et al.*** Structural basis for the bifunctionality of fructose-1,6-bisphosphate aldolase/phosphatase. *Nature* (**2011**), 478:538-41.

References

[318] **Nishimasu H, *et al.*** The first crystal structure of the novel class of fructose-1,6-bisphosphatase present in thermophilic archaea. *Structure* (**2004**), 12:949-59.

[319] **Cooper RA, *et al.*** Mechanism of phosphoenolpyruvate synthase reaction. *Biochim. Biophys. Acta* (**1967**), 141:211-&.

[320] **Wood HG, *et al.*** Properties of carboxytransphosphorylase; pyruvate, phosphate dikinase; pyrophosphate-phosphofructokinase and pyrophosphate-acetate kinase and their roles in the metabolism of inorganic pyrophospate. (**1977**).

[321] **Burnell JN.** Cloning and characterization of *Escherichia coli* DUF299: a bifunctional ADP-dependent kinase - P_i-dependent pyrophosphorylase from bacteria. *Bmc Biochemistry* (**2010**), 11.

[322] **Kotrba P, *et al.*** Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *Journal of Bioscience and Bioengineering* (**2001**), 92:502-17.

[323] **Schramm A, *et al.*** Pyruvate kinase of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: Physiological role and phylogenetic aspects. *J. Bacteriol.* (**2000**), 182:2001-9.

[324] **Larsen TM, *et al.*** Structure of rabbit muscle pyruvate kinase complexed with Mn^{2+}, K^+, and pyruvate. *Biochem. (Mosc.)* (**1994**), 33:6301-9.

[325] **Laughlin LT, *et al.*** The monovalent cation requirement of rabbit muscle pyruvate kinase is eliminated by substitution of lysine for glutamate 117. *Arch. Biochem. Biophys.* (**1997**), 348:262-7.

[326] **Oria-Hernandez J, *et al.*** Dichotomic phylogenetic tree of the pyruvate kinase family - K^+-dependent and -independent enzymes. *J. Biol. Chem.* (**2006**), 281:30717-24.

[327] **Jurica MS, *et al.*** The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure* (**1998**), 6:195-210.

[328] **Larsen TM, *et al.*** Structure of the Bis(Mg^{2+})-ATP-oxalate complex of the rabbit muscle pyruvate kinase at 2.1 angstrom resolution: ATP binding over a barrel. *Biochem. (Mosc.)* (**1998**), 37:6247-55.

[329] **Morgan HP, *et al.*** Allosteric mechanism of pyruvate kinase from *Leishmania mexicana* uses a rock and lock model. *J. Biol. Chem*. (**2010**), 285:12892-8.

[330] **Valentin G, *et al.*** The allosteric regulation of pyruvate kinase - A site-directed mutagenesis study. *J. Biol. Chem.* (**2000**), 275:18145-52.

[331] **Mattevi A, *et al.*** Crystal structure of *Escherichia coli* pyruvate kinase type I molecular basis of the allosteric transition. *Structure* (**1995**), 3:729-41.

[332] **Rigden DJ, *et al.*** The structure of pyruvate kinase from *Leishmania mexicana* reveals details of the allosteric transition and unusual effector specificity. *J. Mol. Biol.* (**1999**), 291:615-35.

[333] **Suzuki K, *et al.*** Crystal structure of pyruvate kinase from *Geobacillus stearothermophilus*. *J. Biochem. (Tokyo).* (**2008**), 144:305-12.

[334] **Cook WJ, *et al.*** Crystal Structure of *Cryptosporidium parvum* Pyruvate Kinase. *Plos One* (**2012**), 7.

[335] **Johnsen U, *et al.*** Comparative analysis of pyruvate kinases from the hyperthermophilic archaea *Archaeoglobus fulgidus*, *Aeropyrum pernix*, and *Pyrobaculum aerophilum* and the hyperthermophilic bacterium *Thermotoga maritima* - Unusual regulatory properties in hyperthermophilic archaea. *J. Biol. Chem.* (**2003**), 278:25417-27.

[336] **Muirhead H, *et al.*** The structure of cat muscle pyruvate-kinase. *EMBO J.* (**1986**), 5:475-81.

[337] **Tolentino R, *et al.*** Identification of the amino acid involved in the regulation of bacterial pyruvate, orthophosphate dikinase and phosphoenolpyruvate synthetase. *Adv. Biol. Chem.* (**2013**), 3.

[338] **Sakai H.** Possible structure and function of the extra C-terminal sequence of pyruvate kinase from *Bacillus stearothermophilus*. *J. Biochem. (Tokyo).* (**2004**), 136:471-6.

[339] **Herzberg O, *et al.*** Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proc. Natl. Acad. Sci. U. S. A.* (**1996**), 93:2652-7.

[340] **Carroll LJ, *et al.*** Substrate-binding domains in pyruvate, phosphate dikinase. *Biochem. (Mosc.)* (**1994**), 33:1134-42.

[341] **McGuire M, *et al.*** Determination of the nucleotide binding site within *Clostridium symbiosum* pyruvate phosphate dikinase by photoaffinity labeling, site-directed mutagenesis, and structural analysis. *Biochem. (Mosc.)* (**1996**), 35:8544-52.

[342] **McGuire M, *et al.*** Location of the phosphate binding site within *Clostridium symbiosum* pyruvate phosphate dikinase. *Biochem. (Mosc.)* (**1998**), 37:13463-74.

[343] **Ye DM, *et al.*** Investigation of the catalytic site within the ATP-grasp domain of *Clostridium symbiosum* pyruvate phosphate dikinase. *J. Biol. Chem.* (**2001**), 276:37630-9.

[344] **Lim K, *et al.*** Swiveling domain mechanism in pyruvate phosphate dikinase. *Biochem. (Mosc.)* (**2007**), 46:14845-53.

[345] **Xu Y, *et al.*** Location of the catalytic site for phosphoenolpyruvate formation within the primary structure of *Clostridium symbiosum* pyruvate phosphate dikinase .1. Identification of an essential cysteine by chemical modification with 1-C-14 bromopyruvate and site-directed mutagenesis. *Biochem. (Mosc.)* (**1995**), 34:2181-7.

[346] **Yankie L, *et al.*** Location of the catalytic site for phosphoenolpyruvate formation within the primary structure of *Clostridium symbiosum* pyruvate phosphate dikinase .2. Site-directed mutagenesis of an essential arginine contained within an apparent P-loop. *Biochem. (Mosc.)* (**1995**), 34:2188-94.

[347] **Nakanishi T, *et al.*** Crystal structures of pyruvate phosphate dikinase from maize revealed an alternative conformation in the swiveling-domain motion. *Biochem. (Mosc.)* (**2005**), 44:1136-44.

