An *in silico* re-design of the metabolism in *Thermotoga maritima* for increased biohydrogen production

Juan Nogales*†, Steinn Gudmundsson*, Ines Thiele*‡

*Center for Systems Biology, University of Iceland, 101 Reykjavik, Iceland, †current address: Department of Bioengineering, University of California at San Diego, La Jolla, CA 92093-0412

‡To whom correspondence should be addressed: Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland. E-mail: ithiele@hi.is.

Abstract

Microbial hydrogen production is currently hampered by lack of efficiency. We examine how hydrogen production in the hyperthermophilic bacterium *Thermotoga maritima* can be increased *in silico*. An updated genome-scale metabolic model of *T. maritima* was used to i) describe in detail the H_2 metabolism in this bacterium, ii) identify suitable carbon sources for enhancing H_2 production, and iii) to design knockout-strains, which increased the *in silico* hydrogen production up to 20%. A novel synthetic oxidative module was further designed, which connects the cellular NADPH and ferredoxin pools by inserting into the model a NADPH-ferredoxin reductase. We then combined this *in silico* knock-in strain with a knock-out strain design, resulting in an *in silico* production strain with a predicted 125% increase in hydrogen yield. The *in silico* strains designs presented here may serve as blueprints for future metabolic engineering efforts of *T. maritima*.

Keywords

Biosustainability; Biohydrogen; *Thermotoga maritima*; genome-scale model; COBRA methods.

1. Introduction

There is a growing concern about the continued use of fossil fuels for energy generation. Depletion of fossil fuels together with high prices, an ever-increasing demand and global warming, demonstrate the need for alternative energy sources. Major research efforts are therefore underway, which aim at harnessing renewable sources, including solar and wind energy, geothermal resources and hydrogen (H_2). Hydrogen is currently seen as a promising energy carrier, which may provide an efficient alternative to fossil fuels for transportation [1]. Moreover, hydrogen is used in large quantities in the petroleum and chemical industries. The uses include hydrocracking, saturating fats and oils, and the production of raw chemicals, such as hydrochloric acid, ammonia, and methanol [2]. Most of the H_2 currently produced is derived from fossil fuels, mainly by the steam reforming of methane or natural gas [3]. Methods, which do not utilize hydrocarbons as a primary source, include electrolysis of water, thermal decomposition and biological methods. Both electrolysis and thermal methods are energy inefficient and may also depend indirectly on fossil fuels for electricity or heat generation [1]. Hydrogen has the potential to replace fossil fuels, provided that it can be generated economically and in an environmentally friendly manner.

Biological methods employing microbes for H_2 production have received significant attention in the last decade. They possess several advantages over traditional methods, such as the ability to use renewable energy sources as feedstock and they do not rely on high temperatures or pressure. The major drawback of microbial H_2 production so far has been lack of efficiency. Metabolic pathways exist in organisms (e.g., the oxidative pentose

phosphate pathway), which can, theoretically, produce stoichiometric amounts of H_2 from glucose. However, pathways exceeding the yield of 4 mol of H_2 per mol of glucose are not known. It is expected that such theoretical high-yielding pathways would compete with the growth yield. Growth de-coupled production approaches have therefore been suggested for increasing the efficiency in bio-hydrogen production [4, 5].

Thermotoga maritima is one of several hyperthermophilic bacteria, which have received considerable interest recently as potential sources of hydrogen [6]. *T. maritima* is able to produce H_2 in good yields from a wide array of inexpensive polysaccharide sources, such as starch, xylan, and cellulose [7]. In addition, it has an optimal growth temperature around 80°C. Biohydrogen production at high temperatures has several advantages over lower temperatures. They include better pathogenic destruction [8] and a lower risk of contamination by methanogenic archaea [9]. High temperatures shift the equilibrium point of the H_2-pathways in the direction of H_2 production by a factor of up to 4.5, which results in higher H_2 yields [10, 11]. Moreover, hyperthermophiles are less sensible to high H_2 partial pressures [12]. All these properties make *T. maritima* an excellent candidate for H_2 production. *T. maritima* possesses a strong coupling between cell growth and H_2 production. It has higher H_2 yields than most other H_2 producing microorganisms and is close to the Thauer limit (4 mol H_2 per mol glucose, with acetate as the main byproduct) [6, 13, 14].

Systems biology is finding increasing use in metabolic engineering and there are already several examples of successful applications of *in silico* analysis to genome-scale metabolic networks [15-19]. Constraint-based modeling approaches have been

demonstrated to predict the effects of gene knock-ins, gene knockouts, and gene down-regulation on the phenotype with reasonable accuracy [20, 21].

The purpose of this study was to investigate how H_2 production in *T. maritima* could be increased using a systematic *in silico* approach to metabolic engineering, i.e., flux balance analysis [22]. The present computational study includes the identification of optimal carbon sources and a detailed systems biology analysis of the H_2 metabolism in *T. maritima*. It also includes an extensive search for gene knock-outs and knock-ins, which result in increased *in silico* H_2 production.

2. Materials and methods

2.1. Metabolic reconstruction

The recently published genome-scale metabolic reconstruction of *T. maritima* iTZ479 [23] was used to study the H_2 production in this hyperthermophilic bacterium. The model consists of 478 genes, 503 metabolites, 562 intracellular and 83 extracellular reactions. Since the publication of the original reconstruction, a better understanding of the hydrogenase activity in *T. maritima* has been obtained [24]. The model was updated to take this information into account as well as to fill gaps in the central metabolism, including the completion of the Entner-Doudoroff pathway [13, 25]. A detailed description of the model updates is given in Table S1. The model was grown on the *in silico* minimal medium previously defined [23]. In order to compare the effects of different carbon sources on the growth rates and H_2 production, the uptake rates were normalized to 10 mmol/gDW/h of glucose in terms of the number of carbon atoms present. The reconstruction was extended based on the established reconstruction protocol [26].

2.2. Flux balance analysis

H_2 production was studied by applying flux balance analysis [27, 28] to the model. The analysis was carried out by solving the following linear optimization problem (LP)

$$\begin{aligned} &\text{max} && \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v} \\ &\text{subject to} && \mathbf{S}\mathbf{v} = \mathbf{0} \\ & && \boldsymbol{v}_{\mathrm{l}} \le \boldsymbol{v} \le \boldsymbol{v}_{\mathrm{u}} \end{aligned} \qquad (1)$$

where S is an m x n matrix containing all the stoichiometric coefficients in the model of m metabolites and n reactions. The vector $\boldsymbol{v}$ has n elements, which represent the individual flux values for each reaction that are to be determined (decision variables). The constraints $\boldsymbol{S}\boldsymbol{v} = \mathbf{0}$ correspond to steady-state mass conservation. The vector $\boldsymbol{c}$ has n elements and contains zeros for all entries but the reaction(s) that are part of the objective function (e.g., growth rate). The vectors $\boldsymbol{v}_{\mathrm{l}}$ and $\boldsymbol{v}_{\mathrm{u}}$ are vectors with n elements each, which represent the lower and upper bounds on the fluxes, respectively. Analysis of reaction essentiality was performed by fixing the flux through the corresponding reaction to zero and testing whether the corresponding model supported growth. The consumption or production of NADH was simulated by including a sink or source reaction for this metabolite, respectively. The flux across this reaction was forced from -25 to 25 mmol/gDW/h. Negative and positive values correspond to NADH production and consumption, respectively.

There are in general multiple equivalent solutions to (1), i.e., there are infinitely many flux distributions, which correspond to the maximum cellular objective [29, 30]. For

the study of reaction essentiality and knockout analysis this is usually not a problem, since only the objective value is considered. However, for the interpretation of individual flux values or comparisons of fluxes through different pathways, the existence of multiple optimal solutions needs to be taken into account. Several strategies have been proposed to address this issue, including the following two step procedure. The model (1) is solved to find the maximum value for the cellular objective, denoted by Z*. In the second step, the following convex quadratic optimization problem is solved

$$\begin{aligned} \text{min} \quad & \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} \\ \text{subject to} \quad & \mathbf{S}\mathbf{v} = \mathbf{0} \\ & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{Z}^{*} \\ & \boldsymbol{v}_{\mathrm{l}} \le \boldsymbol{v} \le \boldsymbol{v}_{\mathrm{u}} \end{aligned} \qquad (2)$$

The resulting flux distribution $\boldsymbol{v}$ is the smallest overall flux, in the sum of squares sense [31, 32], which achieves the maximum cellular objective. Minimizing the overall flux in this way has the additional benefit of removing loops from the network. The strict convexity of the objective function ensures that the solution is unique [33] . In the following, whenever we compare fluxes between reactions or pathways, we refer to flux values obtained by this two-step process.

2.3. Design of knockouts

Although growth de-coupled production strategies could increase the biohydrogen efficiency [4, 5], the coupled synthesis of the target chemical to biomass yield is usually of interest for metabolic engineering. This allows for easy selection of overproducing strains by selecting

the strains with the fastest growth. In addition, these strains will exhibit a stable phenotype since mutations, which decrease the target chemical production rate will result in a lower growth rate and will subsequently be outcompeted [34]. This coupling can be difficult to achieve in general and may require multiple gene-knockouts [34]. Numerous algorithms, such as OptKnock [35], OptGene [36], RobustKnock [37], and GDLS [38], have been proposed for designing production strains via gene knockouts. The search for knockout mutants is a combinatorial optimization problem and most of the published algorithms involve the solution of one or more mixed-integer linear problems (MILP). General MILPs are known to be hard computational problems and the effort required to solve such problems scales exponentially with the problem size. This means that the time taken to solve MILPs arising from network reconstructions quickly becomes prohibitive. The application of MILP to genome scale reconstructions is further complicated by scaling issues arising from the large range of coefficients of the stoichiometric matrix when the reconstruction contains a biomass reaction. The difficulties manifest themselves as invalid solutions returned by the MILP solvers.

Instead of formulating the mutant search as a MILP, we decided to carry out an exhaustive search over all single, double, and triple knockout mutants. While this strategy is limited to few knockouts, four to five at most for genome-scale models, it avoids the inherent difficulty of solving large, ill-conditioned MILPs. A major benefit of this strategy is that it finds *all* growth-coupled mutants instead of a single mutant returned by most of the MILP based algorithms. In addition, because the generation of multiple knock-outs in *T. maritima* is currently not possible due to a severe lack of vector and knockout systems [39],

our approach of restricting the scope to three knockouts is reasonable. For the *T. maritima* model, the exhaustive search can be done in less than 15 minutes on a desktop PC.

