Integrative Biology

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/ibiology

Insight, innovation, integration

Moorella thermoacetica is a thermophilic acetogen that employs the Wood-Ljungdahl pathway of CO_2-fixation for converting a range of gaseous substrates and organic compounds into acetyl-CoA — a critical precursor for biological production of various commodity chemicals. Thus, *M. thermoacetica* is very promising to use as a biological chassis for industrial biotechnology applications. In this paper, we describe the construction of a genome-scale constraint-based model of *M. thermoacetica* metabolism that help elucidate its metabolic potentials and limitations, as well as the energy conservation process during autotrophy. Most importantly, the genome-wide model now provides a platform to design and engineer efficient microbial cell factories using *M. thermoacetica*.

Investigating *Moorella thermoacetica* Metabolism with a Genome-Scale Constraint-Based Metabolic Model

M. Ahsanul Islam[a], Karsten Zengler[b], Elizabeth A. Edwards[c], Radhakrishnan Mahadevan[c,*], and Gregory Stephanopoulos[a,*]

[a]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[b]Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

[c]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, CANADA

*Corresponding authors: gregstep@mit.edu and krishna.mahadevan@utoronto.ca

Abstract

Moorella thermoacetica is a strictly anaerobic, endospore-forming, and metabolically versatile acetogenic bacterium capable of conserving energy by both autotrophic (acetogenesis) and heterotrophic (homoacetogenesis) modes of metabolism. Its metabolic diversity and the ability to efficiently convert a wide range of compounds, including syngas (CO_2 + H_2) into acetyl-CoA have made this thermophilic bacterium a promising host for industrial biotechnology applications. However, lack of detailed information on *M. thermoacetica*'s metabolism is a major impediment to its use as a microbial cell factory. In order to overcome this issue, a genome-scale constraint-based metabolic model of *Moorella thermoacetica*, *i*AI558, has been developed using its genome sequence and physiological data from published literature. The reconstructed metabolic network of *M. thermoacetica* comprises 558 metabolic genes, 705 biochemical reactions, and 698 metabolites. Of the total 705 model reactions, 680 are gene-associated while the rest are non-gene associated reactions. The model, in addition to simulating both autotrophic and heterotrophic growth of *M. thermoacetica*, revealed degeneracy in its TCA-cycle, a common characteristic of anaerobic metabolism. Furthermore, the model helped elucidate the poorly understood energy conservation mechanism of *M. thermoacetica* during autotrophy. Thus, in addition to generating experimentally testable hypotheses regarding its physiology, such a detailed model will facilitate rapid strain designing and metabolic engineering of *M. thermoacetica* for industrial applications.

Introduction

Acetogenesis is the microbial metabolic process of acetate formation from carbon dioxide (CO_2) and hydrogen (H_2) through a metabolic pathway called the reductive acetyl-CoA pathway, also known as the Wood-Ljungdahl pathway (WLP)[1, 2]. Unlike other prokaryotic CO_2-fixing pathways, the reductive acetyl-CoA pathway is the most simplified, efficient, and only linear pathway to synthesize acetyl-CoA from CO_2 without requiring any recycled intermediates[3-5]. Although many organisms, including methanogens, sulfate reducing bacteria, and methanogenic euryarchaeota, use this ancient autotrophic pathway for CO_2-fixation or acetate-oxidation, only acetogens — a group of obligate anaerobic bacteria — use WLP for simultaneous assimilation of CO_2 into cell-biomass and ATP generation to conserve energy by converting acetyl-CoA to acetate[3, 4]. Acetogens are also capable of conserving energy and producing acetate as an exclusive fermentation product from sugars, including glucose, fructose, and xylose, via WLP by a process termed homoacetogenesis[6, 7]. Thus, acetogens can conserve energy by both electron transport phosphorylation and substrate level phosphorylation[1, 2]. This versatility in energy conservation processes, together with their capability of using a wide range of organic and inorganic compounds as electron donors and acceptors, enable them to play critical roles in many diverse ecosystems, such as soils, sediments, sludge, and intestinal tracts of termites and humans[1-3].

Moorella thermoacetica (formerly *Clostridium thermoaceticum*[8]) is a strictly anaerobic, spore-forming, thermophilic, Gram-positive acetogen that has been extensively studied as a model organism for over 50 years to identify and characterize the enzymes of WLP[2, 9-14]. *M. thermoacetica* uses a wide variety of electron donors and acceptors, including H_2, CO_2, CO,

sugars, alcohols, organic acids, methoxylated aromatic compounds, nitrate, nitrite, thiosulfate, and dimethylsulfoxide [1-3]. Its metabolic diversity and the ability to efficiently convert syngas (CO_2 + H_2) into acetyl-CoA have made this thermophilic bacterium a promising host for industrial biotechnology. In order to harness and utilize *M. thermoacetica*'s diverse metabolic capabilities to engineer platform strains for bio-based industrial applications, detailed information on its metabolic characteristics is an essential necessity. Moreover, the metabolic potentials and possible metabolic bottlenecks of this organism are not well-understood despite having extensive research on *M. thermoacetica* physiology[1-3, 15]. Also, the energy conserving mechanism of this bacterium during autotrophy, although many details only have recently been discovered[16-20], is still debatable. To investigate these issues in details, a genome-scale constraint-based metabolic model of an organism using the constraint-based reconstruction and analysis (COBRA) approach[21-24] is very useful. COBRA utilizes the flux balance analysis (FBA) technique[25, 26] to quantitatively simulate and analyze the metabolic potentials and phenotypes of an organism using its metabolic network reconstructed from a plethora of the organism specific information. Although such a systems-level model is available for another acetogen, *Clostridium ljungdahlii*, no genome-scale model has been developed for *M. thermoacetica* to date.

Herein, we describe the construction of a genome-wide constraint-based metabolic model, *i*AI558, for *Moorella thermoacetica*. The model primarily is based on the organism's genome sequence, and was extensively curated using experimental growth data and published literature regarding its physiology. We used the model to characterize *M. thermoacetica*'s metabolic phenotypes on H_2, CO_2, CO, methanol, glucose, fructose, and xylose. The model was also compared with the published model of *C. ljungdahlii*[27] to identify the differences in central

metabolic processes between these acetogens. Additionally, the metabolic genes involved in *M. thermoacetica*'s energy metabolism were compared with those of the three other related clostridial species — *Clostridium acetobutylicum*, *Clostridium cellulolyticum*, *Clostridium thermocellum* — and another acetogen, *Acetobacterium woodii*. Finally, the model was used to obtain a better understanding of the unresolved energy conserving mechanism of *M. thermoacetica* during its autotrophic metabolism.