[348] **Ritte G, *et al.*** Phosphorylation of C6- and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. *FEBS Lett.* (**2006**), 580:4872-6.

[349] **Burnell JN, *et al.*** Cloning and expression of maize-leaf pyruvate, P_i dikinase regulatory protein gene. *Biochem. Biophys. Res. Commun.* (**2006**), 345:675-80.

[350] **Luoto HH, *et al.*** Membrane-integral pyrophosphatase subfamily capable of translocating both Na^+ and H^+. *Proc. Natl. Acad. Sci. U. S. A.* (**2013**), 110:1255-60.

[351] **Luoto HH, *et al.*** Na^+-translocating membrane pyrophosphatases are widespread in the microbial world and evolutionarily precede H^+-translocating pyrophosphatases. *J. Biol. Chem.* (**2011**), 286:21633-42.

[352] **Huang H, *et al.*** Divergence of structure and function in the haloacid dehalogenase enzyme superfamily: *Bacteroides thetaiotaomicron* BT2127 is an inorganic pyrophosphatase. *Biochem. (Mosc.)* (**2011**), 50:8937-49.

[353] **Lee HS, *et al.*** A novel inorganic pyrophosphatase in *Thermococcus onnurineus* NA1. *FEMS Microbiol. Lett.* (**2009**), 300:68-74.

[354] **Young TW, *et al.*** Bacillus subtilis ORF yybQ encodes a manganese-dependent inorganic pyrophosphatase with distinctive properties: the first of a new class of soluble pyrophosphatase? *Microbiology-Uk* (**1998**), 144:2563-71.

[355] **Cooperman BS, *et al.*** Evolutionary conservation of the active-site of soluble inorganic pyrophosphatase. *Trends Biochem. Sci.* (**1992**), 17:262-6.

[356] **Heikinheimo P, *et al.*** A site-directed mutagenesis study of *Saccharomyces cerevisiae* pyrophosphatase - Functional conservation of the active site of soluble inorganic pyrophosphatases. *Eur. J. Biochem.* (**1996**), 239:138-43.

[357] **Oksanen E, *et al.*** A complete structural description of the catalytic cycle of yeast pyrophosphatase. *Biochem. (Mosc.)* (**2007**), 46:1228-39.

[358] **Pohjanjoki P, *et al.*** Evolutionary conservation of enzymatic catalysis: Quantitative comparison of the effects of mutation of aligned residues in *Saccharomyces cerevisiae* and *Escherichia coli* inorganic pyrophosphatases on enzymatic activity. *Biochem. (Mosc.)* (**1998**), 37:1754-61.

[359] **Salminen T, *et al.*** Structure and function-analysis of *Escherichia coli* inorganic pyrophosphatase - is a hydroxide ion the key to catalysis. *Biochem. (Mosc.)* (**1995**), 34:782-91.

[360] **Samygina VR, *et al.*** The structures of *Escherichia coli* inorganic pyrophosphatase complexed with Ca^{2+} or $CaPP_i$ at atomic resolution and their mechanistic implications. *J. Mol. Biol.* (**2001**), 314:633-45.

[361] **Sivula T, *et al.*** Evolutionary aspects of inorganic pyrophosphatase. *FEBS Lett.* (**1999**), 454:75-80.

[362] **Gomez-Garcia MR, *et al.*** Comparative biochemical and functional studies of family I soluble inorganic pyrophosphatases from photosynthetic bacteria. *Febs Journal* (**2007**), 274:3948-59.

[363] **Andreeva N, *et al.*** Purification and properties of polyphosphatase from *Saccharomyces cerevisiae* cytosol. *Yeast* (**1998**), 14:383-90.

[364] **Ugochukwu E, *et al.*** The crystal structure of the cytosolic exopolyphosphatase from *Saccharomyces cerevisiae* reveals the basis for substrate specificity. *J. Mol. Biol.* (**2007**), 371:1007-21.

[365] **Ahn S, *et al.*** The "open" and "closed" structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii*. *J. Mol. Biol.* (**2001**), 313:797-811.

[366] **Tuominen H, *et al.*** Crystal structures of the CBS and DRTGG domains of the regulatory region of *Clostridium perfringens* pyrophosphatase complexed with the inhibitor, AMP, and activator, diadenosine tetraphosphate. *J. Mol. Biol.* (**2010**), 398:400-13.

[367] **Jamsen J, *et al.*** A CBS domain-containing pyrophosphatase of *Moorella thermoacetica* is regulated by adenine nucleotides. *Biochem. J.* (**2007**), 408:327-33.

[368] **Belogurov GA, *et al.*** A lysine substitute for K^+ - A460K mutation eliminates K^+ dependence in H^+-pyrophosphatase of *Carboxydothermus hydrogenoformans*. *J. Biol. Chem.* (**2002**), 277:49651-4.

[369] **Malinen AM, *et al.*** Na^+-pyrophosphatase: A novel primary sodium pump. *Biochem. (Mosc.)* (**2007**), 46:8872-8.

[370] **Serrano A, *et al.*** H^+-PPases: Yesterday, today and tomorrow. *Iubmb Life* (**2007**), 59:76-83.

[371] **Kellosalo J, *et al.*** The structure and catalytic cycle of a sodium-pumping pyrophosphatase. *Science* (**2012**), 337:473-6.

[372] **Lin S-M, *et al.*** Crystal structure of a membrane-embedded H^+-translocating pyrophosphatase. *Nature* (**2012**), 484:399-403.

[373] **Baltscheffsky M, *et al.*** H^+-proton-pumping inorganic pyrophosphatase: a tightly membrane-bound family. *FEBS Lett.* (**1999**), 452:121-7.

[374] **Nakanishi Y, *et al.*** Mutagenic analysis of functional residues in putative substrate-binding site and acidic domains of vacuolar H^+-pyrophosphatase. *J. Biol. Chem.* (**2001**), 276:7654-60.

[375] **Rea PA, *et al.*** Vacuolar H^+-translocating pyrophosphatases - a new category of ion translocase. *Trends Biochem. Sci.* (**1992**), 17:348-53.

[376] **Mimura H, *et al.*** Membrane topology of the H^+-pyrophosphatase of *Streptomyces coelicolor* determined by cysteine-scanning mutagenesis. *J. Biol. Chem.* (**2004**), 279:35106-12.

[377] **Velayudhan J, *et al.*** Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni*: an essential role for phosphoenolpyruvate carboxykinase. *Microbiology-Sgm* (**2002**), 148:685-94.

[378] **Punta M, *et al.*** The Pfam protein families database. *Nucleic Acids Res.* (**2012**), 40:D290-D301.

[379] **Zhou J, *et al.*** A-typical glycolysis in *Clostridium thermocellum*. *Appl. Environ. Microbiol.* (**2013**), 79:3000-8.