The generation of each knockout mutant involved solving two LP problems. The solution to the first problem provided the maximum growth rate of the knockout mutant. The second LP *minimized* the amount of target chemical while maintaining a growth rate equal to the maximum. For growth-coupled mutants, the objective value of the latter LP is always greater than zero. It is straightforward to use other, e.g., nonlinear, criteria, such as the strength of growth coupling [34] or the biomass-coupled product yield, i.e., the product of the growth rate and product yield (BPCY) [36]. Once all growth-coupled mutants were identified, the ones with the largest minimum H_2 yields were selected. Several different knockouts can give rise to identical or almost identical phenotypes due to the presence of non-essential reactions. We therefore filtered the strain designs by considering two mutants as different if their respective growth rates and H_2 production differed in the third significant digit.

To reduce the computational complexity, flux variability analysis [30, 40] was used to identify dead-end metabolites (i.e., metabolites that can be only produced or consumed) and to remove blocked reactions (i.e., reactions, which cannot carry any flux in the given simulation condition) from the network prior to performing the *in silico* knock out screening. The reactions targeted for knockouts consisted of all reactions in the reduced models but excluded essential and exchange reactions.

3. Results and discussion

3.1. Model refinements and new insights into the H_2 production in *T. maritima*

The main objective of this study was to design efficient H_2 producing strains. We therefore updated the published metabolic reconstruction of *T. maritima* (iTZ479) [23] with recent findings in the H_2 metabolism of *T. maritima*. For instance, the hydrogenase of *T. maritima* has been characterized as a bifurcating hydrogenase enzyme, which uses ferredoxin (fdxr) and NADH in an exact stoichiometric ratio of two, but is unable to use either NADH or fdxr as the sole electron donor [24]. Because the formulation of a correct hydrogenase reaction is critical in this work, a new bifurcating hydrogenase reaction (BHYH2) replaced the NADH-dependent and fdxr-dependent hydrogenase reactions present in iTZ479. Moreover, the two-cluster [4Fe:4S]-ferredoxin included in the original model was replaced by the one-cluster [4Fe:4S]-ferredoxin [41]. The stoichiometric coefficients in the ferredoxin participating reactions (i.e., OOR2 and POR), were modified accordingly. An analysis of the original model revealed that important reactions involved in the H_2 and carbohydrate metabolism were missing. They include reactions for L-lactate and L-alanine secretion since the two metabolites are well known fermentation products of *T. maritima* [13, 25]. The Entner-Doudoroff (ED) pathway was completed by adding the 6-phosphogluconate dehydratase (EDD) reaction to the model on the basis of sequence and biochemical evidences [13, 25]. We added a NADH:ferredoxin oxidoreductase (NFO) reaction to the model. While the encoding gene for this reaction is unknown, ferredoxin-dependent reduction of NAD^+ has been detected in *T. maritima* extracts and it has been attributed to redox balancing [13, 25, 42]. We observed differences in the predicted H_2 production and the growth rate between the original model (Fig. 1, dotted lines) and the updated model (Fig 1., solid lines) growing on glucose (10 mmol/gDW/h). For instance, we found unattainable functional states driven by the NADH dependent hydrogenase activity in the

original model (Fig. 1, green patch), while the functional states in the updated model involved L-lactate and L-alanine secretion (Fig. 1, orange patch). The complete list of modifications included in the updated model, referred to as iTZ479_v2 in the following, is given in Table S1.

The effects of the model updates were tested by using all the carbon sources, which sustained growth of iTZ479 [23]. Since the H_2 production in *T. maritima* is growth-coupled [13], we decided to use maximum biomass as the optimization criterion in the linear program (1) instead of maximum H_2 production [23, 43]. The growth rate and H_2 production, scaled per substrate carbon, for the carbon sources was then computed (Figs S1 and S2). The differences in growth rates and H_2 production between the two models were insignificant for most of the carbon sources (within 3%). However, the updated model was not able to use glycerol and rhamnose as carbon sources. Complete oxidation of glycerol to acetate and CO_2 provides an extra mol of NADH, which corresponds to an fdxr/NADH ratio of one (Fig. 2A). In the original model, the extra NADH was oxidized by the NADH-dependent hydrogenase reaction, which resulted in a 54% increase in H_2 production compared to glucose (Fig. S2) [23]. The inclusion of the bifurcating hydrogenase in the updated model resulted in an over-reduced state, avoiding growth. This behavior is consistent with recent experimental reports, which indicate that while glycerol is initially oxidized by *Thermotoga* sps. [44], it does not enable growth [14]. These findings indicate that *T. maritima* is highly sensitive to changes in the internal redox state and that it is lacking the buffering capacity and flexibility exhibited by other, more robust organisms, such as *Escherichia coli*, *Pseudomonas putida,* and *Saccharomyces cerevisiae* [45-47].

In summary, the inclusion of a bifurcating hydrogenase in the metabolic model of *T. maritima* revealed that the fdxr/NADH ratio required by this enzyme may limit the ability to respond to internal perturbations in the redox state of the cell. Furthermore, the updated model also provided more realistic H_2 yields and growth predictions for glycerol.

3.2. A systems biology analysis of the H_2 metabolism in *T. maritima*

The iTZ479_v2 model captures the ability of *T. maritima* of utilizing a variety of simple and complex carbohydrate substrates for growth [48] (Fig. S1). We identified common patterns in their metabolism and as expected, we found that most of the carbohydrates analyzed were converted to glucose-6-phosphate (g6p) or fructose 6-phosphate (f6p) (Fig. 2A), independent of the initial catabolic reactions. *T. maritima* performs a broad range of phosphorylation reactions for funneling carbohydrates to sugar phosphates. They include free phosphate for the cellobiose metabolism, polyphosphate used by the inorganic phosphate-dependent phospho-fructose kinase and ATP for glucose phosphorylation [49]. It is therefore not surprising that there are still uncertainties about how the differing initial ATP requirements affect the carbohydrate metabolism and the bioenergetics in *T. maritima*. We used iIZ479_v2 in order to study the impact of the initial catabolism of carbohydrates in the metabolism of *T. maritima* (shown in purple in Figure 2). Our analysis predicted significant differences in the growth rate between different carbon sources, but negligible differences in H_2 yields (Fig. S1). Polysaccharides were found to be the most suitable carbon sources, followed by tri- and disaccharides. Lower growth rates were obtained with monosaccharides (Fig. S1). These results are consistent with previous experimental findings, which report that growth on glucose is slower than for

other polysaccharides [48]. In summary, our results point to complex carbohydrates being preferred carbon sources for achieving a higher biomass-coupled product yield (Fig. 2B, S1).

Sugar phosphates can be further metabolized by three different pathways, the Embden-Meyerhof (EM) pathway, the Entner-Doudoroff (ED) pathway, or through the oxidative branch of the pentose phosphate pathway (OPP) (Fig. 2A, solid, dashed and dotted blue lines, respectively). In order to understand the relative importance of these pathways, we carried out an analysis of network robustness by computing the growth rate as a function of the flux across representative reactions of the EM, ED, and OPP pathways (Fig. 2C). The results show that an increase in flux across either the EM or ED pathways correlated with an increase in growth (up to a point), a flux increase in the OPP always had a negative effect on growth. Similar results were obtained for the H_2 production (Fig. 2D). These results suggest that the EM and the ED pathways should be preferred to OPP for sugar phosphate metabolism and for H_2 production, which is strongly growth coupled. ^{13}C-labeling experiments have determined that the glucose metabolism in *T. maritima* proceeds through both the EM (87%) and ED (13%) pathways, although EM is more efficient [50]. Consistently, we found that iTZ479_v2 metabolized the carbohydrates mainly through the EM pathway with 12 – 30% proceeding via the ED pathway, the exact amount depending on the type of carbohydrate. For glucose, 86% of it was metabolized by the EM pathway while the remaining 14% was funneled to the ED pathway (Fig. 5A). Although the EM pathway has been identified as the main glycolytic pathway, our analysis suggests that the small fraction metabolized via the ED pathway is essential for growth. When a reaction essentiality analysis was carried out, glucose 6-phosphate dehydrogenase (G6PDH2), 6-

phosphogluconolactonase (PGL), 6-phosphogluconate dehydratase (EDD), and 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA_R) were all required for growth (Fig. 2A, Table S2). The critical role of the ED pathway appears to be related to the maintenance of an optimal fdxr/NADH ratio for the bifurcating hydrogenase and other cell processes (see below).

In contrast to the initial steps of the carbohydrate metabolism, where multiple metabolic pathways are involved in the catabolism of carbohydrates to glyceraldehyde-3-phosphate (g3p), our analysis predicted that g3p was metabolized exclusively through the energy pay-off phase of glycolysis (Fig. 2A, green lines). Not only was the growth strongly dependent on flux across the payoff phase (Fig. 2E) but a reaction essentiality analysis also showed that all reactions in this pathway were essential for growth (Table S2). On the other hand, we found that while a strong coupling between H_2 production and growth rate was observed, the pay-off phase was not essential for H_2 production. Approximately 50% of the maximal H_2 production rate could be achieved without the participation of this pathway by using pyruvate from the ED pathway, with the NFO reaction being the main source of NADH from fdxr (Fig. 2E).

Optimal growth rate and H_2 production in *T. maritima* is achieved with acetic acid as the main fermentative end product [13] (Fig. 2A, brown line). The formation of L-lactate and other byproducts, such as L-alanine, have negative effects on both the growth rate and H_2 production [13, 25, 50]. This is in agreement with our finding that both the H_2 production and the growth rate were strongly dependent on flux across the pyruvate synthase reaction (POR) (Fig. 2H). POR is the sole source of fdxr in the model and we found

that it was essential for H_2 production, but not for growth. In the absence of POR, pyruvate was funneled to L-lactate, resulting in reduced growth rate, being 35% of the optimal growth rate with acetate as fermentation byproduct, but cell viability (Figure 2H). This result strongly suggests that, excluding the bifurcating hydrogenase, POR is a key reaction for H_2 production in *T. maritima*. In line with these latter results, secretion of either L-alanine or L-lactate (Fig. 2A, solid and dotted orange lines, respectively) led to decreased growth rate and H_2 production as expected (Fig. 2F and 2G). Finally, we studied the effects of elemental sulfur reduction in the model (Fig. 2A, pink lines). Our analysis indicated that for the model in order to grow, a small amount of sulfur had to be reduced, presumably because this pathway was involved in the oxidation of NADPH from the ED pathway (Fig. 2A). In addition, we observed a linear decrease of H_2 with increased secretion of H_2S (Fig. 2M). These results are in agreement with experimental reports of sulfur being one of the main electron sinks in *T. maritima* as well as with the growth-promoting effect of elemental sulfur [13].