Materials and methods

Generation of a draft metabolic model for *Moorella thermoacetica*

Using the publicly accessible *M. thermoacetica* genome sequence[28], the reconstructed metabolic network and a draft metabolic model for the bacterium were automatically generated using the rapid annotation using subsystem technology (RAST) annotation server[29] and the Model SEED platform[30] in SEED[31]— a freely accessible, comprehensive web-based environment for performing comparative genomic analyses and developing highly curated genomic data (Figure 1). First, the genome-sequence of *M. thermoacetica* was uploaded into RAST, which re-annotated the genome using information from the curated SEED database[31]. Then, Model SEED generated a preliminary reconstructed network using the RAST annotations and other published genome-scale reconstructions in SEED, as well as a template biomass reaction[30]. This preliminary network included network gaps, i.e., missing reactions for producing all precursor metabolites present in the biomass reaction. The gaps were reconciled during the auto completion stage[30] by adding necessary intracellular and transport reactions to generate an analysis-ready model. This analysis-ready draft model was capable of simulating *M.*

thermoacetica growth using FBA. However, this draft model, as well as the reconstructed network, required extensive manual curation because it included a template biomass reaction from SEED, which was not specific to the biomass composition of *M. thermoacetica*.

Determination of *Moorella thermoacetica* biomass composition

The detailed biomass composition, including the amount of all cellular macromolecules, such as amino acids, DNA, RNA, fatty acids, and lipids, of *M. thermoacetica* was generated from its genome sequence, biochemical database, and published literature describing its cellular compositions and physiology (see Tables S2 – S9 in the supplementary information 1). The overall percentage composition of macromolecules in the *M. thermoacetica* cell was assumed to be similar to the *Bacillus subtilis* cell, and was obtained from the genome-scale *Bacillus subtilis* model[32] as both *M. thermoacetica* and *B. subtilis* are Gram-positive bacteria. The detailed composition of all 20 amino acids, DNA, and RNA was estimated from the *M. thermoacetica* genome sequence[28], while the composition of fatty acids was obtained from published literature[33]. Due to the lack of *M. thermoacetica*-specific data, the compositions of cell-wall, lipids, and other ions and cofactors were assumed to be the same as the *Bacillus subtilis* model[32]. The non-growth associated maintenance (NGAM) parameter was estimated from the experimental growth rate and yield (see text in the supplementary information 1 for details) of *M. thermoacetica* on glucose, and the growth-associated maintenance (GAM) parameter was iteratively determined using model simulations. All compositions were calculated using the basis of one gram dry cell weight.

Curation of the draft metabolic model of *Moorella thermoacetica*

Extensive curation of the draft metabolic model of *M. thermoacetica* was achieved by following the steps shown in Figure 1. First, the naming of model reactions and metabolites was changed from the Model SEED notations to the BiGG database[34] notations to make them consistent with other curated and published genome-scale models. This conversion process involved the mapping of metabolites and reactions between SEED and BiGG using the BiGG notations as keys. The SEED metabolites and reactions not found in BiGG were manually changed to BiGG notations in the model. The conversion to BiGG notations also helped check the consistency of metabolites and reactions included in the model. Next, the template biomass reaction in the draft model was replaced with the *M. thermoacetica*-specific one containing all biomass compositions, including the NGAM and GAM parameters estimated earlier. Then, FBA was used to generate all biomass precursors using the *M. thermoacetica* minimal growth medium[3] and the reconstructed metabolic network. This improved network was further curated by reconciling with published models of related clostridial species (*C. ljungdahlii*[27], *C. acetobutylicum*[35], *C. cellulolyticum*[36], and *C. thermocellum*[37]). The *M. thermoacetica* genome was compared with the genomes of *A. woodii*[38], *C. ljungdahlii*[39], *C. acetobutylicum*[40], *C. cellulolyticum*[41], and *C. thermocellum*[41] to check the correctness of gene-annotations in the reconstructed network using homology-based inference techniques such as BLAST[42] and phylogenetic profiling[43]. Annotations of the experimentally characterized *M. thermoacetica* metabolic genes, as well as the ones homologous to all compared genomes, were given high confidence; the metabolic genes found in all acetogens only were given medium confidence, while low confidence was assigned to metabolic genes present only in *M. thermoacetica*. Once this curated metabolic network was capable of generating all biomass precursors in the minimal medium, both autotrophic and

heterotrophic modes of metabolism of *M. thermoacetica* were simulated using the genome-wide metabolic model.

In silico analysis of *Moorella thermoacetica* metabolism

FBA was used for *in silico* analysis and simulation of *M. thermoacetica* autotrophic growth on H_2, CO_2, CO, and methanol and heterotrophic growth on glucose, fructose, and xylose. The COBRA toolbox[44] was used for FBA implementation, and all simulations were conducted in MATLAB (the MathWorks Inc., Natick, MA) using the IBM ILOG® CPLEX 12.5.1 optimization solver. The model reactions and genes, as well as the details of constraints used to simulate *M. thermocaetica* growth with *i*AI558, is provided in Tables 3 – 5 in the supplementary information 2. Two sbml model files (*i*AI558_co2_h2.xml and *i*AI558_glucose.xml) are also provided in the supplementary data.

Results and discussion

Moorella thermoacetica reconstructed metabolic network

The reconstructed metabolic network of *M. thermoacetica*, also denoted as *i*AI558 according to the established naming convention[45], comprises 558 metabolic genes, representing 22% genes of a genome of 2.6 Mbp (Table S1 in the supplementary information 1). In total, the network includes 705 biochemical reactions, of which 680 are gene-associated reactions while 25 are non-gene associated; non-gene associated reactions are either spontaneous and diffusion reactions (10 in total), or added to the network during the gap-filling procedure (15 reactions) to produce biomass precursor metabolites. In addition to the biomass demand reaction containing

M. thermoacetica-specific biomass compositions, the network also comprises 59 exchange reactions, including components of the *in silico* minimal growth medium (Table 1) and transporters. To have a better understanding of *M. thermoacetica* metabolic processes, the biochemical reactions of *i*AI558 were further categorized based on the metabolic pathways, also known as model subsystems, in which they were involved. For instance, reactions involved in glycolysis/gluconeogenesis, the tricarboxylic acid (TCA)-cycle, the pentose phosphate pathway, and carbohydrate metabolism were referred to as "central carbon metabolism" reactions, while reactions involved in vitamins and cofactors synthesis were included in the "cofactor metabolism" category. This classification of metabolic reactions led to the identification of 9 model subsystems in the *M. thermoacetica* metabolic model (Figure 2). The subsystem "cofactor metabolism" has the highest number of reactions followed by "amino acid metabolism" and "lipid metabolism" (Figure 2). The fact that *M. thermoacetica* can generate a wide variety of corrinoids[11, 46] probably results in the highest number of reactions in the cofactor metabolism category.