[380] **Sosa A, *et al.*** H^+/PP_i stoichiometry of membrane-bound pyrophosphatase of *rhodospirillum rubrum*. *Arch. Biochem. Biophys.* (**1995**), 316:421-7.

[381] **Lopez-Marques RL, *et al.*** Differential regulation of soluble and membrane-bound inorganic pyrophosphatases in the photosynthetic bacterium *Rhodospirillum rubrum* provides insights into pyrophosphate-based stress bioenergetics. *J. Bacteriol.* (**2004**), 186:5418-26.

[382] **Baykov AA, *et al.*** Pyrophosphate-fueled Na^+ and H^+ transport in prokaryotes. *Microbiol. Mol. Biol. Rev.* (**2013**), 77:267-76.

[383] **Beschastnyi AP, *et al.*** Activities of 6-phosphofructokinases and inorganic pyrophosphatase in aerobic methylotrophic bacteria. *Microbiology* (**2008**), 77:636-8.

[384] **Huang SB, *et al.*** Does anoxia tolerance involve altering the energy currency towards PP_i? *Trends in Plant Science* (**2008**), 13:221-7.

[385] **Pace DA, *et al.*** Overexpression of a cytosolic pyrophosphatase (TgPPase) reveals a regulatory role of PP_i in glycolysis for *Toxoplasma gondii*. *Biochem. J.* (**2011**), 440:229-40.

[386] **Claasen PAM.** 019825-(SES6), HYVOLUTION, Non-thermal production of pure hydrogen from biomass. *http://www.biohydrogen.nl/hyvolution/32288/9/0/20* (**2011**).

[387] **Mansilla C, *et al.*** CO_2-free hydrogen as a substitute to fossil fuels: What are the targets? Prospective assessment of the hydrogen market attractiveness. *Int. J. Hydrogen. Energy* (**2012**), 37:9451-8.

[388] **Whang L-M, *et al.*** Metabolic and energetic aspects of biohydrogen production of *Clostridium tyrobutyricum*: The effects of hydraulic retention time and peptone addition. *Bioresource Technol.* (**2011**), 102:8378-83.

[389] **Xing D, *et al.*** Continuous hydrogen production of auto-aggregative *Ethanoligenens harbinense* YUAN-3 under non-sterile condition. *Int. J. Hydrogen. Energy* (**2008**), 33:1489-95.

[390] **Taguchi F, *et al.*** Efficient hydrogen production from starch by a bacterium isolated from termites. *Journal of Fermentation and Bioengineering* (**1992**), 73:244-5.

[391] **Niu K, *et al.*** Characteristics of fermentative hydrogen production with *Klebsiella pneumoniae* ECU-15 isolated from anaerobic sewage sludge. *Int. J. Hydrogen. Energy* (**2010**), 35:71-80.

[392] **Yokoi H, *et al.*** Characteristics of hydrogen production by aciduric *Enterobacter aerogenes* strain HO-39. *Journal of Fermentation and Bioengineering* (**1995**), 80:571-4.

[393] **Wang X, *et al.*** Biochemical kinetics of fermentative hydrogen production by *Clostridium butyricum* W5. *Int. J. Hydrogen. Energy* (**2009**), 34:791-8.

[394] **Ren N, *et al.*** Dark fermentation of xylose and glucose mix using isolated *Thermoanaerobacterium thermosaccharolyticum* W16. *Int. J. Hydrogen. Energy* (**2008**), 33:6124-32.

[395] **Jayasinghearachchi HS, *et al.*** Biological hydrogen production by extremely thermophilic novel bacterium *Thermoanaerobacter mathranii* A3N isolated from oil producing well. *Int. J. Hydrogen. Energy* (**2012**), 37:5569-78.

[396] **Brenner K, *et al.*** Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* (**2008**), 26:483-9.

[397] **Liu Y, *et al.*** Hydrogen production from cellulose by co-culture of *Clostridium thermocellum* JN4 and *Thermoanaerobacterium thermosaccharolyticum* GD17. *Int. J. Hydrogen. Energy* (**2008**), 33:2927-33.

[398] **Pawar SS, *et al.*** Thermophilic biohydrogen production: how far are we? *Appl. Microbiol. Biotechnol.* (**2013**), 97:7999-8009.

[399] **Zhang Y, *et al.*** Hydrogen supersaturation in extreme-thermophilic (70 degrees C) mixed culture fermentation. *Appl. Energy* (**2013**), 109:213-9.

[400] **O-Thong S, *et al.*** High-rate continuous hydrogen production by *Thermoanaerobacterium thermosaccharolyticum* PSU-2 immobilized on heat-pretreated methanogenic granules. *Int. J. Hydrogen. Energy* (**2008**), 33:6498-508.

[401] **Lee KS, *et al.*** Continuous hydrogen production by anaerobic mixed microflora using a hollow-fiber microfiltration membrane bioreactor. *Int. J. Hydrogen. Energy* (**2007**), 32:950-7.

[402] **Oh SE, *et al.*** Biological hydrogen production using a membrane bioreactor. *Biotechnol. Bioeng.* (**2004**), 87:119-27.

[403] **Kalil MS, *et al.*** Effect of nitrogen source and carbon to nitrogen ratio on hydrogen production using *C. acetobutylicum*. *Am. J. of Biochem. Biotech.* (**2008**), 4:393-401.

[404] **Martinez-Porqueras E, *et al.*** Effect of medium composition on biohydrogen production by the extreme thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Int. J. Hydrogen. Energy* (**2013**), 38:11756-64.

[405] **Bielen AAM, *et al.*** Biohydrogen production by the thermophilic bacterium *Caldicellulosiruptor saccharolyticus*: Current status and perspectives. *Life* (**2013**), 3:52-85.

[406] **Tyurin MV, *et al.*** Electrotransformation of *Clostridium thermocellum*. *Appl. Environ. Microbiol.* (**2004**), 70:883-90.

[407] **Peng H, *et al.*** Electrotransformation of *Thermoanaerobacter ethanolicus* JW200. *Biotechnology Letters* (**2006**), 28:1913-7.

[408] **Santangelo TJ, *et al.*** Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* (**2011**), 81:897-911.

[409] **Sato T, *et al.*** Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.* (**2005**), 71:3889-99.

[410] **Lipscomb GL, *et al.*** Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* (**2011**), 77:2232-8.

[411] **Liu B, *et al.*** Establishment of a genetic transformation system and its application in *Thermoanaerobacter tengcongensis*. *Journal of Genetics and Genomics* (**2012**), 39:561-70.

[412] **Han DM, *et al.*** Construction and transformation of a *Thermotoga-E. coli* shuttle vector. *Bmc Biotechnology* (**2012**), 12.

[413] **Cha M, *et al.*** Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. *Biotechnology for Biofuels* (**2013**), 6.