In summary, our data strongly suggests that the overall H_2 metabolism in *T. maritima* can be decomposed into independent metabolic modules, each having a different impact on both the growth rate and H_2 production (Figs. 2 and 6A). These module include i) the initial carbohydrate metabolism, ii) the sugar phosphate metabolism through the EM, ED, and OPP pathways, iii) the glyceraldehyde-3-phosphate (3gp) metabolism through the pay-off phase of glycolysis, iv) the pyruvate synthase dependent H_2 production module, and v) alternative metabolite secretion modules.

3.3. Internal redox balancing in *T. maritima*

The systems analysis of the H_2 metabolism in *T. maritima* indicated that a complex redox balancing system was needed for providing the exact fdxr/NADH ratio for the bifurcating hydrogenase. In order to shed light on this process we varied the frdx/NADH ratio *in silico* by simulating internal metabolic processes involved in the consumption and production of NADH. We studied the effects of these metabolic states on the metabolism of *T. maritima* and we found three main scenarios when glucose was used as the carbon source (Fig. 3). The first scenario corresponds to a balanced metabolic state, in which the metabolic processes do not modify the NADH/NAD^+ pool provided by glycolysis and in this case the frdx/NADH ratio is exactly two. Under these conditions, no mechanism for redox balancing was required. Glucose was fermented to acetate via EM, while the bifurcating hydrogenase regenerated NAD^+ and oxidized ferredoxin, and maximal growth was achieved (Fig. 3). In a second scenario, identified by the consumption of NADH, the fdxr/NADH ratio was above two and consequently, the bifurcating hydrogenase was not sufficient for redox balancing. The NFO reaction provided NADH from fdxr and was the key reaction in redox balancing under these conditions (Fig. 3). Finally, an increase in the NADH pool resulted in an fdxr/NADH ratio below two, which required the presence of pathways consuming the surplus of NADH. We found under these conditions that NADH was consumed by the funneling of a small fraction of the glucose to the ED pathway, coupled with elemental sulfur reduction. In fact, we found that one mole of NADH was consumed per mole of glucose funneled to the ED pathway. Therefore, the fine tuning between the EM and ED pathways, which yielded fdxr/NADH ratios of two and four, respectively, led to the balancing of the redox state (Fig. 3). This latter scenario was found when iIZ479_v2 grew on glucose (Fig. 3, dotted line) as well as on other carbohydrates.

This result strongly suggests that the *T. maritima* metabolism of sugars as main carbon sources yields in a surplus of NADH and subsequently, in an fdxr/NADH ratio lower than two. By quantifying the NADH consumed required to achieve an fdxr/NADH ratio of two (Fig. 3, black arrow), the surplus of NADH on glucose as carbon source (10 mmol/gDW/h) was estimated to be 1.12 mmol/gDW/h, which correlates well with the flux across the ED pathway under these conditions (Fig. 3). Therefore, the amount of glucose-6-phosphate metabolized by the ED pathway for each carbohydrate could be dependent of the surplus NADH produced. The required redox balancing under these conditions could therefore explain the significant flux through ED found *in vivo* [50], despite the low efficiency of this pathway compared to the EM pathway. In summary, we found that the NFO reaction, the EM, ED, and sulfur reduction pathways are required for internal redox balancing in *T. maritima.*

3.4. Knockout approach toward the *in silico* design of H_2 overproducer strains

In silico models can be used as a platform for examining a wide variety of possible genetic modifications at the genome scale. We therefore performed a thorough *in silico* knock-out analysis by searching over all single, double, and triple reaction knockouts in order to increase H_2 production. Based on the predicted growth rates and H_2 yields (Fig. S1), nine carbon sources were selected prior to carrying out the knockout analysis, consisting of seven glycolytic and two gluconeogenic carbon sources, L-lactate and L-alanine. The glycolytic carbon sources included were glucose (a hexose), xylose (a pentose), sucrose (a disaccharide), raffin (a trisaccharide), cellulose, starch (homopolysaccharides), and xylan (a heteropolysaccharide). The uptake rate for each substrate was carbon

adjusted to correspond to 10 mmol glucose/gDW/h. Analysis of the mutants revealed that the knockout of the H_2S transporter between the cytosol and extracellular compartments (H2St) resulted in several mutants with high H_2 yields. Blocking the transporter prevented the secretion of hydrogen sulfide. This fact limited the use of sulfur as an electron sink, which subsequently led to an increase in H_2 production. The knockout of the H_2S transporter could be emulated by limiting the amount of sulfur present in the medium. We simulated sulfur-limited conditions by adding to the medium, the sulfur amount necessary for supporting maximum growth. Sulfur-limiting conditions were used in the following, unless stated otherwise.

The maximum and minimum flux rates of H_2 in the wild-type strain were computed as a function of the growth rate (Fig. 4, blue curves) as well as for single, double, and triple knockouts (Fig. 4, magenta, green, and red circles), for all the carbon sources under consideration**.** While some mutants resulted in promising production strains**,** numerous inferior mutants were identified (Fig. 4, filled circles), both in terms of growth and H_2 production. While flux balance analysis predicts high H_2 yields in *T. maritima* growing on L-lactate or L-alanine, the results indicate that the yield could not be improved using knockouts alone (Figs. 4D and 4I). For the knockout mutants growing on glycolytic carbon sources, a significant increase in H_2 production was obtained, typically ranging from 13% for a single knockout to 21% for triple knockouts (Figs. 4A, 4C, 4E, 4F, and 4H). In all computed knockouts, the additional production of H_2 was due to the employment of less efficient pathways, which required a higher amount of glucose for ATP production, while the production of reducing equivalents was not affected. As a consequence, a minimal increase in H_2 production was predicted at the expense of reduced growth rates (Fig. 4).

Optimal growth and H_2 production in *Thermotogales* requires the complete oxidation of pyruvate to acetate, providing an extra mole ATP per mole of pyruvate through the reaction catalyzed by acetate thiokinase (ACKr). L-lactate formation is therefore the main limitation (Fig. 2) [13, 25, 50]. Our results show that by modifying the pyruvate metabolism, as well as by blocking L-lactate formation, a significant increase in H_2 production could be obtained for all tested glycolytic carbon sources (Table S3). This was achieved by the combined deletion of the acetate thiokinase (ACKr) and L-lactate dehydrogenase (LDH_L) reactions (Fig. 4, gray circles). This finding is in line with reports of engineered LDH_L-deficient *E. coli* and *Thermoanaerobacterium sp*. strains, which exhibit about two-fold increase in H_2 production [51, 52]. The deletion of the *ack* gene encoding for ACKr in *Clostridium tyrobutyricum* increased the H_2 production from glucose by 50% [53]. It is interesting that neither the single knockout of ACKr nor of LDH_L resulted in an increase in H_2 yield *in silico*, which is in contrast to experimental evidence from other species. This apparent discrepancy does not invalidate our results, but rather supports them. For instance, although the single ACKr knockout of *C. tyrobutyricum* resulted in higher H2 production [53], a continuous culture of this ACKr knockout strain could result in an adaptive evolution including an up-regulation of the LDH_L, increasing the L-lactate excretion and the subsequent decrease in H_2 production (Fig. 2A). Conversely, the growth-coupled overproducer strain consisting of a double deletion identified in our study overcomes this drawback. Furthermore, it could be adaptively evolved in a continuous culture, thus leading to faster growth rates and potentially even higher H_2 production rates [34].

Another high yielding strategy was to block the sugar phosphate metabolism in the initial steps of glycolysis, i.e., either the fructose bisphosphate aldolase (FBA) or the triose-phosphate isomerase (TPI) reactions. A FBA knockout mutant (Fig. 4, orange circles) growing on glucose resulted in a 13% increase in H_2 production (Tables 1, S3). This was achieved by redirecting all carbon flux through the glucose-6-phosphate dehydrogenase (G6PD) to the ED pathway (Fig. 5A). The metabolism of glucose via ED was less efficient since only three moles of ATP per mole of glucose were produced compared to four moles provided via EM. Moreover, two of the three moles of ATP were produced by the ACKr, which resulted in an increase in the carbon flux through the POR reaction in order to maximize the ATP production. Subsequently, the production of H_2 increased at the expense of reduced growth rate (Table 1). The TPI mutant prevented the direct conversion of dihydroxyacetone phosphate (dhap) to glyceraldehyde-3-phosphate and in this mutant, dhap was converted to D-fructose 1,6-bisphosphate via gluconeogenesis and the consumption of one extra mole of ATP. The TPI and FBA knockouts behaved similarly, the small increase in the H_2 production was a consequence of the increase in the carbon flux through the POR reaction in order to maximize the ATP production through ACKr (Fig. 5C). No significant further increase in H_2 yields was predicted when the FBA and TPI knockouts were combined with one or two additional knockouts (Table S3). The combination of ACKr/LDH_L knockouts with either FBA or TPI knockouts turned out to be lethal.

Stoichiometric models have been extensively used for the study of H_2 metabolism in oxygenic-photosynthetic [54-56], photoheterotrophic [57], and heterotrophic microorganisms [43, 58]. However, there are not many publications dealing in depth with the *in silico* redesign of the metabolism toward over-production of H_2. In an interesting study, Pharkya and colleges

addressed the search for growth-coupled H_2 overproducer knockouts in *E. coli*, a well-established production system, and in *Clostridium acetobutylicum*, a known hydrogen producer [59]. They found that while a triple knockout (ACK, ATPsynthase, and 2-oxoglutarate dehydrogenase) in *E. coli* yielded 2.95 mol H_2/mol glucose, a double knockout (ACK and butyrate transport) in *C. acetobutylicum* yielded up to 3.17 mol H_2/mol glucose [59]. Interestingly, this study also indicated that the blocking of ACK and the use of a less efficient metabolism was required to increase the H_2 production. However the results of that study also suggested that a more complex redesign would be required to obtain yields above 4 mol H_2/mol glucose.

In summary, we propose two possible metabolic engineering strategies for increasing the H_2 production in *T. maritima*: i) by modifying the pyruvate metabolism (Fig. 6B) or ii) by blocking the EM pathway (Fig. 6C). In addition, our results indicate that only a modest increase in H_2 yields can be achieved by knockouts alone and that more complex modifications are likely to be needed in order to obtain a significant increase in yields.