Comparative analysis of *Moorella thermoacetica* metabolism

The *M. thermoacetica* genome sequence was compared with the genomes of two other model acetogens and three related clostridial species (Figure S1 in the supplementary information 1) during the construction of *i*AI558. The comparative analysis showed that *C. ljungdahlii* has the highest number (1034) of *M. thermoacetica* orthologs followed by *C. cellulolyticum* with 879 orthologous genes. Both *A. woodii* and *C. thermocellum* include 854 orthologs, while *C. acetobutylicum* has the lowest number of orthologous genes (829) among the 5 genomes compared (Figure S1 in the supplementary information 1). Thus the presence of only 41% (1034

of 2523) and 34% (854 of 2523) orthologous genes in *C. ljungdahlii* and *A. woodii* suggests the different nature of *M. thermoacetica* metabolism in spite of all being acetogens. This difference is further evident from the presence of only ~300 orthologous metabolic genes in *C. ljungdahlii* and *A. woodii* (Figure S1 in the supplementary information 1) out of 558 *M. thermoacetica* metabolic genes. Because WLP/ the reductive acetyl-CoA pathway is a characteristic feature of all acetogens[2, 11, 20], the genes involved in the pathway should be conserved among them.

However, a comparison of the *M. thermoacetica* genes involved in WLP and autotrophic energy metabolism showed that not all genes are conserved among the 5 genomes compared, including the 2 acetogens (Figure 3). Out of the total 46 *M. thermoacetica* genes, only 23 are orthologous to *A. woodii* while *C. ljungdahlii* includes 18 orthologs. Among the other three clostridial species, *C. cellulolyticum* and *C. thermocellum* possess 17 and 14 orthologs, respectively, but *C. acetobutylicum* has only 7 (Figure 3). Notably, no orthologs were identified for 11 *M. thermoacetica* genes: Moth_1193, Moth_1195, and Moth_1196 encoding a putative electron bifurcating 5, 10-methylenetetrahydrofolate reductase; Moth_2312 encoding a formate dehydrogenase; Moth_1181 encoding a phosphotransacetylase; Moth_0451, Moth_0452, Moth_2185, Moth_2189, and Moth_2195 encoding a putative proton-translocating ferredoxin hydrogenase; and Moth_1885 encoding an $NADP^+$-reducing hydrogenase (Figure 3). Due to this unique gene repertoire, the encoded enzymes involved in WLP and autotrophic energy metabolism in *M. thermoacetica* are different from *C. ljungdahlii* and *A. woodii*. For instance, the formate dehydrogenase in *M. thermoacetica* is $NADP^+$-specific[11, 28], but the substrate specificity of the enzyme is different in *C. ljungdahlii*[39] and *A. woodii*[38]. Also, *M. thermoacetica* has a putative proton-translocating, energy-converting hydrogenase (Ech) complex/ formate-

hydrogen lyase complex[16, 17] involved in energy conservation[23], while *C. ljungdahlii* and *A. woodii* contain the energy-conserving Rnf complex instead[20, 47].

Interestingly, orthologs for only 3 genes were identified in all genomes and these included Moth_0109 (formate-tetrahydrofolate ligase), Moth_0940 (acetate kinase), and Moth_1516 (bifunctional 5, 10-methenyltetrahydrofolate cyclohydrolase/ 5, 10-methylenetetrahydrofolate dehydrogenase); this finding suggests a potential involvement of these genes in other metabolic functions apart from WLP. In fact, both Moth_0109 and Moth_1516 are also required for folate biosynthesis, while Moth_0940 is involved in pyruvate metabolism. Despite having a number of WLP genes, the absence of orthologs for Moth_1197 (5-methyltetrahydrofolate corrinoid iron-sulfur protein methyltransferase), and Moth_1198 – Moth_1203 (carbon monoxide dehydrogenase/ acetyl-CoA synthase complex) (Figure 3) is possibly responsible for not having a functional WLP in other clostridial genomes, except in *C. ljungdahlii*. Thus the metabolic gene repertoire and hence the metabolism of *M thermoacetica*, although shares similarity with the 5 genomes compared, is quite different.

Comparison of the *M. thermoacetica* model (*i*AI558) with the *C. ljungdahlii* model (*i*HN637)

We compared *i*AI558 with the recently published genome-scale model of *C. ljungdahlii*, *i*HN637[27] to obtain an in-depth look at the metabolic differences between these model acetogens. Although the *C. ljungdahlii* genome has 1.7 times more protein coding genes than *M. thermoacetica* (4184 genes[39] vs. 2523 genes[28]) and *i*HN637 contains more metabolic genes than *i*AI558, the latter actually contains and predicts functions for more metabolic genes (22% or 558 of 2523) than the former (15% or 637 of 4184) from the perspective of respective genomes.

Thus, *M. thermoacetica* is metabolically more diverse than *C. ljungdahlii* even though the *C. ljungdahlii* genome is nearly twice as big (4.63 Mbp[39] vs 2.62 Mbp[28]) as the *M. thermoacetica* genome. The metabolic versatility of *M. thermoacetica* is further evident from the range of compounds it can use as electron donors and acceptors[2, 3] to support its growth. However, the developed model cannot use all growth supporting compounds to simulate *M. thermoacetica* metabolism because pathway-level details are available only for a few substrates, including H_2, CO_2, CO, sugars (glucose, fructose, and xylose), and methanol[2, 3, 11, 28].

Both *M. thermoacetica* and *C. ljungdahlii* employ WLP during the autotrophic mode of metabolism; however, the potential energy-conserving mechanism during autotrophy is different[20] in these organisms due to the difference in involved enzymes. As described in *i*HN637[27], *C. ljungdahlii* requires a proton-translocating Rnf complex reaction during its autotrophic growth, while a proton-translocating Ech complex reaction is essential for *i*AI558 to simulate *M. thermoacetica*'s autotrophic growth. Also, both models contain two electron-bifurcating-type transhydrogenase reactions (FRNDPRr and HYDFDNr); although both reactions are $NADP^+$-dependent in *i*HN637[27], one is $NADP^+$-dependent (FRNDPRr) and the other is NAD^+-dependent (HYDFDNr) in *i*AI558. Notably, only the *M. thermoacetica* enzymes catalyzing these reactions were experimentally characterized[16, 18]. Comparison of the central metabolic pathways revealed the presence of complete glycolysis/ gluconeogenesis and pentose phosphate pathways, but a degenerate TCA-cycle in both *M. thermoacetica* and *C. ljungdahlii* models (Figure 4); the nature of degeneracy is different though. The *M. thermoacetica* TCA-cycle is missing the fumarate reductase (FRD) reaction, while the succinyl-CoA synthetase (SUCOAS) and 2-oxoglutarate synthase (OOR) reactions are absent from the *C. ljungdahlii*

TCA-cycle (Figure 4) due to missing genes in respective genomes. Thus, the degenerate TCA-cycle in both models provides only the precursors for biosynthetic or anabolic reactions without any participation in catabolic ATP generation.