[414] **Kanai T, *et al.*** Hydrogen production by the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Journal of the Japan Petroleum Institute* (**2013**), 56:267-79.

[415] **Lin L, *et al.*** Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production. *Biotechnology Advances* (**2013**), 31:827-37.
[416] **Vardar-Schara G, *et al.*** Metabolically engineered bacteria for producing hydrogen via fermentation. *Microbial Biotechnology* (**2008**), 1:107-25.
[417] **Yu RS, *et al.*** Blocking the butyrate-formation pathway impairs hydrogen production in *Clostridium perfringens*. *Acta Biochim. Biophys. Sinica* (**2013**), 45:408-15.
[418] **Conners SB, *et al.*** An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol.* (**2005**), 187:7267-82.
[419] **Watanabe S, *et al.*** Metabolic fate of L-lactaldehyde derived from an alternative L-rhamnose pathway. *Febs Journal* (**2008**), 275:5139-49.
[420] **Saikusa T, *et al.*** Metabolism of 2-oxoaldehydes in bacteria: purification and characterization of methylglyoxal reductase from *Escherichia coli*. *Agric. Biol. Chem.* (**1987**), 51:1893-9.
[421] **Rhee H, *et al.*** Metabolism of 2-ketoaldehydes in bacteria oxidative conversion of methylglyoxal to pyruvate by an enzyme from *Pseudomonas putida*. *Agric. Biol. Chem.* (**1987**), 51:1059-66.
[422] **Sun JS, *et al.*** Heterologous expression and maturation of an NADP-dependent NiFe -hydrogenase: A key enzyme in biofuel production. *Plos One* (**2010**), 5.
[423] **Nogales J, *et al.*** An in silico re-design of the metabolism in *Thermotoga maritima* for increased biohydrogen production. *Int. J. Hydrogen. Energy* (**2012**).
[424] **Meyer J.** [FeFe] hydrogenases and their evolution: a genomic perspective. *Cell. Mol. Life Sci.* (**2007**), 64:1063-84.
[425] **Vignais PM, *et al.*** Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* (**2001**), 25:455-501.
[426] **Shima S, *et al.*** A third type of hydrogenase catalyzing H_2 activation. *Chemical Record* (**2007**), 7:37-46.
[427] **Vignais PM, *et al.*** Occurrence, classification, and biological function of hydrogenases: An overview. *Chem. Rev.* (**2007**), 107:4206-72.
[428] **Meuer J, *et al.*** Purification and catalytic properties of Ech hydrogenase from *Methanosarcina barkeri*. *Eur. J. Biochem.* (**1999**), 265:325-35.
[429] **Meuer J, *et al.*** Genetic analysis of the archaeon *Methanosarcina barkeri* Fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation. *Proc. Natl. Acad. Sci. U. S. A.* (**2002**), 99:5632-7.
[430] **Schuchmann K, *et al.*** A bacterial electron-bifurcating hydrogenase. *J. Biol. Chem.* (**2012**), 287:31165-71.
[431] **Poehlein A, *et al.*** An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *Plos One* (**2012**), 7.
[432] **Wang SN, *et al.*** A reversible electron-bifurcating ferredoxin- and NAD-dependent FeFe -hydrogenase (HydABC) in *Moorella thermoacetica*. *J. Bacteriol.* (**2013**), 195:1267-75.
[433] **Huang H, *et al.*** Electron-bifurcation Involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H_2 plus CO_2. *J. Bacteriol.* (**2012**), 194:3689-99.
[434] **Stams AJM.** Metabolic interactions between anaerobic-bacteria in methanogenic environments. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* (**1994**), 66:271-94.
[435] **Maru BT, *et al.*** Glycerol fermentation to hydrogen by *Thermotoga maritima*: Proposed pathway and bioenergetic considerations. *Int. J. Hydrogen. Energy* (**2013**), 38:5563-72.
[436] **Stec B, *et al.*** MJ0109 is an enzyme that is both an inositol monophosphatase and the 'missing' archaeal fructose1,6-bisphosphatase. *Nat. Struct. Biol.* (**2000**), 7:1046-50.
[437] **Mertens E.** ATP versus pyrophosphate: Glycolysis revisited in parasitic protists *Parasitology Today* (**1993**), 9:122-6.
[438] **Feist AM, *et al.*** Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology* (**2009**), 7:129-43.
[439] **Joyce AR, *et al.*** The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology* (**2006**), 7:198-210.
[440] **Kauffman KJ, *et al.*** Advances in flux balance analysis. *Curr. Opin. Biotechnol.* (**2003**), 14:491-6.
[441] **Lee JM, *et al.*** Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics* (**2006**), 7:140-50.
[442] **Zeidan AA.** Hydrogen production by *Caldicellulosiruptor* species. *Doctoral Thesis, Division of Applied Micobiology, Faculty of Engineering, Lund University* (**2011**).
[443] **Jones PR.** Improving fermentative biomass-derived H_2 production by engineering microbial metabolism. *Int. J. Hydrogen. Energy* (**2008**), 33:5122-30.
[444] **Veit AM, *et al.*** Constructing and testing the thermodynamic limits of synthetic NAD(P)H:H_2 pathways. *Microbial Biotechnology* (**2008**), 1:382-94.
[445] **Atomi H, *et al.*** Application of hyperthermophiles and their enzymes. *Curr. Opin. Biotechnol.* (**2011**), 22:618-26.
[446] **Bhalla A, *et al.*** Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technol.* (**2013**), 128:751-9.
[447] **Woodward J, *et al.*** In vitro hydrogen production by glucose dehydrogenase and hydrogenase. *Nat. Biotechnol.* (**1996**), 14:872-4.
[448] **Woodward J, *et al.*** Biotechnology - Enzymatic production of biohydrogen. *Nature* (**2000**), 405:1014-5.
[449] **Ye XH, *et al.*** Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. *Chemsuschem* (**2009**), 2:149-52.

Dutch summary – Nederlandse samenvatting

Waterstof (H_2) is sinds haar officiële ontdekking in de 18de eeuw uitgegroeid tot een zeer belangrijke bulkchemicalie, met toepassingen in diverse processen binnen de glas-, voedsel-, elektronica- en metaal- industrie. Daarnaast wordt H_2 in grote hoeveelheden gebruikt tijdens de raffinage van olie en vormt het een belangrijk basisbestanddeel voor de productie van andere bulkchemicaliën zoals methanol en ammoniak. H_2 kan ook direct worden gebruikt als brandstof, in een proces waarbij alleen water ontstaat. In tegenstelling tot de verbranding van fossiele brandstoffen, wordt er bovendien geen broeikasgas (CO_2) gevormd. Naast de genoemde toepassingen kan H_2 ook fungeren als algemene energiedrager, dat wil zeggen dat H_2 kan worden gebruikt voor het opslaan van energie en als transportmedium. Op basis van laatstgenoemde eigenschap is H_2 door sommigen naar voren geschoven als de centrale energiedrager in een toekomstige waterstofeconomie.