3.5. *In silico* design of synthetic oxidative-modules for increased H_2 production

Metabolic engineering strategies, which modify pyruvate metabolism in bacteria in order to increase the H_2 yields, have been extensively employed [51-53]. We therefore focused the subsequent analysis on blocking of the glycolytic pathway and the utilization of alternative pathways for the sugar phosphate metabolism. In particular, it seemed worthwhile to study the funneling of carbons through the oxidative branch of the pentose phosphate pathway. While no flux across the OPP was found in experimental flux distribution experiments [50], theoretical estimates indicate that metabolizing glucose via

the OPP could increase the H_2 yields up to eight mole H_2 per mole glucose [4, 60]. We found that both the growth rate and the H_2 production decreased when carbon flux was redirected through the OPP (Fig. 2C-D). This finding is in agreement with experiments performed with a glucose-6-phosphate isomerase (PGI) knockout *E. coli* strain, which metabolized glucose through the OPP and exhibited a reduced growth rate [61]. This phenotype was explained by the lack of an efficient $NADP^+$ regeneration system [61]. In order to test if the reduced growth rate driven by the OPP was due to the lack of a $NADP^+$ regeneration system, we decided to include a non-native NADPH-dependent hydrogenase reaction into the model[62].

When this reaction was added to the model, a significant increase in the flux across the OPP and a high H_2 production rate was predicted when H_2 production was used as the optimization criterion (data not shown). This finding is in line with the aforementioned *E. coli*, study [61] and suggests that a lack of reactions consuming NADPH hampers the flow through the OPP. We therefore hypothesize that the inclusion of a non-native reaction connecting the NADPH and NADH/ferredoxin pools could increase the flux across the OPP and thus, could increase the H_2 yields. Since the over-expression of non-native hydrogenases is still a considerable challenge [43, 63, 64], even more so in *T. maritima* for which no appropriate genetic tools exist, we decided to analyze a simpler strategy. We simulated the knock-in of the thermostable NADPH-NADH transhydrogenase (NADP_H2) and NADPH-Ferredoxin reductase (FNOR_HT) from *Thermus thermophilus* (Q5SLT5) and *Hydrogenobacter thermophilus* TK-6 [65], respectively. The inclusion of either NADP_H2 or FNOR_HT reactions did not result in significant changes compared to the wild type. When optimizing for growth rate, we found that the strains grew equally fast as the wild type,

while exhibiting only a marginal increase in H_2 production (Fig. 2L, Table 1). However, when H_2 production was selected as an optimization criterion, the inclusion of either NADP_H or FNOR_HT yielded a two-fold increase in H_2 production (Fig. 2K). The additional H_2 was obtained by an increased flux across the OPP pathway, as expected, (Fig. 2K) and in a POR-independent manner (Fig. 2I). Pyridine nucleotide transhydrogenases have been employed to successfully improve strain productivity by enabling NADH/NADPH interconversion [66, 67]. However, since the directionality of the transhydrogenase reaction largely depend on the NADH/NADPH ratio and due to the presence of additional NADH or NADPH consuming reactions, the over-expression of transhydrogenases may result in unexpected phenotypes [68]. Thus, it seems reasonable to assume that the presence of an effective $NADP^+$ regeneration system is necessary for increasing the flux across the OPP but that it may not be sufficient.

In an attempt to design additional H_2 over-producer strains based on the OPP, while avoiding undesirable phenotypes, we repeated the growth-coupled knockout analysis using the two knock-in strains. The increase in H_2 production of single knockout mutants in the NADP_H2 knock-in strain was negligible, and we identified again the FBA and TPI mutants as the mutants with highest yields (Table 1). When double knockouts strains were analyzed *in silico*, the computations resulted in new knockout strains, in which the glucose metabolism was completely redesigned. These double knockouts resulted in substantial increases in H_2 yields while maintaining acceptable growth rates (Tables 1, S4). For instance, blocking the direct conversion of glucose into f6p, achieved through the deletion of the glucose-6-phosphate isomerase (PGI) and the knockout of the fructose kinase (HEX7) or the D-glucose aldose-ketose-isomerase (XYLI2), led to an increase in H_2

production by more than 40% (Tables 1, S4). The high H_2 yields were achieved by metabolizing glucose mainly through the OPP (Fig. 5A) and in a POR-independent manner (Fig. 5D).

We then repeated the knockout analysis in the context of a FNOR_HT knock-in, and obtained similar results as for single knockouts (Tables 1, S5). The double knockout strains involving PGI and HEX7/XYLI2 showed similar H_2 yields to those obtained with the NADPH_H2 knock-in strain (Fig. 6D, Tables 1, S5). However, in the presence of FNOR_HT, the ED pathway became non-essential (Fig. 2L), since the optimal oxidation-reduction state could be provided by reduced ferredoxin derived from the NADPH. Subsequently, we identified a new double knockout mutant, deficient in FBA and 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA_R), with blocked EM and ED pathways. In this mutant, the combined action of FNOR_HT and the knockouts redirected the glucose flux exclusively through the OPP (Fig. 5A). The combined action of FNOR_HT and NFO provided the required fdxr/NADH ratio for the bifurcating hydrogenase and as a consequence, this double mutant resulted in a 125% increase in H_2 production in a POR-independent manner (Fig. 5E, Table 1). The predicted H_2, CO_2, and acetate yields per mole glucose were 7.6, 3.8, and 0.8, respectively, which superseded the Thauer limit for H_2 production with acetate as byproduct. In fact, these values are close to the theoretical maximum yield obtained by funneling sugar phosphates through the OPP (Fig. 5B, 6E) [60] and they represent almost two thirds of the maximum stoichiometric yield of hydrogen from glucose (i.e., 12 mole H_2) [62].

While we are not aware of any *in vivo* attempts employing this strategy for overproducing H_2 in other bacteria, there is experimental evidence suggesting that this approach could be implemented. For instance, *E. coli*'s *pgi* knock-out mutant has been described to metabolize glucose mainly through the OPP, which introduced a redox imbalance due to excess NADPH leading to reduced growth compared to the wild-type [61]. It is interesting to note that over-expressing the soluble UdhA transhydrogenase improved the growth rate of this mutant significantly [61]. Furthermore, recurrent mutations enhancing the activity of the transhydrogenase were observed in strains isolated from adaptive evolution experiments performed on the *E. coli pgi* knock-out strains as well as decreased acetate excretion [69]. These experiments highlight the synergism (blocking of glycolysis and inclusion of an efficient NADPH regeneration system) required to increase the flux across the OPP and support the reduced acetate secretion rate computed in those OPP-based H_2-overproducer strains (Fig. 5A, 5B). Finally, is noteworthy that similar behavior has been described for yeast *pgi* knock-outs over-expressing the UdhA transhydrogenase [70], which could indicate that this strategy can be indeed extended to other organisms.

3.6. Expanding the metabolic capabilities of *T. maritima in silico*

Glycerol is the main by-product from the hydrolysis of fats and is currently generated in large amounts in the biodiesel industry. As a result glycerol has gone from being a commodity chemical to a waste chemical in less than a decade. This is one of the reasons why glycerol has been identified as a promising carbon source for industrial microbiology in the future [71]. The complete pathway for glycerol metabolism is encoded

in the *T. maritima* genome and a previous *in silico* analysis suggested that high yields of H_2 could be obtained by using glycerol as a carbon source [23]. Unfortunately, a recent experiment demonstrated that *T. maritima* is unable use glycerol as a carbon source [14], which could be due to the over-reduced state described previously. An unexpected consequence of the FNOR_HT inclusion was that the model was able to grow on glycerol. Analysis of flux values indicated that a significant fraction of glycerol was funneled to the OPP via gluconeogenesis. As a result, the extra ferredoxin levels derived from NADPH to balance the excess of NADH produced during the first steps of the glycerol metabolism were able to support growth. In addition, we obtained very high H_2 yields, which were twice as high as those found for the other carbohydrates (Table 1). We then repeated the knock out analysis using glycerol as carbon source, combined with the FNOR_HT knock-in. While, no single mutants outperformed the wild type strain, we found that the H_2 production in the double mutants ACKr/LDH_L increased slightly (Table 1). Since glycerol was metabolized through the payoff phase via dhap (Fig. 2, grey line), the double mutant FBA/EDA_R identified previously did not increase H_2 production, however, a triple mutant, which additionally blocked the phosphate isomerase reaction (TPI) did. Glycerol was metabolized in this mutant to g3p through OPP via gluconeogenesis, similar to the metabolism of glucose in the double FBA/EDA_R knockout. In fact, this strategy gave H_2 yields close to the maximum theoretical yields of five moles H_2 per mole glycerol (Table 1). The results suggest that glycerol could be exploited as an inexpensive carbon source by means of a single gene insertion (FNOR_HT) into *T. maritima*, producing H_2 in high yields.

4. Conclusions

We have presented an updated genome-scale metabolic model of the hyperthermophile *T. maritima* and employed it to gain insights into the H_2-metabolism of this bacterium. The model was then used to re-design the H_2 metabolism *in silico* to improve the H_2 yields. Systems analysis revealed that several individual modules were involved in the H_2 metabolism, with POR being a key reaction. We showed that the H_2 production could be increased by up to 21% by using three knockouts for *T. maritima* growing on cellulose, sucrose, starch and xylan. In addition, the combination of gene knock-outs with a single inclusion of either NADP_H2 or FNOR_HT completely re-designs the native H_2 metabolism. A new synthetic H_2 producing module, driven by the OPP, could increase the H_2 production by 125% when combined with knockouts of the FBA and EDA_R reactions. This strain design is non-intuitive and requires a comprehensive *in silico* systems metabolic engineering approach. In addition, the sole inclusion of FNOR_HT allowed the model to grow on glycerol, a cheap and abundant carbon source while producing H_2 in with high yields. It seems reasonable to hope that genetic engineering strategies for *T. maritima* will become available in the near future [72] and the implementation of some of the proposed strains here could be of great interest.

Acknowledgements

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant DE-SC0002009. JN was funded, in part, by the Spanish MEC through National Plan of I-D+i 2008-2011. The authors thank Dr. Ronan M.T. Fleming for valuable discussions.