Model-based simulation of *Moorella thermoacetica* growth

The genome-scale model of *M. thermoacetica*, *i*AI558 was used to simulate the bacterium's autotrophic growth on H_2, CO_2, CO, and methanol, and heterotrophic growth on glucose, fructose, and xylose, as well as the ATP production rate and yield on these substrates (Figure 5). The model simulations were conducted using the *in silico* minimal medium (Table 1), and the simulated growth rates were compared with the available experimental growth data (Figure 5A) of *M. thermoacetica* in an actual minimal medium[15]. The sbml files of both autotrophic and heterotrophic models (*i*AI558_co2_h2.xml and *i*AI558_glucose.xml) are presented in the supplementary data. The model was first trained to simulate the growth rate on glucose using the experimental data[15], and then used to predict the growth rates on other substrates. The predicted growth rates on H_2-CO_2 and CO are similar to the experimental data (Figure 5A), which indicates the model's capability to simulate the organism's growth behavior on different substrates. Notably the growth rate and yield, as well as the ATP production rate and yield are the highest on CO among the autotrophic substrates (Figure 5A and 5B) because *i*AI558 uses the carbon monoxide dehydrogenase (CODH) reaction to generate extra reducing equivalents (reduced ferredoxin) only when CO is used as growth substrate. Clearly, the autotrophic growth rates and yields of *M. thermoacetica* are slower than those in the heterotrophic mode (Figure 5A) because less ATP is produced in the former mode of metabolism than the latter (Figure 5B). This

phenomenon can be further clarified from analyzing the model simulated flux data as discussed in the following section.

The predicted reaction fluxes for *M. thermoacetica*, generated using *i*AI558 during the autotrophic growth on H_2-CO_2 and heterotrophic growth on glucose, were analyzed and showed in Figures 6 and 7, respectively. During autotrophy, *M. thermoacetica* metabolizes H_2-CO_2 primarily using WLP, and gluconeogenesis supersedes the glycolytic process in the central metabolism (Figure 6); hence, fluxes through these reactions are higher (Figure 6) than the other central metabolic reactions. *M. thermoacetica* generates no net ATP through WLP; rather it requires ATP for gluconeogenesis (Figure 6) during the autotrophic growth. Thus, it generates ATP using only the ATP synthase (ATPS4r) reaction (Figure 6) through the chemiosmotic mechanism[48, 49], and the proton (H^+) gradient required for ATPS4r to work is generated by the translocation of one proton (H^+) by the ferredoxin hydrogenase (FRHD) reaction (Figure 6). The model, therefore, shows that *M. thermoacetica* conserves energy by electron transport phosphorylation (ETP) or anaerobic respiration during its autotrophic mode of metabolism. Since ETP is less efficient in ATP generation than substrate level phosphorylation (SLP)[50-52], *M. thermoacetica* generates less energy and grows slower than that observed during its heterotrophic growth.

On the other hand, glycolysis is the dominant process as indicated by higher fluxes through these reactions (Figure 7) during *M. thermoacetica*'s heterotrophic growth on glucose using *i*AI558. Interestingly, higher fluxes are also observed through the WLP reactions indicating their essentiality in *M. thermoacetica*'s metabolism even during heterotrophy. The degenerate TCA-

cycle cannot produce ATP; so as the ATP synthase reaction (ATPS4r) due to the lack of H^+ gradient in the absence of H_2. However, this deficiency is obviated by sufficient ATP production through SLP using the glycolytic reactions (Figure 7). Since SLP or glucose fermentation produces more ATP than ETP, the heterotrophic growth of *M. thermoacetica* is faster than its autotrophic growth. In fact, the higher efficiency of glycolytic reactions in ATP production during sugar fermentation by *M. thermoacetica* is further evident from much higher ATP yields on glucose, fructose, and xylose than on one carbon substrates (Figure 5B). These results again corroborate the notion that fermentation is more efficient than anaerobic respiration or ETP[50-52]. The sbml files of *i*AI558 (*i*AI558_co2_h2.xml and *i*AI558_glucose.xml) for simulating the described results are given in the supplementary data.

Insights into the energy conservation mechanism during autotrophy in *Moorella thermoacetica* using *i*AI558

The energy conservation mechanism of *M. thermoacetica* during its autotrophic growth on one carbon compounds is not well-understood yet. *M. thermoacetica* employs WLP during autotrophy, but the pathway generates no net ATP as discussed before. Thus, it is conceivable that the organism requires an electron transport chain (ETC) to generate ATP through ETP. However, the structure of ETC and the corresponding energy conservation mechanism are still debatable. Although the *M. thermoacetica* genome encodes cytochrome and menaquinone[28, 53, 54], apparent thermodynamic limitations[17] of these membrane-bound electron carriers make them unlikely to be a part of the potential *M. thermoacetica* ETC. The genome also encodes genes for a putative energy-converting hydrogenase (Ech) complex[28], which likely can also function as a putative formate-hydrogen lyase complex[16, 17] due to the proximity to a formate dehydrogenase

gene[28]. Based on these gene-annotations, two energy conservation mechanisms have been proposed[17, 20] so far, and we used *i*AI558 to analyze their feasibility for ATP generation in the model (Figure 8).

The first mechanism, proposed by Mock et al[11], included a formate-hydrogen lyase (FHL) and a methylene-tetrahydrofolate reductase (MTHFR) reaction (Figure 8A), where the coupling of a membrane-integral FHL with an electron-bifurcating MTHFR helps translocate two protons out of the cell membrane to generate proton gradient for ETP. The two electron-bifurcating transhydrogenase reactions — electron bifurcating ferredoxin: NAD hydrogenase (HYDFDNr) and electron bifurcating ferredoxin: NADP oxidoreductase (FRNDPRr) — also play essential roles in autotrophic energy conservation as H_2 is oxidized by HYDFDNr in *M. thermoacetica*[16, 18]. To investigate how this mechanism generates ATP in *i*AI558 on H_2 and CO_2 as growth substrates, we conducted model simulations by allowing only CO_2 and H_2 exchange fluxes to vary. The results showed no change in ATP flux with the change in CO_2 and H_2 exchange fluxes (Figure 8A).