Echter, op dit moment is meer dan 90% van de mondiale H_2-productie gebaseerd op het gebruik van fossiele brandstoffen. Gezien de negatieve effecten van het gebruik van fossiele brandstoffen (bv. uitstoot van het broeikasgas CO_2) en het niet hernieuwbare karakter van deze grondstof, wordt de productie van H_2 uit fossiele brandstoffen niet als een duurzaam proces beschouwd. Om deze reden richten diverse onderzoeken zich op de ontwikkeling van alternatieve H_2-productieprocessen die gebaseerd zijn op hernieuwbare grondstoffen.

In een biologisch H_2-productieproces (bioH_2), ook wel aangeduid als 'donker-fermentatie', worden fermentatieve micro-organismen gebruikt om H_2 te genereren uit koolwaterstofrijke industriële afvalstromen of suikerrijke biomassa (bijv. planten). Hierbij gaat de voorkeur uit naar het gebruik van biomassa-typen die niet in competitie zijn met voedselproductie.

Sommige thermofiele anaërobe micro-organismen bezitten bepaalde eigenschappen die er voor zorgen dat deze organismen uitermate geschikt zijn voor gebruik in een op donker-fermentatie gebaseerd bioH_2-productieproces (**hoofdstuk 1**). Zo beschikken de thermofiele bacteriën *Caldicellulosiruptor saccharolyticus* en *Thermotoga maritima* over de capaciteit om complexe biomassa af te breken tot makkelijk fermenteerbare di- of mono-sachariden (suikers). Bovendien leidt de stofwisseling van deze organismen slechts tot een beperkt aantal eindproducten, waarbij de afbraakroute die de meeste ATP (een belangrijke interne energiebron van de organismen) oplevert ook nog eens resulteert in de hoogsteH_2-opbrengst. Met andere woorden, deze micro-organismen willen het liefst H_2 maken, en ze zijn er nog goed in ook! Onder ideale omstandigheden zetten deze thermofielen eenvoudige suikers om in H_2, azijnzuur en CO_2 met een H_2-opbrengst die de theoretisch maximaal haalbare opbrengst van 4 H_2 per hexose benadert.

Het onderzoek beschreven in dit proefschrift richt zich hoofdzakelijk op de moleculaire mechanismen die zich afspelen binnen de H_2-vormende thermofiele anaërobe bacteriën *Caldicellulosiruptor saccharolyticus* en *Thermotoga maritima*. Een beter begrip inzake i) de metabole routes betrokken bij de afbraak van koolwaterstoffen ii) de mechanismen van recycling van redox-carriers en iii) de invloed van verschillende groeicondities op de fermentatie van deze micro-organismen geeft een dieper inzicht in de factoren die de H_2-opbrengst van de huidige, op donker-fermentatie gebaseerde, bioH_2-productie processen limiteren. Bevindingen die ons een stap dichterbij een grootschalige industriële toepassing van donker-fermentatie brengen.

Ruwe glycerol wordt in grote hoeveelheden gevormd als bijproduct tijdens biodieselproductie. Vanwege de relatieve hoge gereduceerde staat van koolstof in glycerol is deze goedkope grondstof aantrekkelijk voor de productie van hernieuwbare brandstoffen, inclusief bioH_2. De data gepresenteerd in **hoofdstuk 2** laten voor het eerst zien dat *T. maritima* in staat is op glycerol te groeien in zowel batch- als chemostaat- fermentaties. *T. maritima* fermenteert glycerol voornamelijk tot azijnzuur, CO_2 en H_2. De H_2-opbrengst van deze omzetting nadert de theoretische limiet van 3 H_2 per glycerol, een rendement dat gemiddelde 3 maal zo hoog is als de opbrengst waargenomen voor mesofiele processen! De waargenomen H_2-opbrengst suggereert dat de elektronen die vrijkomen tijdens de oxidatie van glycerol 3-phosphate uiteindelijk worden overgedragen naar H_2, maar het exacte mechanisme voor elektronenoverdracht is nog niet bekend. Dit hoofdstuk presenteert een mogelijke route voor glycerolafbraak en het betrokken mechanisme voor elektronenoverdracht in *T. maritima*. De analyse van beschikbare genomen suggereert dat de mogelijkheid tot groei op glycerol als een algemene eigenschap van alle *Thermotoga*-soorten kan worden beschouwd.

In een donker-fermentatie proces wordt de vorming van H_2 geremd door verhoogde H_2-niveaus, wat betekent dat ophoping van H_2 in een systeem een verlaging van de totale H_2-opbrengst te weeg brengt. In **hoofdstuk 3** is de reactie van *C. saccharolyticus* op blootstelling aan verhoogde H_2-niveaus onderzocht in glucose-gelimiteerde chemostaatcultivaties. Diverse groeicondities zijn onderzocht om een onderscheid te kunnen maken tussen de invloed van i) roersnelheid, ii) type reactor-flushing en iii) type flushinggas op de H_2 vormende capaciteiten van *C. saccharolyticus*. Het blijkt dat onder vergelijkbare groeicondities, met een hoge roersnelheid, het spoelen van de reactor met H_2-gas ongeveer even efficiënt is als spoelen met N_2-gas. Deze waarneming bevestigt dat de opgeloste H_2-concentratie de bepalende factor is onder dergelijke condities en niet de H_2-concentratie in de gasfase. Gepresenteerde transcriptiedata laten zien dat, als reactie op verhoogde H_2-niveaus, de transcriptie niveaus van genen die betrokken zijn bij de recycling van gegenereerde reductie-

equivalenten worden verhoogd (transcriptie up-regulatie), zoals bijvoorbeeld de genen die coderen voor lactaat-dehydrogenase, alcohol-dehydrogenase en de NADH- en ferredoxine- afhankelijke hydrogenasen. Deze bevindingen komen overeen met de waargenomen veranderingen in het fermentatieprofiel van voornamelijk azijnzuur en H_2 onder lage H_2-niveaus naar azijnzuur, melkzuur, ethanol en H_2 onder verhoogde H_2-niveaus. Daarnaast zijn veranderingen in de relatieve transcriptie-niveaus waargenomen voor genen gerelateerd aan het centrale koolstofmetabolisme, vetzuurbiosynthese en diverse trans-membraan-transportsystemen. Tevens levert het onderzoek bewijs voor de betrokkenheid van specifieke transcriptie-regulatoren onder condities van verhoogde H_2-spanning, zoals de redox-toestand-gevoelige Rex-regulator, de ijzer-opname-regulator Fur en de vetzuurbiosynthese-regulator FapR.

