Exploring Hydrogenotrophic Methanogenesis: A Genome Scale Metabolic Reconstruction of *Methanococcus maripaludis S2*

**Matthew A. Richards1,2, Thomas J. Lie3, Juan Zhang4, John A. Leigh3\*, Nathan D. Price2\***

1Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL USA

2Institute for Systems Biology, Seattle, WA, USA

3Department of Microbiology, University of Washington, Seattle, WA, USA

4Jiangnan University, China

\*Corresponding authors: [nprice@systemsbiology.org](mailto:nprice@systemsbiology.org), [leighj@u.washington.edu](mailto:leighj@u.washington.edu)

# Abstract

Hydrogenotrophic methanogenesis occurs in multiple environments ranging from the intestinal tracts of animals to anaerobic sediments and hot springs. Energy conservation in hydrogenotrophic methanogens was long a mystery; only recently, it was reported that net energy conservation for growth depends on electron bifurcation. In this work we focus on *Methanococcus maripaludis*, a well-studied hydrogenotrophic marine methanogen. To better understand hydrogenotrophic methanogenesis and compare it with methyltrophic methanogenesis that utilizes oxidative phosphorylation rather than electron bifurcation, we have built iMR540, a genome scale metabolic reconstruction that accounts for 540 of the 1722 protein-coding genes of *M. maripaludis* strain S2. Our reconstructed network uses recent literature to not only portray the central electron bifurcation reaction, but also incorporate vital biosynthesis and assimilation pathways, including unique cofactor and coenzyme syntheses. We show that our model accurately predicts experimental growth and gene knockout data. Furthermore, we use our reconstruction to probe the implications of electron bifurcation by showing its essentiality and investigating the effects of ferredoxin specificity on the network. Additionally, we demonstrate a novel method of applying thermodynamic constraints to a metabolic model to quickly estimate overall organism free energy. Finally, we describe a reconstruction-specific computational toolbox we created to improve usability. Together, our results provide a computational network for exploring hydrogenotrophic methanogenesis and confirm the importance of electron bifurcation in this process.

# Importance

Understanding and applying hydrogenotrophic methanogenesis could be key for developing new bio-energy technologies around methane gas. Although a significant portion of biological methane is generated through this pathway, existing methanogen models portray more traditional energy conservation mechanisms found in other methanogens. We have constructed a genome-scale metabolic network of *Methanococcus maripaludis* that explicitly shows this environmentally ubiquitous process, including energy conservation through electron bifurcation. Our reconstruction demonstrates the importance of electron bifurcation in central metabolism, providing both a window into hydrogenotrophic methanogenesis and a hypothesis-generating platform to fuel metabolic engineering efforts.

# Introduction

Methane plays a critical role in the global carbon cycle and as a greenhouse gas, is 21 times more potent than carbon dioxide (1) in absorbing and emitting energy. Additionally, it is a candidate bridge fuel (2) because it burns comparatively cleaner than traditional fossil fuels. Advancing technology also enables this gas to be converted to high energy density liquid fuels with a lower carbon footprint (3). Methane is the second most abundant greenhouse gas after carbon dioxide (4) and is produced in the environment by biological and non-biological sources (5). Methanogens are the largest biological contributors of methane, producing about 1 Gt of methane gas per year (6). This group of microorganisms from the domain Archaea grow on carbon dioxide or one or two carbon compounds using enzymes containing unique biological co-factors (7, 8).

Though phylogenetically and metabolically diverse, methanogens can be separated into two groups based on the presence or absence of cytochromes (6). The cytochrome-lacking methanogens (sometimes referred to as hydrogenotrophic methanogens) mainly use H2, and sometimes formate, as sources of electrons for CO2 reduction to methane. In contrast, cytochrome-containing (or methylotrophic) methanogens utilize acetate and methylated compounds for methanogenic growth with a minority also being able to use H2 and CO2. Although both groups have similar central pathways of CO2 reduction, they possess differing modes of energy coupling (9) at the last methanogenic step involving heterodisulfide reductase (Hdr).

The reduction of the CoM-S-S-CoB heterodisulfide with H2 or reduced electron carriers is exergonic and can be directly or indirectly coupled to energy generation. In the methylotrophic methanogens, a membrane-associated cytochrome-containing Hdr (HdrDE) receives reducing equivalents from a methanogen-specific membrane-soluble electron shuttle, methanophenazine, for reduction of the heterodisulfide. This results in proton extrusion and the creation of a membrane potential for ATP generation (10, 11). However, in the hydrogenotrophic methanogens, the Hdr (HdrABC) is cytoplasmic and no membrane potential is generated. Instead, Hdr mediates a bifurcation of electron flow in which the exergonic heterodisulfide reduction is coupled to and drives the endergonic reduction of a ferredoxin used for the first step of methanogenesis (12).