In a recent review, Schuchmann and Muller[20] discussed the second hypothetical mechanism of the *M. thermoacetica* ETC. According to this mechanism, the ferredoxin hydrogenase (FRHD) reaction (Figure 8B) catalyzed by the Ech complex translocates four protons and couples with electron-bifurcating MTHFR, in which an oxidized ferredoxin is assumed to be its second electron acceptor[20] (Figure 8B). This mechanism also proposed different reaction stoichiometries for the HYDFDNr reaction (Figure 8B) than that used in the first mechanism. Simulation of ATP production using this mechanism in *i*AI558 on CO_2 and H_2 substrates also illustrated that ATP

fluxes are independent of the change in availability of CO_2 and H_2 in the medium. Thus, both the proposed mechanisms seemed to be incompatible with *i*AI558 to simulate ATP production and energy conservation during *M. thermoacetica*'s autotrophic growth.

However, the second mechanism of ETC worked for *i*AI558 only when stoichiometries of the FRHD and HYDFDNr reactions were changed to be different from the previous case (Figure 8C). The ratio of translocated H^+/ produced H_2 of the Ech complex remains the same as proposed originally[55], but the FRHD reaction now translocates one proton instead of four, and the HYDFDNr reaction uses the same stoichiometries as mentioned in the first mechanism. The model simulations showed a linear increase in ATP production with the increased supply of CO_2 and H_2 in the medium (Figure 8C). These results, therefore, suggest that the two previously proposed mechanisms actually make the *M. thermoacetica* ETC so efficient that the model can generate sufficient ATP without the requirement of any growth substrate due to creating thermodynamically infeasible cycles (see Tables 1 and 2 in the supplementary information 2). This scenario changes only when the reaction stoichiometries for the second mechanism were changed to be compatible with *i*AI558. Thus, a genome-scale constraint-based model is quite useful to obtain better insights into the disputable ETC mechanisms of the well-studied *M. thermoacetica*.

Conclusions

Construction of the genome-wide metabolic model for *M. thermoacetica* helped obtain a systems-level understanding of its metabolism. The predictive capability of the model was tested by comparing the organism's *in silico* autotrophic and heterotrophic growth rates with the

experimental growth data. In spite of significant uncertainties and debate, model-based analysis of the proposed ETC mechanisms suggests a feasible energy conservation process for *M. thermoacetica* during autotrophy. Although no strain-design strategies for metabolic engineering of *M. thermoacetica* were discussed and analyzed here, the developed model can now be leveraged to materialize such initiatives. In addition, the model predictions can be used to explore new research opportunities. For instance, the degenerate nature of the TCA-cycle, as well as the suggested stoichiometries for the possible *M. thermoacetica* ETC reactions can be tested using the ^{13}C-labelled flux analysis technique. Also, further experimental information on *M. thermoacetica* growth on other substrates apart from those discussed in this paper can help extend, modify, and validate *i*AI558. In addition to generating experimentally testable hypotheses regarding its physiology, such a detailed model will be instrumental in expediting the design of useful metabolic engineering strains of *M. thermoacetica* suitable for industrial applications.

Acknowledgements

The authors would like to thank Harish Nagarajan of the Department of Bioengineering, University of California San Diego, and Benjamin M. Woolston of the Department of Chemical Engineering, Massachusetts Institute of Technology (MIT) for insightful discussion. Support by the MIT Energy initiative and the DOE Grant (#DE-FOA-0001060) is gratefully acknowledged.

References

1. H. L. Drake, in *Acetogenesis*, ed. H. L. Drake, Chapman & Hall, 1994.
2. H. L. Drake, A. S. Gossner and S. L. Daniel, *Ann. N. Y. Acad. Sci.*, 2008, 1125, 100-128.
3. H. L. Drake and S. L. Daniel, *Res. Microbiol.*, 2004, 155, 869-883.
4. G. Fuchs, *Annu. Rev. Microbiol.*, 2011, 65, 631-658.
5. A. G. Fast and E. T. Papoutsakis, *Current Opinion in Chemical Engineering*, 2012, 1, 380-395.
6. F. E. Fontaine, W. H. Peterson, E. McCoy, M. J. Johnson and G. J. Ritter, *J. Bacteriol.*, 1942, 43, 701-715.
7. G. Diekert and G. Wohlfarth, *Antonie Van Leeuwenhoek*, 1994, 66, 209-221.
8. M. D. Collins, P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe and J. A. Farrow, *Int. J. Syst. Bacteriol.*, 1994, 44, 812-826.
9. H. A. Barker and M. D. Kamen, *Proc Natl Acad Sci U S A*, 1945, 31, 219-225.
10. S. W. Ragsdale, *Ann. N. Y. Acad. Sci.*, 2008, 1125, 129-136.
11. S. W. Ragsdale and E. Pierce, *Biochim. Biophys. Acta*, 2008, 1784, 1873-1898.
12. L.-F. Li, L. Ljungdahl and H. G. Wood, *J. Bacteriol.*, 1966, 92, 405-412.
13. I. Yamamoto, T. Saiki, S. M. Liu and L. G. Ljungdahl, *J. Biol. Chem.*, 1983, 258, 1826-1832.
14. L. G. Ljungdahl, *Annu. Rev. Microbiol.*, 2009, 63, 1-25.
15. S. L. Daniel, T. Hsu, S. I. Dean and H. L. Drake, *J. Bacteriol.*, 1990, 172, 4464-4471.
16. H. Huang, S. Wang, J. Moll and R. K. Thauer, *J. Bacteriol.*, 2012, 194, 3689-3699.
17. J. Mock, S. Wang, H. Huang, J. Kahnt and R. K. Thauer, *J. Bacteriol.*, 2014, 196, 3303-3314.
18. S. Wang, H. Huang, J. Kahnt and R. K. Thauer, *J. Bacteriol.*, 2013, 195, 1267-1275.
19. V. Müller, *Appl. Environ. Microbiol.*, 2003, 69, 6345-6353.
20. K. Schuchmann and V. Muller, *Nat Rev Micro*, 2014, 12, 809-821.
21. N. E. Lewis, H. Nagarajan and B. O. Palsson, *Nat Rev Micro*, 2012, 10, 291-305.
22. A. M. Feist, M. J. Herrgard, I. Thiele, J. L. Reed and B. O. Palsson, *Nat. Rev. Microbiol.*, 2009, 7, 129-143.
23. I. Thiele and B. O. Palsson, *Nat. Protocols*, 2010, 5, 93-121.
24. N. D. Price, J. L. Reed and B. O. Palsson, *Nat. Rev. Microbiol.*, 2004, 2, 886-897.
25. J. D. Orth, I. Thiele and B. O. Palsson, *Nat Biotech*, 2010, 28, 245-248.
26. A. Varma and B. O. Palsson, *Nat. Biotechnol.*, 1994, 12, 994-998.
27. H. Nagarajan, M. Sahin, J. Nogales, H. Latif, D. Lovley, A. Ebrahim and K. Zengler, *Microbial Cell Factories*, 2013, 12, 118.
28. E. Pierce, G. Xie, R. Barabote, E. Saunders, C. Han, J. Detter, P. Richardson, T. Brettin, A. Das, L. Ljungdahl and S. Ragsdale, *Environ. Microbiol.*, 2008, 10, 2550-2573.
29. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke and O. Zagnitko, *BMC Genomics*, 2008, 9, 75.
30. C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay and R. L. Stevens, *Nat. Biotechnol.*, 2010, 28, 977-982.