In **hoofdstuk 4** wordt de invloed van verhoogde H_2 concentraties op eindproduct-vorming in *C. saccharolyticus* onderzocht met betrekking tot groei op i) ammonium-vrij medium en ii) een lage versus hoge glucoseconcentratie. Daarnaast wordt de route van rhamnose-fermentatie en de betrokken mechanismen van recycling van redox-carriers besproken. Een toename in substraat-concentratie (glucose) gedurende chemostaatcultivaties onder verhoogde H_2-niveaus leiden duidelijk tot een toename in de melkzuur-vorming terwijl de ethanolproductie gelijk blijft. Dit resultaat laat duidelijk zien dat voor *C. saccharolyticus* de vorming van melkzuur het primaire mechanisme is om redox-stress af te wenden. Het achterwegen laten van NH_4^+ in het groeimedium resulteerde in onvolledige substraatomzetting tijdens groei in een chemostaat onder verhoogde H_2-niveaus. De fermentatieprofielen van *C. saccharolyticus* tijdens groei op rhamnose onder de aan-/afwezigheid van CO (een remmer van hydrogenase-activiteit) suggereren de overdracht van elektronen tussen gereduceerd ferredoxine en $NAD(P)^+$. Echter, er is tot dusver geen gen in het genoom van *C. saccharolyticus* geïdentificeerd dat codeert voor een eiwit dat deze reactie zou kunnen katalyseren. De mogelijkheid tot groei op rhamnose voor verschillende *Caldicellulosiruptor*-soorten is sterk gecorreleerd met de aanwezigheid van twee specifieke genclusters. Verder geeft **hoofdstuk 4** een uitgebreid overzicht van i) de hydrolytische capaciteiten (suikerafbraak), ii) het suikermetabolisme, iii) H_2-vormende capaciteiten en iv) het mechanisme betrokken bij de gemengde zuren-productie van *C. saccharolyticus*.

Hoofdstuk 5 beschrijft het onderzoek naar de rol van pyrofosfaat (PP_i) in het centrale metabolisme van *C. saccharolyticus* als zowel energiedrager (fosforyl-donor) als enzymremmer. In overeenkomst met voorspellingen op basis van het geannoteerde genoom, laten cel-extracten zowel pyruvaat-kinase, pyruvaat-fosfaat-dikinase, en ATP- als PP_i- afhankelijke fosfofructokinase-activiteit zien. Daarnaast is er alleen pyrofosfatase-activiteit waar te nemen in de membraanfractie van het celextract, terwijl er geen significante pyrofosfatase-activiteit kan worden gedetecteerd in de cytosol-fractie. De aanwezig van PP_i- afhankelijke activiteit van

pyruvaat-fosfaat-dikinase en PP_i-fosfofructokinase in *C. saccharolyticus* gegroeid op glucose, suggereren een katabole rol voor deze enzymen. Deze enzymen vormen parallelle routen voor koolstofmetabolisme in *C. saccharolyticus* analoog met de klassieke glycolytische enzymen pyruvaat-kinase en ATP-fosfofructokinase. De clustering van het pyruvaat-fosfaat-dikinase-gen samen met alle genen die coderen voor de enzymen betrokken bij de C3 tak van de Embden-Meyerhof-Parnas-route voor glycolyse (met uitzondering van pyruvaat-kinase) lijkt een unieke eigenschap van *Caldicellulosiruptor* soorten. Onderzoek gericht op de dynamiek van de PP_i-niveaus in *C. saccharolyticus* tijdens groei op glucose laten zien dat tijdens de exponentiële groeifase relatief hoge PP_i-niveaus zijn waar te nemen terwijl een relatief laag PP_i-nivieau samenvalt met de overgang van exponentiële naar stationaire fase. Daarnaast heeft PP_i een remmende werking op de enzymactiviteit van pyruvaat-kinase. Al met al laat de gepresenteerde data zien dat er een belangrijke rol voor PP_i is weggelegd in het centrale koolstof- en energiemetabolisme van *C. saccharolyticus*.

Betrokkenheid van pyrofosfaat als fosforyl-donor in de glycolyse kan tot een energetisch voordeel leiden voor de cel. In de literatuur zijn verschillende vermeldingen te vinden over de betrokkenheid van PP_i in zowel glycolyse als gluconeogenese, i.e. op het niveau van PP_i-afhankelijk fosfofructokinase en pyruvaat-fosfaat-dikinase, maar het is onbekend of dit slechts sporadisch voorkomt of dat het als een algemeen fenomeen moet worden beschouwd. De vergelijkende genoomanalyse gepresenteerd in **hoofdstuk 6** geeft een overzicht van de verdeling van PP_i-afhankelijke fosfofructokinase en pyruvaat-fosfaat-dikinase, en de gerelateerde enzym-subtypen, over de genomen van 70 archaea, 30 eukarya en 395 bacteria. Daarnaast zijn de correlaties (het wel of niet gemeenschappelijk aanwezig zijn van de coderende genen in een genoom) tussen PP_i-afhankelijke fosfofructokinase en pyruvaat-fosfaat-dikinase onderzocht met betrekking tot i) elkaar, ii) de aanwezigheid van cytosol- en membraangebonden- pyrofosfatase-coderende genen en iii) genen van de klassieke glycolyse/gluconeogenese gerelateerde enzymen: ATP-afhankelijk fosfofructokinase, fructose-1,6-bisfosfatase, pyruvaat-kinase en pyruvaat-water-dikinase. Een, op COG-classificatie gebaseerde, *ab initio* enzym-subtypen-classificatie identificeerde naast de reeds bekende enzym-subtypen enkele nieuwe enzym-subtypen. De helft van de onderzochte genomen bevatten genen die coderen voor de niet-klassieke glycolyse/gluconeogenese gerelateerde enzymen, PP_i-afhankelijke fosfofructokinase en pyruvaat-fosfaat-dikinase. Daarnaast laten de waargenomen distributiepatronen een correlatie tussen de simultane aanwezigheid van de niet-klassieke PP_i- afhankelijke enzymen en de membraangebonden pyrofosfatase zien. De analyse toont aan dat de betrokkenheid van PP_i in glycolyse/gluconeogenese moet worden beschouwd als een wijdverspreid fenomeen dat is waar te nemen in diverse organismen uit de drie verschillende domeinen van het leven.