References

[1] Veziroglu A, Macario R. Fuel cell vehicles: State of the art with economic and environmental concerns. Int J Hydrogen Energ. 2011;36:25-43.
[2] Baade WF, Parekh UN, Raman VS. Hydrogen. New York: John Wiley & Son; 2001.
[3] Rao KK, Cammack R. Producing hydrogen as a fuel. London and New York: Taylor & Francis; 2001.
[4] Hallenbeck PC, Benemann JR. Biological hydrogen production; fundamentals and limiting processes. Int J Hydrogen Energ. 2002;27:1185-93.
[5] Keasling JD, Benemann J, Pramanik J, Carrier TA, Jones KL, Dien SJ. A Toolkit for Metabolic Engineering of Bacteria BioHydrogen. In: Zaborsky OR, Benemann JR, Matsunaga T, Miyake J, San Pietro A, editors.: Springer US; 1999. p. 87-97.
[6] Chou C-J, Jenney Jr FE, Adams MWW, Kelly RM. Hydrogenesis in hyperthermophilic microorganisms: Implications for biofuels. Metab Eng. 2008;10:394-404.
[7] Chhabra SR, Shockley KR, Conners SB, Scott KL, Wolfinger RD, Kelly RM. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium Thermotoga maritima. The Journal of biological chemistry. 2003;278:7540-52.
[8] Sahlstrom L. A review of survival of pathogenic bacteria in organic waste used in biogas plants. Bioresource technology. 2003;87:161-6.
[9] van Groenestijn JW, Hazewinkel JHO, Nienoord M, Bussmann PJT. Energy aspects of biological hydrogen production in high rate bioreactors operated in the thermophilic temperature range. Int J Hydrogen Energ. 2002;27:1141-7.
[10] Kongjan P, Min B, Angelidaki I. Biohydrogen production from xylose at extreme thermophilic temperatures (70 degrees C) by mixed culture fermentation. Water research. 2009;43:1414-24.
[11] Veit A, Akhtar MK, Mizutani T, Jones PR. Constructing and testing the thermodynamic limits of synthetic NAD(P)H:H2 pathways. Microbial biotechnology. 2008;1:382-94.
[12] van Niel EWJ, Claassen PAM, Stams AJM. Substrate and product inhibition of hydrogen production by the extreme thermophile, Caldicellulosiruptor saccharolyticus. Biotechnol Bioeng. 2003;81:255-62.
[13] Schröder C, Selig M, Schönheit P. Glucose fermentation to acetate, CO_2; and H_2; in the anaerobic hyperthermophilic eubacterium ;*Thermotoga maritima*: involvement of the Embden-Meyerhof pathway. Archives of Microbiology. 1994;161:460-70.
[14] Eriksen N, Riis M, Holm N, Iversen N. H2 synthesis from pentoses and biomass in *Thermotoga spp*. Biotechnol Lett. 2011;33:293-300.
[15] Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, et al. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng. 2005;91:643-8.
[16] Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. Applied and environmental microbiology. 2005;71:7880-7.
[17] Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol. 2007;3:149.
[18] Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci U S A. 2007;104:7797-802.

[19] Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nature chemical biology. 2011;7:445-52.
[20] Oberhardt MA, Palsson BO, Papin JA. Applications of genome-scale metabolic reconstructions. Mol Syst Biol. 2009;5:320.
[21] Feist AM, Palsson BO. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotech. 2008;26:659-67.
[22] Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotechnol. 2010;28:245-8.
[23] Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, et al. Three-Dimensional Structural View of the Central Metabolic Network of Thermotoga maritima. Science. 2009;325:1544-9.
[24] Schut GJ, Adams MWW. The Iron-Hydrogenase of *Thermotoga maritima* Utilizes Ferredoxin and NADH Synergistically: a New Perspective on Anaerobic Hydrogen Production. J Bacteriol. 2009;191:4451-7.
[25] Huber R, Langworthy TA, König H, Thomm M, Woese CR, Sleytr UB, et al. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Archives of Microbiology. 1986;144:324-33.
[26] Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5:93-121.
[27] Fell DA, Small JA. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Journal of Biochemistry. 1986;238:781 - 6.
[28] Savinell JM, Palsson BO. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. Journal of Theoretical Biology. 1992;154:421-54.
[29] Thiele I, Fleming RM, Bordbar A, Schellenberger J, Palsson BO. Functional characterization of alternate optimal solutions of Escherichia coli's transcriptional and translational machinery. Biophys J. 2010;98:2072-81.
[30] Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic engineering. 2003;5:264-76.
[31] Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol. 2007;3.
[32] Grafahrend-Belau E, Schreiber F, Koschutzki D, Junker BH. Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. Plant physiology. 2009;149:585-98.
[33] Fletcher R. Practical methods of optimization. 2 ed: John Wiley and Sons; 1987.
[34] Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metabolic engineering. 2010;12:173-86.
[35] Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng. 2003;84:647-57.
[36] Patil KR, Rocha I, Forster J, Nielsen J. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics. 2005;6:308.
[37] Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics. 2009;26:536-43.
[38] Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, et al. Large-scale identification of genetic design strategies using local search. Mol Syst Biol. 2009;5:296.
[39] Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, Kelly RM. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic Thermotoga species. FEMS Microbiol Rev. 2006;30:872-905.
[40] Gudmundsson S, Thiele I. Computationally Efficient Flux Variability Analysis. BMC Bioinformatics. 2010;11.

[41] Blamey JM, Mukund S, Adams MWW. Properties of a thermostable 4Fe-ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. FEMS Microbiology Letters. 1994;121:165-9.
[42] Verhaart MRA, Bielen AAM, Oost Jvd, Stams AJM, Kengen SWM. Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. Environmental Technology. 2010;31:993-1003.
[43] Oh Y-K, Kim H-J, Park S, Kim M-S, Ryu DDY. Metabolic-flux analysis of hydrogen production pathway in *Citrobacter amalonaticus* Y19. Int J Hydrogen Energ. 2008;33:1471-82.
[44] van der Lelie D, Van Ooteghem SA, Jones A, Dong B, Mahajan D. H-2 production and carbon utilization by Thermotoga neapolitana under anaerobic and microaerobic growth conditions. Biotechnol Lett. 2004;26:1223-32.
[45] Ebert BE, Kurth F, Grund M, Blank LM, Schmid A. Response of *Pseudomonas putida* KT2440 to Increased NADH and ATP Demand. Applied and environmental microbiology. 2011;77:6597-605.
[46] Holm AK, Blank LM, Oldiges M, Schmid A, Solem C, Jensen PR, et al. Metabolic and Transcriptional Response to Cofactor Perturbations in *Escherichia coli*. Journal of Biological Chemistry. 2010;285:17498-506.
[47] Hou J, Lages NF, Oldiges M, Vemuri GN. Metabolic impact of redox cofactor perturbations in *Saccharomyces cerevisiae*. Metab Eng. 2009;11:253-61.
[48] Conners SB, Montero CI, Comfort DA, Shockley KR, Johnson MR, Chhabra SR, 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.
[49] Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, Kelly RM. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic Thermotoga species. FEMS Microbiology Reviews. 2006;30:872-905.
[50] Selig M, Xavier KB, Santos H, Schönheit P. Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. Archives of Microbiology. 1997;167:217-32.
[51] Yoshida A, Nishimura T, Kawaguchi H, Inui M, Yukawa H. Enhanced hydrogen production from glucose using *ldh* and *frd* inactivated *Escherichia coli* strains. Applied Microbiology and Biotechnology. 2006;73:67-72.
[52] Li S, Lai C, Cai Y, Yang X, Yang S, Zhu M, et al. High efficiency hydrogen production from glucose/xylose by the *ldh*-deleted *Thermoanaerobacterium* strain. Bioresource technology. 2010;101:8718-24.
[53] Liu X, Zhu Y, Yang S-T. Construction and Characterization of *ack* Deleted Mutant of *Clostridium tyrobutyricum* for Enhanced Butyric Acid and Hydrogen Production. Biotechnology Progress. 2006;22:1265-75.
[54] Montagud A, Navarro E, Fernandez de Cordoba P, Urchueguia J, Patil K. Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. BMC Systems Biology. 2010;4:156.
[55] Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proceedings of the National Academy of Sciences. 2012.
[56] Gomes de Oliveira Dal'Molin C, Quek L-E, Palfreyman R, Nielsen L. AlgaGEM - a genome-scale metabolic reconstruction of algae based on the Chlamydomonas reinhardtii genome. BMC Genomics. 2011;12:S5.
[57] Hadicke O, Grammel H, Klamt S. Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. BMC Systems Biology. 2011;5:150.
[58] Show KY, Lee DJ, Tay JH, Lin CY, Chang JS. Biohydrogen production: Current perspectives and the way forward. International Journal of Hydrogen Energy. 2012.
[59] Pharkya P, Burgard AP, Maranas CD. OptStrain: A computational framework for redesign of microbial production systems. Genome Research. 2004;14:2367-76.

[60] de Vrije T, Mars A, Budde M, Lai M, Dijkema C, de Waard P, et al. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Applied Microbiology and Biotechnology. 2007;74:1358-67.
[61] Canonaco F, Hess TA, Heri S, Wang T, Szyperski T, Sauer U. Metabolic flux response to phosphoglucose isomerase knock-out in Escherichia coli and impact of overexpression of the soluble transhydrogenase UdhA. FEMS Microbiology Letters. 2001;204:247-52.
[62] Woodward J, Orr M, Cordray K, Greenbaum E. Biotechnology: Enzymatic production of biohydrogen. Nature. 2000;405:1014-5.
[63] Sun J, Hopkins RC, Jenney FE, Jr., McTernan PM, Adams MWW. Heterologous Expression and Maturation of an NADP-Dependent [NiFe]-Hydrogenase: A Key Enzyme in Biofuel Production. PloS one. 2010;5:e10526.
[64] Abo-Hashesh M, Wang R, Hallenbeck PC. Metabolic engineering in dark fermentative hydrogen production; theory and practice. Bioresource technology. 2011;102:8414-22.
[65] Ikeda T, Nakamura M, Arai H, Ishii M, Igarashi Y. Ferredoxin-NADP+ reductase from the thermophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus* TK-6. FEMS Microbiology Letters. 2009;297:124-30.
[66] Wong MS, Causey TB, Mantzaris N, Bennett GN, San K-Y. Engineering poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer composition in *E. coli*. Biotechnol Bioeng. 2008;99:919-28.
[67] Kabus A, Georgi T, Wendisch V, Bott M. Expression of the *Escherichia coli pntAB* genes encoding a membrane-bound transhydrogenase in *Corynebacterium glutamicum* improves L-lysine formation. Applied Microbiology and Biotechnology. 2007;75:47-53.
[68] Nissen TL, Anderlund M, Nielsen J, Villadsen J, Kielland-Brandt MC. Expression of a cytoplasmic transhydrogenase in *Saccharomyces cerevisiae* results in formation of 2-oxoglutarate due to depletion of the NADPH pool. Yeast. 2001;18:19-32.
[69] Charusanti P, Conrad TM, Knight EM, Venkataraman K, Fong NL, Xie B, et al. Genetic Basis of Growth Adaptation of *Escherichia coli* after Deletion of *pgi*, a Major Metabolic Gene. PLoS Genet. 2010;6:e1001186.
[70] Heux S, Cadiere A, Dequin S. Glucose utilization of strains lacking PGI1 and expressing a transhydrogenase suggests differences in the pentose phosphate capacity among Saccharomyces cerevisiae strains. FEMS Yeast Research. 2008;8:217-24.
[71] Contiero J, da Silva GP, Mack M. Glycerol: A promising and abundant carbon source for industrial microbiology. Biotechnol Adv. 2009;27:30-9.
[72] Frock A, Petrus A, White D, Singh R, Loder A, Blum P, et al. Biohydrogenesis in the Thermotogales. Abstract presented at the DOE Genomic Science Awardee meeting IX2011.