31. R. Overbeek, T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang and M. Cohoon, *et al*, *Nucleic Acids Res.*, 2005, 33, 5691-5702.
32. Y. K. Oh, B. Palsson, S. Park, C. Schilling and R. Mahadevan, *J. Biol. Chem.*, 2007, 282, 28791-28799.
33. K. Yamamoto, R. Murakami and Y. Takamura, *FEMS (Fed Eur Microbiol Soc) Microbiol Lett*, 1998, 161, 351-358.
34. J. Schellenberger, J. Park, T. Conrad and B. Palsson, *BMC Bioinformatics*, 2010, 11, 213.
35. R. S. Senger and E. T. Papoutsakis, *Biotechnol. Bioeng.*, 2008, 101, 1036-1052.
36. F. Salimi, K. Zhuang and R. Mahadevan, *Biotechnology Journal*, 2010, 5, 726-738.
37. S. Roberts, C. Gowen, J. P. Brooks and S. Fong, *BMC Systems Biology*, 2010, 4, 31.
38. A. Poehlein, S. Schmidt, A.-K. Kaster, M. Goenrich, J. Vollmers, A. Thürmer, J. Bertsch, K. Schuchmann, B. Voigt, M. Hecker, R. Daniel, R. K. Thauer, G. Gottschalk and V. Müller, *PLoS ONE*, 2012, 7, e33439.
39. M. Köpke, C. Held, S. Hujer, H. Liesegang, A. Wiezer, A. Wollherr, A. Ehrenreich, W. Liebl, G. Gottschalk and P. Dürre, *Proceedings of the National Academy of Sciences*, 2010, 107, 13087-13092.
40. J. Nölling, G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, G. T. C. S. C. Production, Finishing, G. T. C. W. M. Bioinformatics Teams Gtc Sequencing Center, U. S. U. o. t. H. S. Department of Pathology, T. N. I. o. H. B. M. The National Center for Biotechnology Information, D. d. G. B. T. c. F. Insa, B. Department of, R. U. H. T. Cell Biology, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin and D. R. Smith, *J. Bacteriol.*, 2001, 183, 4823-4838.
41. V. M. Markowitz, I. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. Ivanova and N. Kyrpides, *Nucleic Acids Res.*, 2012, 40, D115-122.
42. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res*, 1997, 25, 3389-3402.
43. M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg and T. O. Yeates, *Proc Natl Acad Sci U S A*, 1999, 96, 4285-4288.
44. J. Schellenberger, R. Que, R. Fleming, I. Thiele, J. Orth, A. Feist, D. Zielinski, A. Bordbar, N. Lewis, S. Rahmanian, J. Kang, D. Hyduke and B. Palsson, *Nature Protocols*, 2011, 6, 1290-1307.
45. J. L. Reed, T. Vo, C. Schilling and B. Palsson, *Genome Biol*, 2003, 4, R54.
46. E. Stupperich, H.-J. Eisinger and S. Schurr, *FEMS (Fed Eur Microbiol Soc) Microbiol Lett*, 1990, 87, 355-360.
47. V. Müller, F. Imkamp, E. Biegel, S. Schmidt and S. Dilling, *Ann. N. Y. Acad. Sci.*, 2008, 1125, 137-146.
48. P. Mitchell, *Nature*, 1961, 191, 144-148.
49. P. Mitchell, *J. Bioenerg.*, 1972, 3, 5-24.
50. A. Kröger, S. Biel, J. Simon, R. Gross, G. Unden and C. Lancaster, *Biochim. Biophys. Acta*, 2002, 15553, 23-38.
51. A. Kröger, V. Geisler, E. Lemma, F. Theis and R. Lenger, *Arch. Microbiol.*, 1992, 158, 311-314.
52. M. T. Madigan, J. M. Martinko and J. Parker, *Brock biology of microorganisms*, Benjamin Cummings; 13th Edition, 2010.

53. M. Gottwald, J. R. Andreesen, J. LeGall and L. G. Ljungdahl, *J. Bacteriol.*, 1975, 122, 325-328.
54. A. Das, J. Hugenholtz, H. Van Halbeek and L. G. Ljungdahl, *J. Bacteriol.*, 1989, 171, 5823-5829.
55. R. Hedderich, *J. Bioenerg. Biomembr.*, 2004, 36, 65-75.
56. M. I. Krzywinski, J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra, *Genome Res.*, 2009, DOI: 10.1101/gr.092759.109.

Figure legends

Figure 1. *M. thermoacetica* metabolic model (*i*AI558) development procedure. The model development steps can be broadly divided into two categories: model generation and model curation. The steps included in model generation are automatic, while both automatic and manual steps are involved in model curation. See text for details.

Figure 2. *M. thermoacetica* metabolic model (*i*AI558) subsystems. The reactions of *i*AI558 are categorized into 9 model subsystems based on the metabolic pathways in which the reactions are involved. Each colored ribbon indicates the relationship between the number of reactions and a model subsystem in this circos[56] figure. The "cofactor metabolism" subsystem contains the highest number of model reactions followed by the "amino acid metabolism" and "lipid metabolism" subsystems.

Figure 3. Comparative analysis of the Wood-Ljungdahl pathway (WLP) genes of *M. thermoacetica*. The sequences of 46 *M. thermoacetica* genes involved in WLP and autotrophic metabolism were compared to 2 acetogens and 3 related clostridial genomes using BLAST[42] and phylogenetic profiling[43]. The *M. thermoacetica* genes are shown on the right side of the circle, and the genomes are placed on the left. The identified ortholog of an *M. thermoacetica* gene in a genome is represented by linking that gene to the corresponding genome by a colored ribbon in this circos[56] figure.