Acknowledgements – Dankwoord

Wetenschap is door de bomen het bos zien, en ik hou van wandelen. Op de middelbare school werd het me al snel duidelijk dat wetenschap zich uitstekend leent voor lange zwerftochten. Via een bergje hier en een dalletje daar, Nijmegen passerend en dwars door de Overbetuwe, belandde ik uiteindelijk bovenop de berg. Een omgeving waar ik me direct op mijn plek voelde. Dat een onderzoek niet altijd de uitgezette paden volgt is inherent aan wetenschap. Anderzijds, de kans om af en toe de gele (of gouden) weg te verlaten en een kijkje dieper in het bos te nemen kan ik nu eenmaal niet links laten liggen. Hoe dan ook, aan het eind van mijn tocht kan ik met zekerheid zeggen dat mijn pad mij veel voldoening heeft gebracht. Daarom kijk ik nu met gezonde weemoed en vol goede herinneringen achterom naar mijn tijd bij Micro, en daarbij ben ik alle mensen die mij hebben geholpen, in welke vorm dan ook, zeer dankbaar. Goed. Genoeg metaforisch gewauwel, tijd voor mijn dankwoord.

Beste Servé, ik ben jou als mijn dagelijkse begeleider, de meeste dank verschuldigd. Vanaf mijn eerste publicatie tot aan de laatste letter van dit proefschrift, heb je mij altijd het gevoel gegeven dat je je betrokken voelde bij mijn onderzoek. Ik kon altijd bij je terecht met mijn vragen. Weliswaar kwam ik vaak met meer vragen uit je kamer dan dat ik ermee naartoe nam, maar, je hebt me zelden bij een bushokje laten staan. Bovendien, als jij er niet was geweest lag ik waarschijnlijk nog steeds in een hotelkamer op de Azoren te slapen. Verder hoop ik dat ik je niet te veel heb gepijnigd met mijn soms complexe manier van dingen op papier zetten ("wat wil je hier nu eigenlijk mee zeggen?"). Ik heb met veel plezier met je samengewerkt, bedankt.

John, hoewel je uitspraak "Als je je een middagje kwaad maakt heb je het zo gedaan" voor mij niet altijd opging, weerspiegelt het wel je toewijding en je directe manier van wetenschap bedrijven. Het feit dat je op een internationaal congres in een korte broek en sandalen stond te presenteren (Extremophiles 2010 Azoren, ik heb foto's) demonstreerde voor mij waar het eigenlijk allemaal om draait… no-nonsens wetenschap. Dank voor deze belangrijke inzichten, in het wetenschapper-zijn. Uiteraard ook veel dank voor je bijdragen aan de stroomlijning van mijn manuscripten.

Willem, jouw immer tentoongestelde efficiëntie en doelgerichtheid zijn belangrijke leerpunten voor elke beginnende PhD-student, zo ook voor mij. Ik heb je enthousiasme en steun rond de PhD-studiereizen, met name de studiereis naar China en Japan in 2011, zeer gewaardeerd. Dank voor het delen van je wetenschappelijke inzichten.

Hierbij wil ik jullie alle drie ook nadrukkelijk bedanken voor het faciliteren van een werkplek tijdens de afrondende fase van mijn promotieonderzoek.

Most of the work presented in this thesis resulted from a diversity of international collaborations. A personal thank you to all the co-authors for your scientific contribution, which eventually led to the publication of our joined work.

From Lund University, Lund, Sweden; Karin Willquist, Ed van Niel and Jakob Engman. Karin and Ed it was a pleasure meeting you in person during your visits to the Laboratory of Microbiology.

From North Carolina State University, Raleigh (NC), USA; Sara Blumer-Schuette, Robert Kelly and Amy vanFossen. Sara thanks for the e-mail discussions concerning the technical aspects of the micro-array data set.

From Universitat Rovira i Virgili, Tarragona, Spain; Biniam Maru, Francesc Medina and Magda Constanti. Biniam, our close collaboration resulted in two nice papers; it was a real pleasure to meet you and many thanks for your lessons in Ethiopian culture and Ethiopian food of course!

Ook ben ik mijn co-auteurs bij Microbiologie, Marcel Verhaart en Fons Stams erg dankbaar. Fons onze koffietafelgesprekken over *Caldi*, *Thermotoga* en natuurlijk waterstofproductie waren zeer aangenaam. Marcel, ook al was je enigszins verbaasd je eigen naam op de publicaties te zien verschijnen, sommige stukken van mijn werk borduurden voort op delen van jouw werk en, ere wie ere toekomt, jij bent degene die me met *Caldi* heeft leren werken, dus bedankt.

Furthermore, a big THANK YOU to all the people from the Laboratory of Microbiology who helped to create a fruitful working environment and a nice atmosphere. A special thanks to the people who worked at the fermentor lab: Marco, Marcel, Katrin, Wouter, John R. en Tom. Katrin and Wouter thanks for sharing the frustrations concerning the metabolomics work. Bedankt kamergenoten jong en oud: Mark L., Marco, John R., Marcel, Amos, Elleke en Tom (ook wel bekent als Dr Love, Jut en Jul, Knabbel en Babbel, Amos, Ester en Oscar) voor de sfeer in de kamer, aan de koffietafel en tijdens de lunchwandelingetjes. Mark L. and PierPaolo, many thanks for the help with some of the molecular biology and FPLC work, unfortunately for me that work didn't end up in this thesis. Mark D. en Erik veel dank voor jullie inzet tijdens jullie thesis-periode als MSc-studenten onder mijn supervisie, ik heb veel van jullie geleerd. Al is het meeste werk wat we samen hebben gedaan, helaas, niet in dit proefschrift beland. Also a special thanks to Mark M., Edze (& Stineke, veel dank voor jullie ongelimiteerde gastvrijheid), Matthijs, Pawel, Nicolas, Bart, Vincent, Bas S., Janine, Flavia, Farrakh, Christian, Martin, Peer, Detmer and Thomas 'the lone ranger' Kruse (I particularly enjoyed our joint dinners at the lab after 'working' hours discussing whatever topic popped up and many thanks for your Danish hospitality).

Also thanks to the organisers and participants of the MIB related social activities, including the PhD study travels, Veluweloop, We-day, FC Kaas matches, MicroMovies sessions, Oud-Micro bijeenkomsten, and board game congregations. Thanks to my co-

organizers of the 2011 PhD trip to China and Japan: Edze, Martin, Mauricio, Teun, Jing and Jimmy, what a time! Sjon, Erwin, Matthijs, John vd O., Edze, Thomas en Soulman Peer, dank voor het met jullie kunnen delen van een van mijn grootste hobby's, muziek.

Uiteraard geen werkbare omgeving zonder hulp van het ondersteunend personeel. Anja en Carolien bedankt voor de adviezen rondom, en het regelen van, alle administratieve zaken. Jannie, bedankt voor al het schone lab-glaswerk. Ton bedankt voor de hulp bij het werk met de HPLC's en GC's. Sjon dank voor het regelen van alle bestellingen. Wim, ontzettend bedankt voor je duizend handen.