Figure legends

Figure 1. Model refinements. The minimum and maximum H_2 production is shown as a function of the growth rate when 10 mmol/gDW/h of glucose was used as carbon source.

Figure 2. Systems analysis of the H_2 metabolism in *T. maritima.* The main reactions involved in the H_2 metabolism and those identified as key steps in the re-design of the *T. maritima* metabolism are shown with black squares. Metabolites are represented as colored balls according to the pathway they are involved in. Extracellular metabolites are indicated by red balls.

Figure 3. Growth rate and flux across the main metabolic pathways involved in redox balancing as a function of the fdxr/NADH ratio.

Figure 4. Analysis of knockout mutants.

Figure 5. **The main characteristics of the *in silico* overproducing strains.** NADP_H2 and FNOR_HT represent the respective knock in *in silico* strains. Relative flux across the EM pathway, ED pathway, and OPP (**A**). Secretion of H_2, CO_2, and acetate (mmol/gDW/h) (**B**). The H_2 production and the flux across the pyruvate synthase for selected knockout mutants of the wild-type (**C**), NADP_H2 knock-in (**D**) and FNOR_HT knock-in (**E**).

Figure 6. A qualitative description of the metabolism of the *T. maritima* wild type and the re-designed strains. The thickness of the pathways indicates the amount of flux present.

Table legends

Table 1. **Growth rate and hydrogen production from glucose and glycerol for different mutant strains**. Growth rate (h^{-1}). H_2, CO_2, acetate, PFK, EDD, GND, RPE

represent flux values of the corresponding reactions (mmol/gDW/h). The PFK reaction represents EM pathway, EDD represents the ED pathway, GND represents the OPP pathway and RPE represents the non-oxidative branch of the pentose phosphate pathway. BPCY is the biomass-coupled product yield (mmol/gDW/h^2) and ΔH_2 is the fractional increase in H_2 production over the wild-type.

Supporting material

Figure legends

Figure S1: Growth rate, hydrogen production and biomass-coupled product yield for each of the carbon sources, which supported growth in the updated model.

Figure S2: Growth rate, hydrogen production and biomass-coupled product yield for each of the carbon sources, which supported growth in the original model.

Table legends

Table S1: A description of model updates.

Table S2: Essential reactions in *T. maritima* grown on glucose.

Table S3: Hydrogen yields for knockout strains based on the wild type.

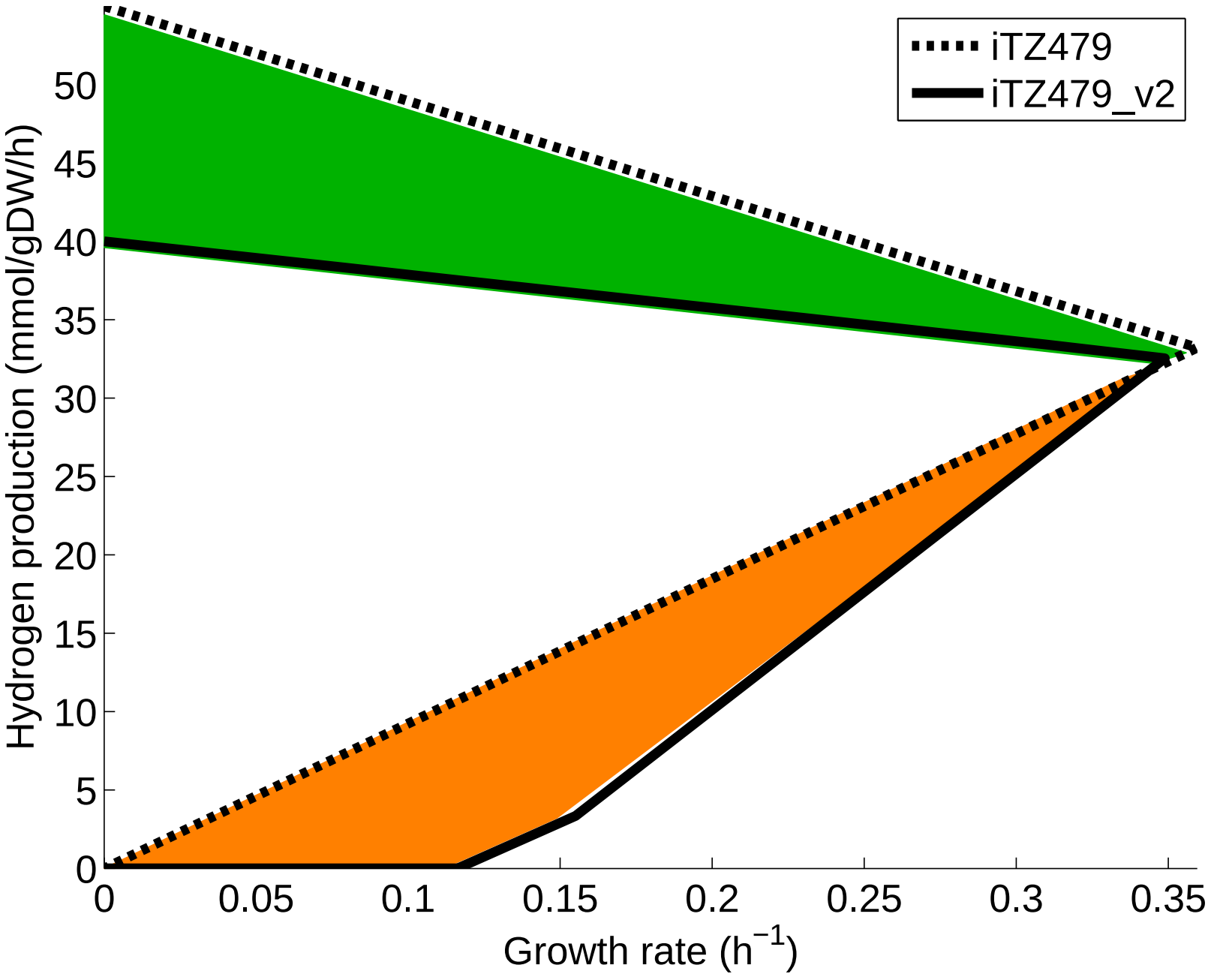
Table S4: Hydrogen yields for knockout strains based on the NADPH_2 knock-in mutant.

Table S5: Hydrogen yields for knockout strains based on the FNOR_HT knock-in mutant.