Figure 4. Central metabolic pathways of *M. thermoacetica* and *C. ljungdahlii*. The differences in the central carbon metabolism of *M. thermoacetica* and *C. ljungdahlii* are identified only in the degenerate TCA-cycle. The reactions and metabolites are denoted by the upper case and lower case letters, respectively. Abbreviations: r5p, ribose-5-phosphate; RPI, ribose-5-phosphate isomerase; ru5p-D, ribulose-5-phosphate; RPE, ribulose-5-phosphate-3-epimerase; xu5p-D, xylulose-5-phosphate; TKT1, transketolase; s7p, sedoheptulose-7-phosphate; g3p, glyceraldehyde-3-phosphate; TALA, transaldolase; f6p, fructose-6-phosphate; TKT2, transketolase; e4p, erythrose-4-phosphate; glc-D, glucose; atp, ATP; adp, ADP; HEX1, hexokinase; g6p, glucose-6-phosphate; PGI, glucose-6-phosphate isomerase; PFK, phosphofructokinase; fdp, fructose-1-6-bisphosphate; FBA, fructose-bisphosphate aldolase; dhap, dihydroxy-acetone-phosphate; TPI, triose-phosphate isomerase; g3p, glyceraldehyde-3-phosphate; GAPD, glyceraldehyde-3-phosphate dehydrogenase; nad, NAD; nadh, NADH; 13dpg, 3-phospho-D-glyceroyl-phosphate; PGK, phosphoglycerate kinase; 3pg, 3-Phospho-D-glycerate; PGM, phosphoglycerate mutase; 2pg, D-glycerate-2-phosphate; ENO, enolase; pep, phosphoenolpyruvate; PYK, pyruvate kinase; pyr, pyruvate; PC, pyruvate carboxylase; oaa,

oxaloacetate; CS, citrate synthase; cit, citrate; ACONTa, aconitase/citrate hydro-lyase; acon, aconitase; ACONTb, aconitase/isocitrate hydro-lyase; icit, isocitrate; ICDH, isocitrate dehydrogenase; akg, 2-oxoglutarate; OOR, 2-oxoglutarate synthase; fdxox, oxidized ferredoxin; fdxrd, reduced ferredoxin; succoa, succinyl-CoA; SUCOAS, succinyl-CoA synthetase; succ, succinate; FRD, fumarate reductase; fum, fumarate; FUM, fumarase; mal-L, L-malate; and MDH, malate dehydrogenase.

Figure 5. *M. thermoacetica* growth and ATP simulation with *i*AI558. *M thermoacetica in silico* growth rates and yields (A), and ATP production rates and yields (B) were simulated with *i*AI558 using the COBRA toolbox[44]. The simulated growth rates were compared with the available experimental growth data[15] shown as an inset of (A).

Figure 6. *M. thermoacetica* autotrophic growth simulation. The simulated reaction-fluxes during the autotrophic growth of *M. thermoacetica* are visualized by heat maps on an abridged version of its metabolic network. The reactions and metabolites are denoted by the upper case and lower case letters, respectively. Abbreviations: r5p, ribose-5-phosphate; RPI, ribose-5-phosphate isomerase; ru5p-D, ribulose-5-phosphate; RPE, ribulose-5-phosphate-3-epimerase; xu5p-D, xylulose-5-phosphate; TKT1, transketolase; s7p, sedoheptulose-7-phosphate; g3p, glyceraldehyde-3-phosphate; TALA, transaldolase; f6p, fructose-6-phosphate; TKT2, transketolase; e4p, erythrose-4-phosphate; glc-D, glucose; EX_glc-D(e), glucose exchange; atp, ATP; adp, ADP; HEX1, hexokinase; g6p, glucose-6-phosphate; PGI, glucose-6-phosphate isomerase; FBP, fructose-1,6-bisphosphatase; fdp, fructose-1,6-bisphosphate; FBA, fructose-bisphosphate aldolase; dhap, dihydroxy-acetone-phosphate; TPI, triose-phosphate isomerase; g3p, glyceraldehyde-3-phosphate; GAPD, glyceraldehyde-3-phosphate dehydrogenase; nad, NAD; nadh, NADH; 13dpg, 3-phospho-D-glyceroyl-phosphate; PGK, phosphoglycerate kinase; 3pg, 3-Phospho-D-glycerate; PGM, phosphoglycerate mutase; 2pg, D-glycerate-2-phosphate; ENO, enolase; pep, phosphoenolpyruvate; PPCK, phosphoenolpyruvate carboxykinase; pyr, pyruvate; PC, pyruvate carboxylase; oaa, oxaloacetate; CS, citrate synthase; cit, citrate; ACONTa, aconitase/citrate hydro-lyase; acon, aconitase; ACONTb, aconitase/isocitrate hydro-lyase; icit, isocitrate; ICDH, isocitrate dehydrogenase; akg, 2-oxoglutarate; OOR, 2-oxoglutarate synthase; fdxox, oxidized ferredoxin; fdxrd, reduced ferredoxin; succoa, succinyl-CoA; SUCOAS, succinyl-CoA synthetase; succ, succinate; FRD, fumarate reductase; fum, fumarate; FUM, fumarase; mal-L, L-malate; MDH, malate dehydrogenase; co2, carbon dioxide; EX_co2(e), carbon dioxide exchange; for, formate; FDH, formate dehydrogenase; thf, tetrahydrofolate; FTHFL, formate-tetrahydrofolate ligase; 10fthf, 10-formyltetrahydrofolate; MTHFC, 5,10-methenyltetrahydrofolate cyclohydrolase; methf, 5,10-methenyltetrahydrofolate; MTHFD, 5,10-methylenetetrahydrofolate dehydrogenase; mlthf, 5,10-methylenetetrahydrofolate; MTHFR, 5,10-methylenetetrahydrofolate reductase; 5mthf, 5-methyltetrahydrofolate; MTRCFSP, 5-methyltetrahydrofolate corrinoid iron-sulfur protein methyltransferase; mcfesp, methylcorrinoid iron-sulfur protein; CODH/ACS, carbon monoxide dehydrogenase/acetyl-CoA synthase complex; coa, coenzyme-A; cfesp, corrinoid iron-sulfur protein; accoa, acetyl-CoA; co, carbon monoxide; PFOR, pyruvate-ferredoxin oxidoreductase; PTAr, phosphotransacetylase; actp, acetyl phosphate; ACKr, acetate kinase; ac, acetate; EX_h2(e), hydrogen exchange; FRNDPRr, ferredoxin: NADP oxidoreductase; h2, hydrogen; HYDFDNr, ferredoxin: NAD hydrogenase; FRHD, ferredoxin hydrogenase; and ATPS4r, ATP synthase.