Zeer dankbaar ben ik ook voor de morele ondersteuning buiten de muren van Micro. Ik zou graag alle leden van karateverenging Chikara willen bedanken voor de fysieke uitdagingen en sociale aangelegenheden. Roman en Erik van Malignant Vision (R.I.P), dank voor de melodieuze death metal en de sporadische filosofische biertjes bij café Jos. Leo, dank voor PhD-perikelen gerelateerde gesprekken en uiteraard de 20 jaar voortdurende Duin-samenspraken. Vincent, Denise, Arend en Michiel als oud-studiegenoten en inmiddels oud-PhD-lotgenoten zijn onze bijeenkomsten me zeer waardevol geworden. Jop, dank je voor het aanhoudend aanbieden van hulp voor het tot succes volbrengen van mijn promotie onderzoek, sorry, maar ik moest het echt zelf doen. Robbert, veel dank voor alle activiteiten waarbij ik mijn poefschrift even uit het hoofd kon zetten, hierin ben je een ware, ik mag wel zeggen duistere meester.

Ook wil ik graag mijn directe familie bedanken. Broer Bas, Trudy, Albert, Tini en natuurlijk Vader B. en Mama (ik kan hier toch moeilijk moemoe neer zetten, niet?). Nadat ik in 2009 eindelijk mijn officiële zelfstandigheidsverklaring kreeg (heus waar, ik heb hem nog) sloeg ik op werkgebied definitief een pad in dat moeilijk voor jullie was te volgen. Misschien dat dit boekje het een en ander kan ophelderen, zo niet (Bas heeft in ieder geval, na het controleren van de introductie, al aangegeven het nooit meer te willen zien) dan mogen jullie me altijd blijven 'lastigvallen' met vreemde vragen.

Lieve Yorma, als ik alle steun die je me tijdens mijn promotieonderzoek hebt gegeven, hier zou moeten verwoorden, dan zouden de printkosten van dit boekje de pan uit rijzen. Dus dat doe ik niet, ik hou het kort. Ik wil je bedanken voor al die keren dat je me naar mijn werk hebt 'laten' gaan, maar vooral ook voor al die keren dat je me thuis hebt 'laten' blijven.

Bram

About the author

Abraham (Bram) Antonius Maria Bielen was born on the 3rd of April 1981, in Venray, the Netherlands. In 2000 he obtained his pre-university secondary education diploma (VWO) from 'Het Dendron college' in Horst. Led by his general interest in physics, chemistry, biology and mathematics he joined the multidisciplinary study-program Natural Sciences, at the Radboud University Nijmegen. At the department of Biophysics (Radboud University Nijmegen) he performed his BSc internship (2004), investigating the contribution of the vestibular organ to the vertical perception of humans. After his BSc he continued with his master (research variant), focussing on the interdisciplinary area between physics and biology. In 2006 he performed his internship at the Applied Neurophysiology research group (UMC St. Radboud Nijmegen) during which he performed in silico analysis of possibilities to enhance the representation of individual muscle activities in the surface electro-myographic signal. During his research thesis (2007), performed at the department of Plant Cell Biology (Radboud University Nijmegen), he focussed on the construction of promoter-reporter gene constructs for key fruit development studies in *Solanum lycopersicum* (tomato). In 2008 he started his PhD research in the Bacterial Genetics group, within the Laboratory of Microbiology at Wageningen University, mainly focussing on redox-carrier recycling, energy metabolism and metabolic network regulation in thermophilic hydrogen producing bacteria. Currently, he works as a Postdoc at the Laboratory of Organic Chemistry, Wageningen University, on the development of a biosensor.

List of publications

The representation of muscles in surface EMG: results from an electrophysiological muscle model and a biomechanical trunk model. D.F. Stegeman, D. Staudenmann, **B. Bielen**, J.P. Dijk, J.H. van Dieen. In R. Grieshaber, M. Stadeler & H.C. Scholle (Eds.), Prävention von arbeistbedingten Gesundheitsgefahren und Erkrankungen, 2009: 237-242.

A.A.M. Bielen#, K. Willquist#, J. Engman, J. van der Oost, E.W.J. van Niel, S.W.M. Kengen. Pyrophosphate as a central energy carrier in the hydrogen producing extremely thermophilic *Caldicellulosiruptor saccharolyticus*. FEMS Microbiology Letters, 2010 307(1): 48-54.
doi: 10.1111/j.1574-6968.2010.01957.x

M.R.A. Verhaart, **A.A.M. Bielen**, J. van der Oost, A.J.M. Stams, S.W.M.Kengen. Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. Environmental technology, 2010 31(8-9): 993-1003.
doi: 10.1080/09593331003710244

Regulation of tomato fruit pericarp development by an interplay between CDKB and CDKA1 cell cycle genes. A. Czerednik, M. Busscher, **B.A.M. Bielen**, M. Wolters-Arts, R.A. de Maagd and G.C. Angenent. Journal of Experimental Botany. 2012 63 (7): 2605-2617.
doi: 10.1093/jxb/err451

B.T. Maru, **A.A.M. Bielen**, S.W.M. Kengen, M. Constantí, F. Medina. Biohydrogen production from glycerol using *Thermotoga* spp. Energy Procedia, 2012 29: 300–307.
doi:10.1016/j.egypro.2012.09.036

A.A.M. Bielen, M.R.A. Verhaart, J. van der Oost and S.W.M. Kengen. Biohydrogen production by the thermophilic bacterium *Caldicellulosiruptor saccharolyticus*: current status and perspectives. Life, 2013 3(1): 52-85.
doi:10.3390/life3010052

A.A.M. Bielen, M.R.A. Verhaart, A.L. VanFossen, S.E. Blumer-Schuette, A.J.M. Stams, J. van der Oost, R.M. Kelly and S.M.W. Kengen. A thermophile under pressure: Transcriptional analysis of the response of *Caldicellulosiruptor saccharolyticus* to different H_2 partial pressures. International Journal of Hydrogen Energy, 2013 38 (4): 1837-1849.
doi: 10.1016/j.ijhydene.2012.11.082

Glycerol fermentation to hydrogen by *Thermotoga maritima*: proposed pathway and bioenergetic considerations. B.T. Maru#, **A.A.M. Bielen**#, M. Constantí, F. Medina, S.W.M. Kengen. International Journal of Hydrogen Energy, 2013 38(14): 5563-5572.
doi: 10.1016/j.ijhydene.2013.02.130

# These authors contributed equally

The research described in this thesis was performed at the Laboratory of Microbiology, Wageningen University, the Netherlands. The work of A.A.M. Bielen was financially supported by the 2008 Wageningen UR IPOP-program.

Cover design: Zeno van den Broek
Concept cover design: Abraham A.M. Bielen

Thesis layout: Abraham A.M. Bielen

Printed by: GVO drukkers & vormgevers B.V. | Ponsen & Looijen, The Netherlands

The financial support from the Laboratory of Microbiology, Wageningen University for printing this thesis is gratefully acknowledged.