*Methanococcus maripaludis* (13) belongs to this group of hydrogenotrophic methanogens. Compared to the larger genomes of methylotrophic methanogens, its genome is relatively small and contains only 1722 protein coding genes (14). It grows robustly with a doubling time of 2 hours (13) and is genetically tractable (15), and thus has been an ideal candidate for studying methanogenesis, unique co-factors and their biosyntheses (16), and gene regulation (17).To avoid environmental fluctuations that can affect gene regulation, a system for continuous culture of *M. maripaludis* (18) has been established for steady state transcriptomic (19) and proteomic (20) studies of *M. maripaludis* strains. Several groups have also employed larger systems biology approaches to perform predictive studies using this organism (21). With these tools in place, and the ability for expression of heterologous genes in *M. maripaludis* (22, 23), the metabolic engineering of *M. maripaludis* for various industrial use is the obvious next step.

Genome scale metabolic reconstructions are powerful tools that map and elucidate metabolic pathways. They are organism-specific knowledge bases that can be used for simulating steady state growth via flux balance analysis (FBA) (24) by generating constraint-based models. Using these models, we can hypothesize different metabolic scenarios that can then be tested experimentally. They have helped guide metabolic engineering efforts to produce industrial biochemicals in multiple organisms (25, 26). Similarly, a genome scale metabolic reconstruction for *M. maripaludis* would not only promote a better understanding of methanogenesis but also support metabolic engineering efforts that could harness the unique metabolism of this hydrogenotrophic methanogen. Other groups have already created metabolic models of *M. maripaludis*; as part of a mutualistic community model with *D. vulgaris* (27)and under axenic conditions (28). In the former case, the model of *M. maripaludis* represented only core metabolism and was used primarily to investigate interactions between the two different species rather than probe the depths of the organism’s metabolism (27). The latter case was the first genome-scale metabolic reconstruction of *M. maripaludis* (28), an important step towards understanding *M. maripaludis* metabolism.

In our model, iMR540, we made important updates and refinements to various pathwaysbased on recent literature*.* The most critical was the electron bifurcation step that has been described above as it explains the ability for this organism to grow despite the lack of a proton-exporting electron transport chain. This also includes eliminating methanophenazine utilization and synthesis, which is part of the membrane bound electron transport system of the methylotrophic methanogens and is absent in hydrogenotrophic methanogens (6). Additional features include a corrected sulfur assimilation pathway (29), and the addition of various biosynthesis pathways for all of the unique coenzymes involved in methanogenesis (30). We increased genome coverage by employing likelihood-based gap filling, a technique that fills reaction gaps based on gene homology rather than on parsimony (31). Our reconstruction is the first manually-curated genome scale reconstruction to employ likelihood based gap filling. Furthermore, we expanded the scope of our reconstruction beyond stoichiometric considerations by creating a new method to approximate overall model free energy. This is an especially salient consideration for methanogenic archaea, which can grow close to the thermodynamic limits that support life (32). A well-established method of applying free energy constraints involves applying the second law of thermodynamics to metabolic models to restrict reaction directionalities in the direction of negative free energy generation (33, 34). Rather than apply thermodynamic constraints to every metabolic reaction, we created a method that predicts overall free energy generated during steady state growth based solely on standard free energies and effective concentrations of external metabolites. In combining these novel thermodynamic considerations with stoichiometric information, iMR540 provides a means to predict energetically feasible strain designs, enhancing our metabolic engineering capabilities with *M. maripaludis.*

# Methods

## Genome Scale Reconstruction Procedure

The process of genome scale metabolic network reconstruction has been reviewed previously (35) and begins with annotating an organism genome using gene-protein-reaction (GPR) relationships stored in a reaction database. Several databases are available for this purpose (36–38); we chose the Department of Energy Systems Biology Knowledgebase (Kbase; www.kbase.us), a suite of tools that includes the Model SEED reaction database (38). We created our first draft reconstruction using the stored Kbase genome for *M. maripaludis S2* (genome id: kb|g.575) and the automated reconstruction method (“Reconstruct Genome-scale Metabolic Model”). For this initial reconstruction, we used the default gram negative biomass composition and filled knowledge gaps using likelihood based gap filling (method currently not supported in Kbase Narrative Interface). This yielded the first full draft of the metabolic reconstruction that could predict growth when simulated as a model.

We expanded and refined the model by manually adding information from literature sources. In cases where reactions from literature were part of the Model SEED database, we labeled the reactions using SEED identifiers, names, subsystems, and EC numbers. For other cases where we encountered reactions that were not part of the Model SEED we created unique reaction identifiers and names, then added subsystem information based on our knowledge of the metabolic network. We also adhered to SEED identifiers, names, formulas, and charges for metabolites whenever possible and had very few cases where we specified our own values. Metabolites were compartmentalized using standard tags for cytosol (“c0”) and extracellular (“e0”) compartments. These tags additionally identify *M. maripaludis* as “Organism 0” in the possible future case where we could add other organisms to create a community metabolic reconstruction. Exchange reactions used for introducing metabolites to the extracellular compartment were standardized in “EX\_{metabolite ID}[e0]” format. Comprehensive information on the reactions, metabolites, and genes in our reconstruction can be found in Supplementary Materials.