Figure 7. *M. thermoacetica* heterotrophic growth simulation. The simulated reaction-fluxes during the heterotrophic growth of *M. thermoacetica* are visualized by heat maps on an abridged version of its metabolic network. The reactions and metabolites are denoted by the upper case and lower case letters, respectively. Abbreviations: r5p, ribose-5-phosphate; RPI, ribose-5-phosphate isomerase; ru5p-D, ribulose-5-phosphate; RPE, ribulose-5-phosphate-3-epimerase; xu5p-D, xylulose-5-phosphate; TKT1, transketolase; s7p, sedoheptulose-7-phosphate; g3p, glyceraldehyde-3-phosphate; TALA, transaldolase; f6p, fructose-6-phosphate; TKT2, transketolase; e4p, erythrose-4-phosphate; glc-D, glucose; atp, ATP; adp, ADP; HEX1, hexokinase; g6p, glucose-6-phosphate; PGI, glucose-6-phosphate isomerase; PFK, phosphofructokinase; fdp, fructose-1-6-bisphosphate; FBA, fructose-bisphosphate aldolase; dhap, dihydroxy-acetone-phosphate; TPI, triose-phosphate isomerase; g3p, glyceraldehyde-3-phosphate; GAPD, glyceraldehyde-3-phosphate dehydrogenase; nad, NAD; nadh, NADH; 13dpg, 3-phospho-D-glyceroyl-phosphate; PGK, phosphoglycerate kinase; 3pg, 3-Phospho-D-glycerate; PGM, phosphoglycerate mutase; 2pg, D-glycerate-2-phosphate; ENO, enolase; pep, phosphoenolpyruvate; PYK, pyruvate kinase; pyr, pyruvate; PC, pyruvate carboxylase; oaa, oxaloacetate; CS, citrate synthase; cit, citrate; ACONTa, aconitase/citrate hydro-lyase; acon, aconitase; ACONTb, aconitase/isocitrate hydro-lyase; icit, isocitrate; ICDH, isocitrate dehydrogenase; akg, 2-oxoglutarate; OOR, 2-oxoglutarate synthase; fdxox, oxidized ferredoxin; fdxrd, reduced ferredoxin; succoa, succinyl-CoA; SUCOAS, succinyl-CoA synthetase; succ, succinate; FRD, fumarate reductase; fum, fumarate; FUM, fumarase; mal-L, L-malate; MDH, malate dehydrogenase; co2, carbon dioxide; EX_co2(e), carbon dioxide exchange; for, formate; FDH, formate dehydrogenase; thf, tetrahydrofolate; FTHFL, formate-tetrahydrofolate ligase; 10fthf, 10-formyltetrahydrofolate; MTHFC, 5,10-methenyltetrahydrofolate cyclohydrolase; methf, 5,10-methenyltetrahydrofolate; MTHFD, 5,10-methylenetetrahydrofolate dehydrogenase; mlthf, 5,10-methylenetetrahydrofolate; MTHFR, 5,10-methylenetetrahydrofolate reductase; 5mthf, 5-methyltetrahydrofolate; MTRCFSP, 5-methyltetrahydrofolate corrinoid iron-sulfur protein methyltransferase; mcfesp, methylcorrinoid iron-sulfur protein; CODH/ACS, carbon monoxide dehydrogenase/acetyl-CoA synthase complex; coa, coenzyme-A; cfesp, corrinoid iron-sulfur protein; accoa, acetyl-CoA; co, carbon monoxide; PFOR, pyruvate-ferredoxin oxidoreductase; PTAr, phosphotransacetylase; actp, acetyl phosphate; ACKr, acetate kinase; ac, acetate; EX_h2(e), hydrogen exchange; FRNDPRr, ferredoxin: NADP oxidoreductase; h2, hydrogen; HYDFDNr, ferredoxin: NAD hydrogenase; FRHD, ferredoxin hydrogenase; and ATPS4r, ATP synthase.

Figure 8. Analysis of the autotrophic energy conservation mechanism of *M. thermoacetica*. Implementation of the autotrophic energy conservation mechanism in *i*AI558 proposed by Mock et al[17] and Schuchmann and Muller[20] is shown in (A) and (B), respectively, while (C) shows the mechanism that is actually implemented in the model. The reactions and metabolites are denoted by the upper case and lower case letters, respectively. Abbreviations: FHL, formate hydrogen lyase; for, formate; co2, carbon dioxide; h2, hydrogen; mlthf, 5,10-methylenetetrahydrofolate; MTHFR, 5,10-methylenetetrahydrofolate reductase; 5mthf, 5-methyltetrahydrofolate; nad, NAD; nadh, NADH; nadp, NADP; nadph, NADPH; FRNDPRr, ferredoxin: NADP oxidoreductase; fdxox, oxidized ferredoxin; fdxrd, reduced ferredoxin; HYDFDNr, ferredoxin: NAD hydrogenase; and FRHD, ferredoxin hydrogenase.

Table 1. Composition of the *In Silico* Minimal Medium of *Moorella thermoacetica*

Reaction Name	Abbreviation	Reaction
Glucose exchange	EX_glc-D(e)	glc-D[e] <=>
Fructose exchange	EX_fru(e)	fru[e] <=>
Xylose exchange	EX_xyl-D(e)	xyl-D[e] <=>
Carbon dioxide exchange	EX_co2(e)	co2[e] <=>
Carbon monoxide exchange	EX_co(e)	co[e] <=>
Hydrogen exchange	EX_h2(e)	h2[e] <=>
Methanol exchange	EX_meoh(e)	meoh[e] <=>
Nicotinic acid exchange	EX_nac(e)	nac[e] <=>
Calcium exchange	EX_ca2(e)	ca2[e] <=>
Cobalt exchange	EX_cobalt2(e)	cobalt2[e] <=>
Iron (III) exchange	EX_fe3(e)	fe3[e] <=>
Water exchange	EX_h2o(e)	h2o[e] <=>
Hydrogen sulfide exchange	EX_h2s(e)	h2s[e] <=>
Potassium exchange	EX_k(e)	k[e] <=>
Magnesium exchange	EX_mg2(e)	mg2[e] <=>
Ammonium exchange	EX_nh4(e)	nh4[e] <=>
Inorganic phosphate exchange	EX_pi(e)	pi[e] <=>

Supplementary information 1.pdf: This pdf file includes supplementary figures and tables, including the details of *M. thermoacetica* biomass composition, and NGAM and GAM parameters calculation.

Supplementary information 2.xlsx: This excel file contains 6 tables describing all genes, proteins, reactions, and metabolites, as well as the simulation results and constraints of the model.

iAI558_co2_h2.xml: This sbml file of *M. thermoacetica* model can be used to simulate *M. thermoacetica* autotrophic growth.

iAI558_glucose.xml: This sbml file of *M. thermoacetica* model can be used to simulate *M. thermoacetica* heterotrophic growth.