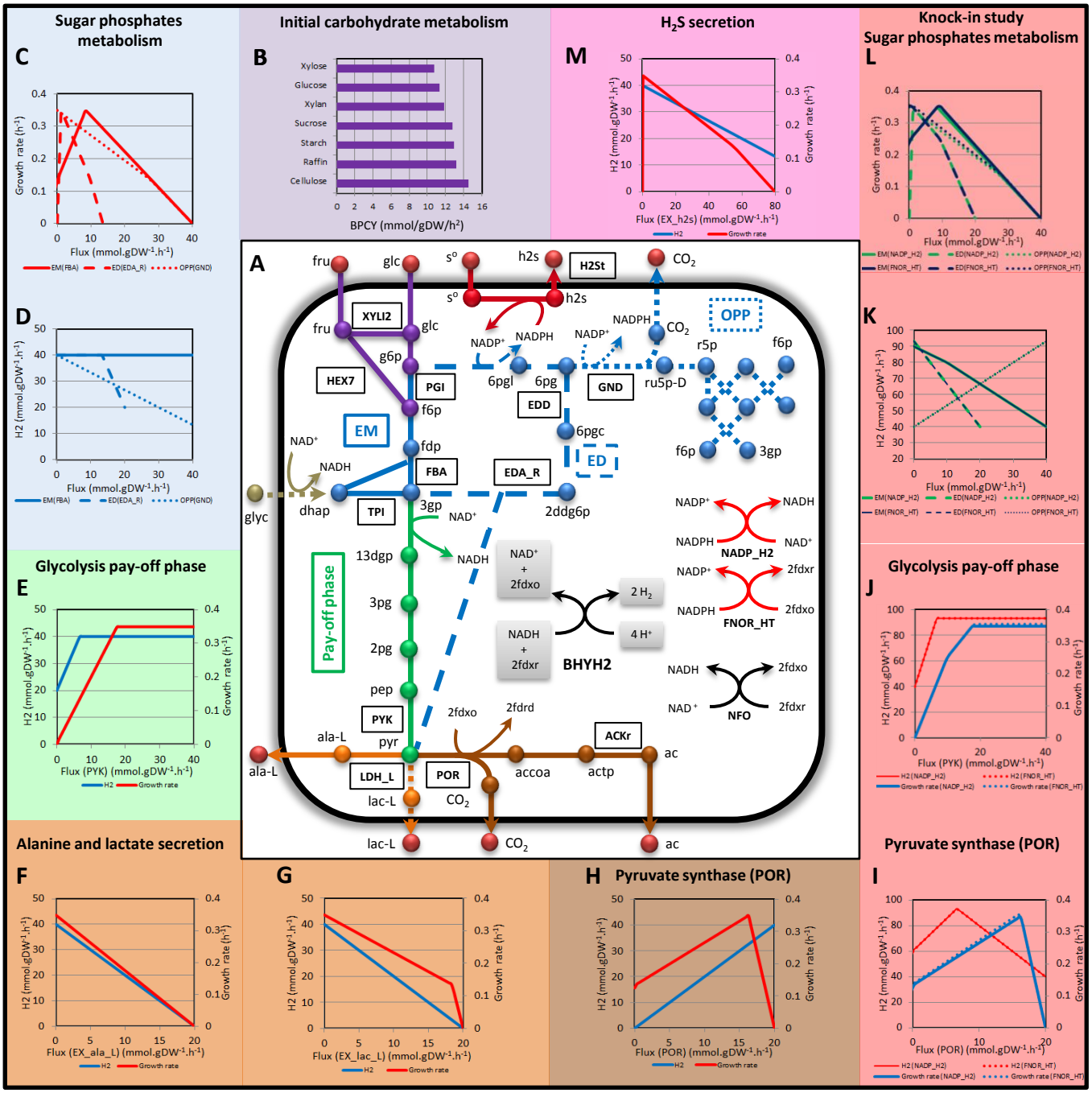
Table

Carbon source	Glucose				Glucose					Glucose						Glycerol		
Knock in	Wild type				NADP_H2					FNOR_HT						FNOR_HT		
Knock out	-	FBA	TPI	ACKr / LDH_L	-	FBA	TPI	HEX7 / PGI	ACKr / LDH_L	-	FBA	TPI	HEX7 / PGI	FBA / EDA_R	ACKr / LDH_L	-	ACKr/ LDH_L	FBA/EDA_R/TPI
Growth rate	0.35	0.13	0.25	0.14	0.35	0.24	0.25	0.28	0.14	0.35	0.24	0.25	0.28	0.12	0.14	0.34	0.18	0.12
H2	32.53	36.83	34.74	37.07	32.95	35.07	35.03	46.45	37.22	34.06	34.62	35.03	47.57	76.36	37.63	67.65	70.33	96.83
BPCY	11.35	4.84	8.54	5.08	11.50	8.42	8.61	12.94	5.10	12.09	8.31	8.61	13.42	9.32	5.33	22.97	12.61	11.82
ΔH2	1.00	1.13	1.07	1.14	1.00	1.06	1.06	1.41	1.13	1.00	1.02	1.03	1.40	2.24	1.10	1.00	1.04	1.43

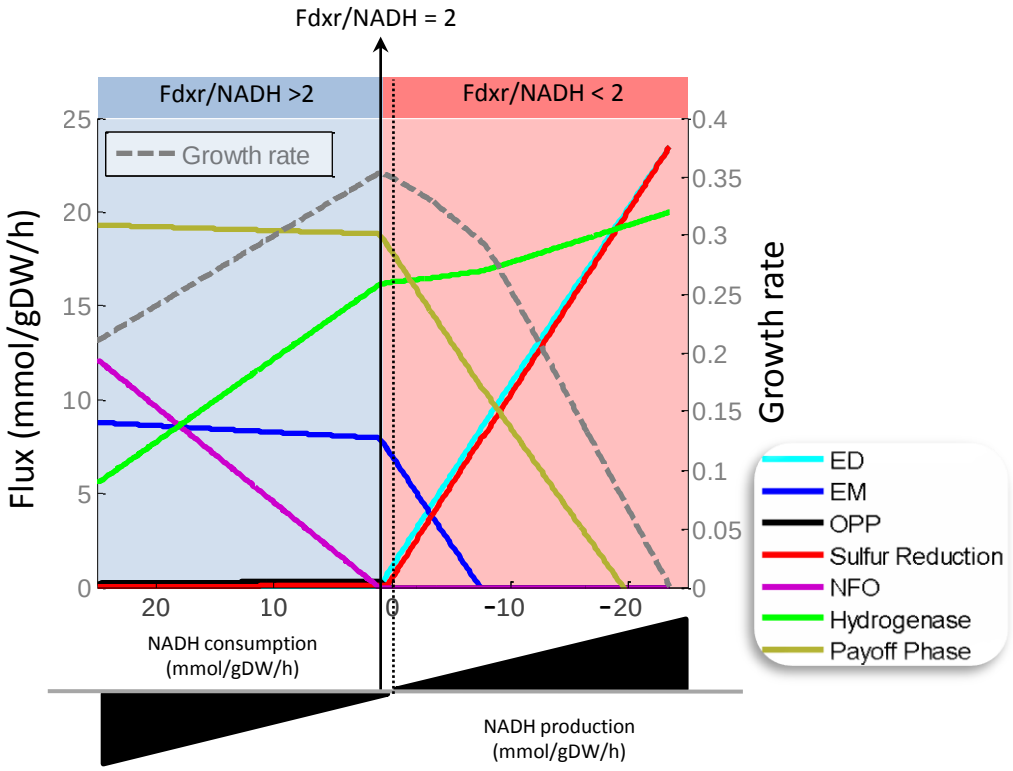
Figure


iTZ479
iTZ479_v2
Hydrogen production (mmol/gDW/h)
0
5
10
15
20
25
30
35
40
45
50
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Growth rate (h^{-1})

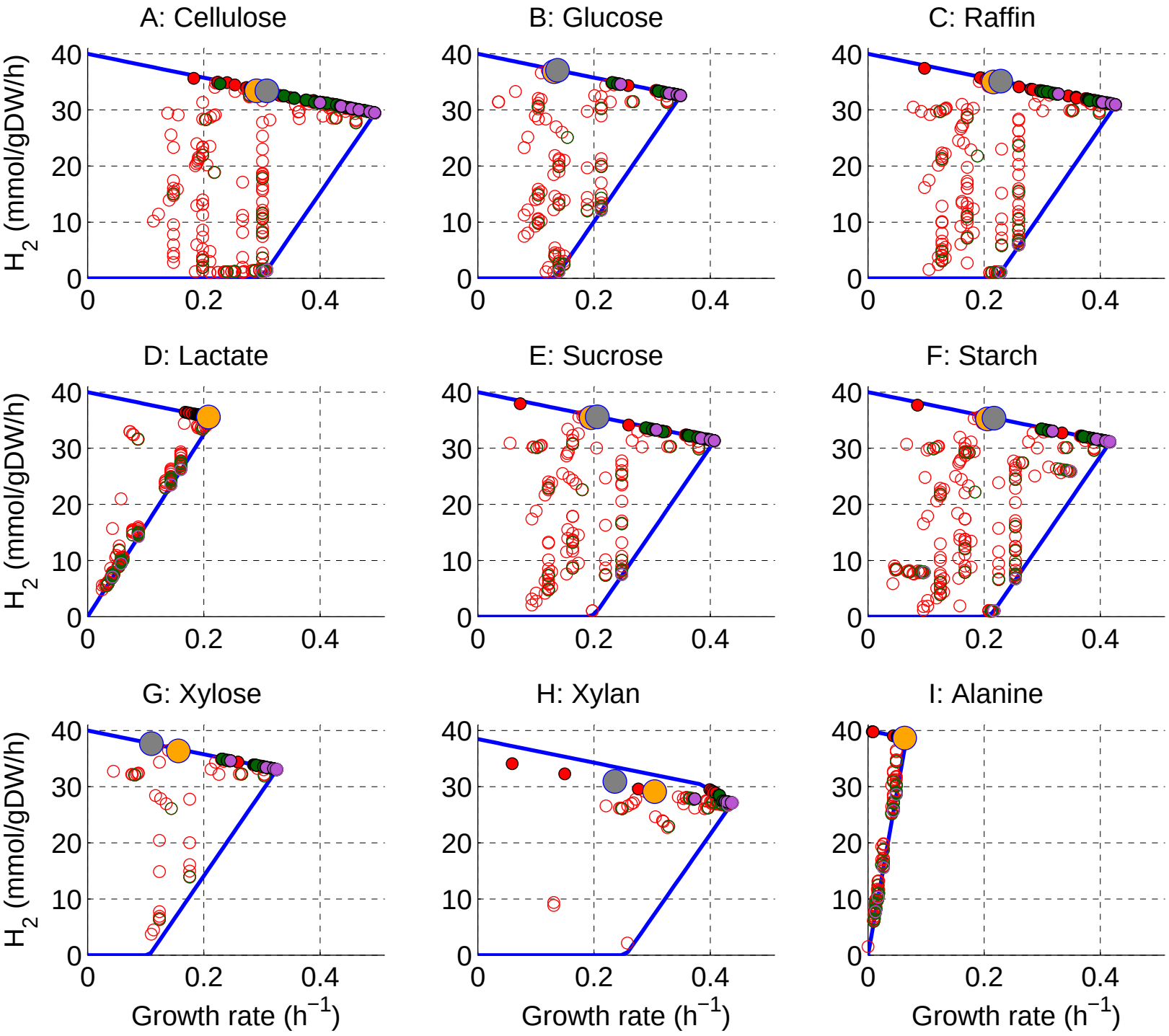
Figure


Sugar phosphates metabolism
C
Growth rate (h-1)
Flux (mmol.gDW-1.h-1)
EM(FBA)
ED(EDA_R)
OPP(GND)
D
H2 (mmol.gDW-1.h-1)
Flux (mmol.gDW-1.h-1)
EM(FBA)
ED(EDA_R)
OPP(GND)
Initial carbohydrate metabolism
B
Xylose
Glucose
Xylan
Sucrose
Starch
Raffin
Cellulose
BPCY (mmol/gDW/h2)
A
fru
glc
s°
h2s
H2St
CO2
XYLI2
HEX7
PGI
EM
FBA
TPI
dhap
glyc
NAD+
NADH
3gp
13dgp
3pg
2pg
pep
Pay-off phase
PYK
pyr
ala-L
LDH_L
lac-L
POR
CO2
accoa
ACKr
actp
ac
NADPH
NADP+
OPP
6pgl
6pg
GND
ru5p-D
r5p
f6p
3gp
EDD
6pgc
ED
EDA_R
2ddg6p
NADP_H2
FNOR_HT
NFO
BHYH2
2fdxo
2fdxr
2fdrd
2 H2
4 H+
H2 (mmol.gDW-1.h-1)
Growth rate (h-1)
H2
Growth rate
H2S secretion
M
Flux (EX_h2s) (mmol.gDW-1.h-1)
Knock-in study
Sugar phosphates metabolism
L
Flux (mmol.gDW-1.h-1)
EM(NADP_H2)
ED(NADP_H2)
OPP(NADP_H2)
EM(FNOR_HT)
ED(FNOR_HT)
OPP(FNOR_HT)
K
Flux (mmol.gDW-1.h-1)
Glycolysis pay-off phase
E
Flux (PYK) (mmol.gDW-1.h-1)
Glycolysis pay-off phase
J
Flux (PYK) (mmol.gDW-1.h-1)
H2 (NADP_H2)
H2 (FNOR_HT)
Growth rate (NADP_H2)
Growth rate (FNOR_HT)
Alanine and lactate secretion
F
Flux (EX_ala_L) (mmol.gDW-1.h-1)
G
Flux (EX_lac_L) (mmol.gDW-1.h-1)
Pyruvate synthase (POR)
H
Flux (POR) (mmol.gDW-1.h-1)
Pyruvate synthase (POR)
I
Flux (POR) (mmol.gDW-1.h-1)

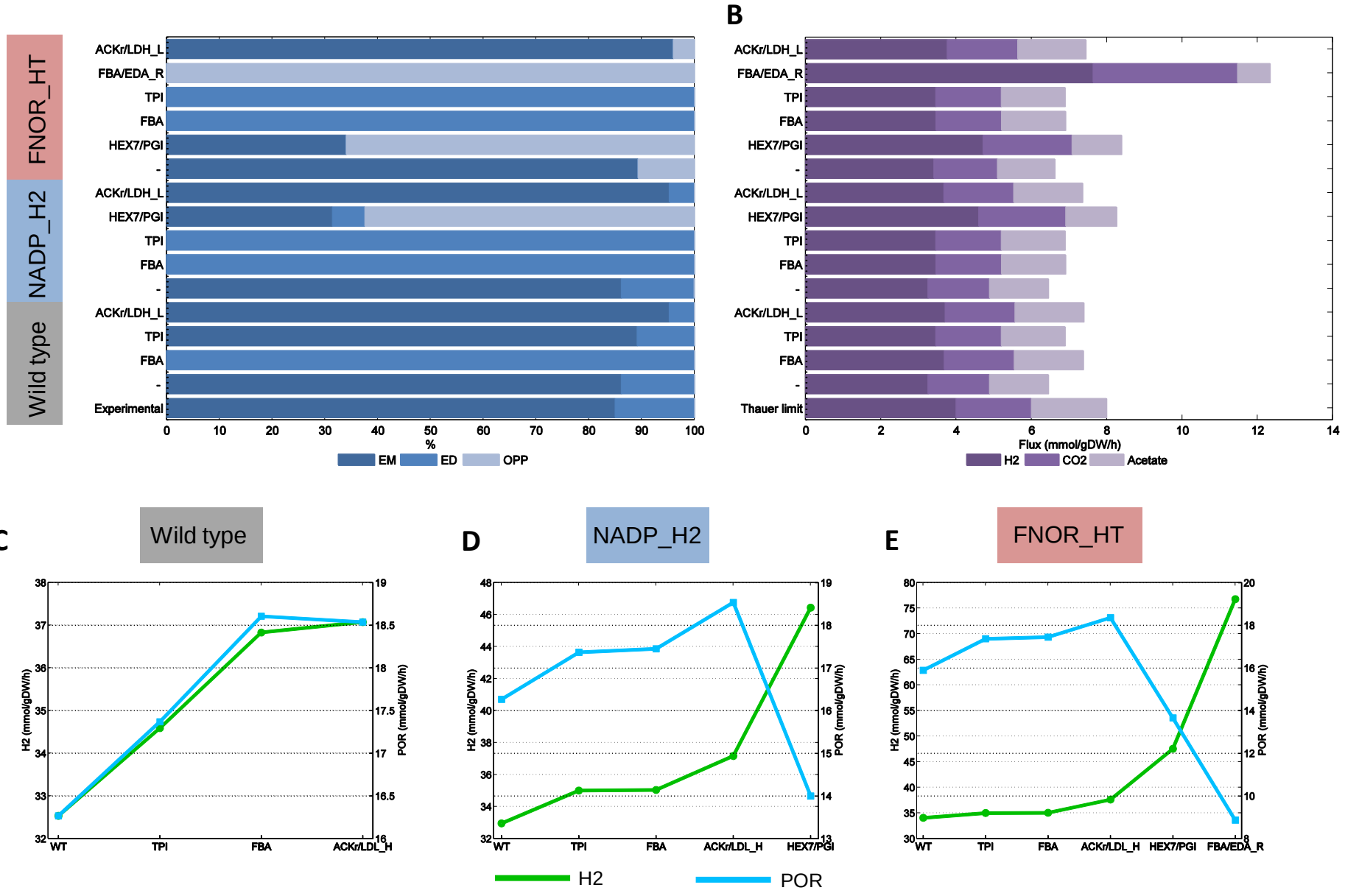
Figure


Fdxr/NADH = 2
Fdxr/NADH >2
Fdxr/NADH < 2
Growth rate
Flux (mmol/gDW/h)
Growth rate
25
20
15
10
5
0
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
20
10
0
-10
-20
NADH consumption
(mmol/gDW/h)
NADH production
(mmol/gDW/h)
ED
EM
OPP
Sulfur Reduction
NFO
Hydrogenase
Payoff Phase

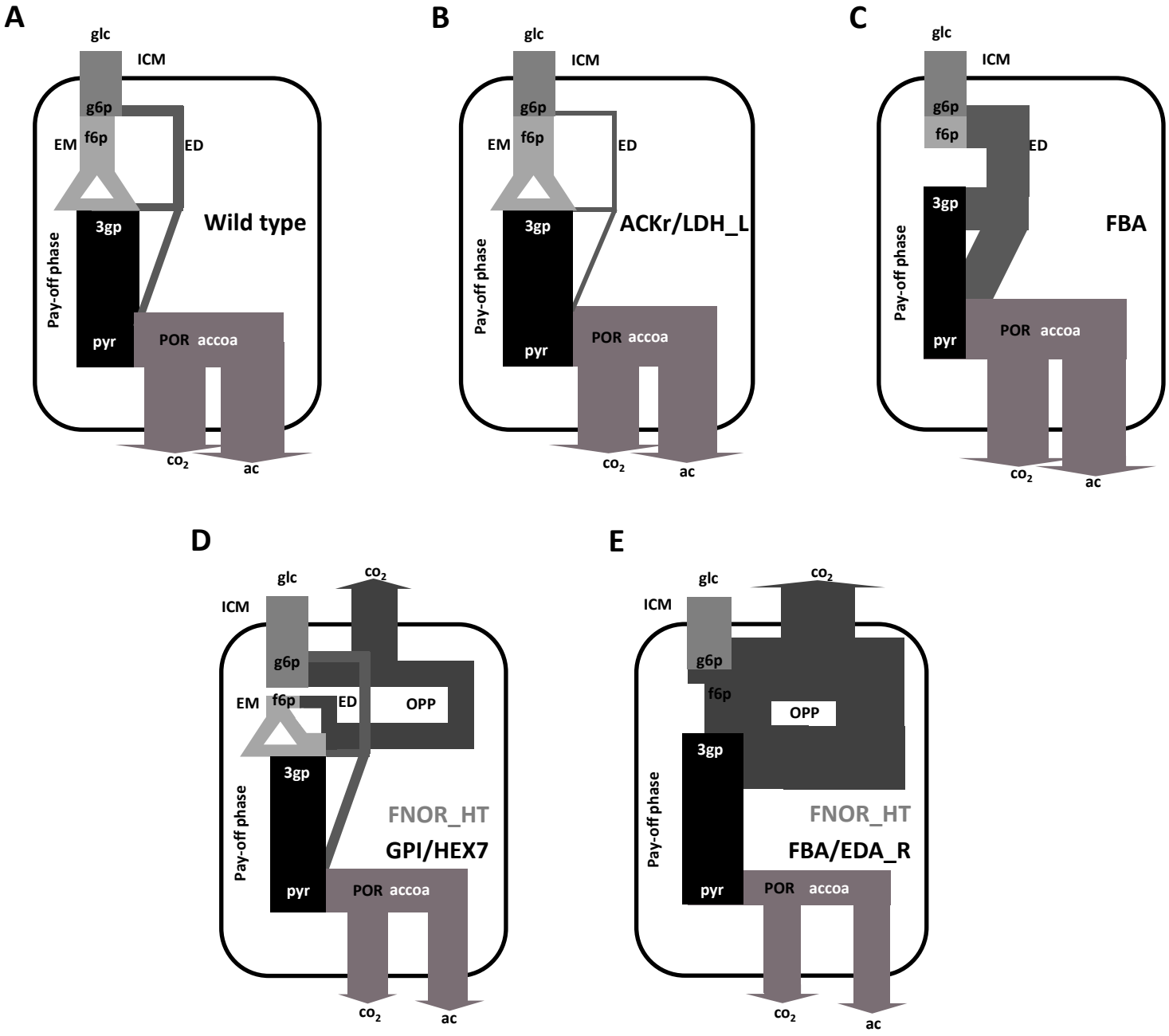
Figure


A: Cellulose
B: Glucose
C: Raffin
D: Lactate
E: Sucrose
F: Starch
G: Xylose
H: Xylan
I: Alanine
H2 (mmol/gDW/h)
Growth rate (h−1)
0
10
20
30
40
0.2
0.4

Figure


A
FNOR_HT
NADP_H2
Wild type
ACKr/LDH_L
FBA/EDA_R
TPI
FBA
HEX7/PGI
-
ACKr/LDH_L
HEX7/PGI
TPI
FBA
-
ACKr/LDH_L
TPI
FBA
-
Experimental
0 10 20 30 40 50 60 70 80 90 100
%
EM ED OPP
B
ACKr/LDH_L
FBA/EDA_R
TPI
FBA
HEX7/PGI
-
ACKr/LDH_L
HEX7/PGI
TPI
FBA
-
ACKr/LDH_L
TPI
FBA
-
Thauer limit
0 2 4 6 8 10 12 14
Flux (mmol/gDW/h)
H2 CO2 Acetate
C
Wild type
H2 (mmol/gDW/h)
POR (mmol/gDW/h)
WT TPI FBA ACKr/LDL_H
D
NADP_H2
H2 (mmol/gDW/h)
POR (mmol/gDW/h)
WT TPI FBA ACKr/LDL_H HEX7/PGI
E
FNOR_HT
H2 (mmol/gDW/h)
POR (mmol/gDW/h)
WT TPI FBA ACKr/LDL_H HEX7/PGI FBA/EDA_R
H2
POR

Figure


A
glc
ICM
g6p
f6p
EM
ED
3gp
Pay-off phase
pyr
POR accoa
Wild type
co2
ac
B
glc
ICM
g6p
f6p
EM
ED
3gp
Pay-off phase
pyr
POR accoa
ACKr/LDH_L
co2
ac
C
glc
ICM
g6p
f6p
ED
3gp
Pay-off phase
pyr
POR accoa
FBA
co2
ac
D
glc
co2
ICM
g6p
EM
f6p
ED
OPP
3gp
Pay-off phase
pyr
POR accoa
FNOR_HT
GPI/HEX7
co2
ac
E
glc
co2
ICM
g6p
f6p
OPP
3gp
Pay-off phase
pyr
POR accoa
FNOR_HT
FBA/EDA_R
co2
ac