## Model Simulations with Flux Balance Analysis

To make rigorous quantitative growth predictions, a genome scale metabolic reconstruction can be simulated as a model. Reactions and their participating metabolites in the metabolic network are connected via the stoichiometric matrix (S), which contains the stoichiometric coefficients for each metabolite (row) in each reaction (column). The S-matrix is used as the basis of a model via the principles of metabolite mass conservation by recognizing that time-dependent accumulation of metabolites in the system (b) is equivalent to the product of the S-matrix and the vector of reaction fluxes (v)

|  |  |
| --- | --- |
|  | [1] |

In flux balance analysis (FBA), we further simplify this differential system by assuming our organism is in steady state growth; thus b=0 and the system is linear (39). This assumption bounds our model system to a large solution space that can further be constrained by applying upper and lower bounds to each reaction flux:

|  |  |
| --- | --- |
|  | [2] |

To find feasible flux distributions that represent likely physiological states within this solution space, we solved our model by optimizing the biomass objective function, a simulation of maximum cell growth yield (40). We further constrained possible flux distributions by minimizing the squared sum of fluxes, effectively forcing our model to find solutions that minimize the total flux in the system while maximizing growth. All model simulations were performed using the COBRA toolbox 2.0 (41) in MATLAB [7.14.0.739] (The MathWorks Inc., Natick, MA).

To encourage model transparency (42) and assist future users in simulating condition-specific models, we created several functions that create these models, simulate maximum growth with the aforementioned constraints, and print relevant information from the flux distribution (Supplementary Materials). We also wrote numerous functions to help modify the reaction network, retrieve specific useful pieces of information from model simulations, and diagnose issues that may arise during model use. For several of these functions, we used the Paint4Net toolbox (43) to draw flux maps that show the direction and magnitude of fluxes in a given FBA solution. A limited number of our functions are included with this manuscript in their current versions (Supplementary Materials) with the full set of up-to-date tools available on Github (https://github.com/marichards/methanococcus).

## Gene Knockout Phenotype Simulations

Because a model is based around the stoichiometry of reactions contained in the S-matrix, knocking out a gene is akin to knocking out all reactions that depend on the gene. Thus, performing a gene knockout phenotype simulation in a metabolic model requires that model reactions be linked to genes via GPR relationships. We performed gene knockout simulations using our function “simulateKOPanel.m” (Supplementary Materials), which relies heavily on the “deleteModelGenes.m” function in the COBRA Toolbox 2.0 (41) as well as several of our own functions. Our experimental test set included 18 knockout genotypes across 4 different growth conditions, with 30 total wet lab experiments across these conditions (44–49). We simulated growth phenotypes for all 72 combinations of knockout genotypes and growth conditions and then evaluated these growth phenotypes as lethal/non-lethal with a threshold of 10% wild type growth. Predictive accuracy was assessed by comparing predictions on the 30 known phenotypes with wet lab data. We further evaluated our model’s performance using the Matthews correlation coefficient (MCC), a metric that evaluates correlation based on a -1 to 1 scale (50):

|  |  |
| --- | --- |
|  | [3] |

## Thermodynamic Calculations

We added standard free energies of formation (1 mM, 25ºC, 1 bar, pH=7, ionic strength = 0.1 M) from the Equilibrator database (51) to all exchange reactions for which these values could be reliably estimated via the group contribution method (52). To incorporate these values into our reconstruction, we expanded the standard model structure to include a “freeEnergy” numerical array with length equal that of the “reactions” array. For calculating overall free energy of a flux distribution, we created an “optimizeThermoModel.m” code (Supplementary Materials) that is built around the “optimizeCbModel.m” code in the COBRA Toolbox 2.0 (41). Our script accepts effective concentrations (mM) for specified exchange metabolites, assumes standard activities of 1 mM for unspecified metabolites, and uses these values to calculate effective metabolite free energies based on the reconstruction’s stored values for each exchange reaction. Prior to performing FBA, we add these free energies to the exchange reactions, which ordinarily have the form:

We alter these exchanges such that production of a metabolite “creates” free energy equivalent to the metabolite’s free energy of formation:

Here, is the stoichiometric coefficient of a new metabolite “dG” that is used to sum model free energy. Because exchange reactions must satisfy mass balance by necessarily entering or exiting the model without creating new metabolites, adding free energies to the model creates an imbalance that we must correct. We restore model balance by allowing “dG” to exit the model via its own exchange reaction (GIBBS\_kJ\_GDW):

Measuring the total flux of the exchange reaction gives an estimation of total free energy being generated in an FBA solution on a per cell mass basis. We have incorporated this thermodynamic calculation into all of our available model simulations (Supplementary Materials); thus by default, we calculate and print overall model free energy in every flux distribution.

## Dry Cell Weight and Growth Yield Measurements

Wild type *M. maripaludis S2* cells were grown in McNA medium—a chemically defined medium for growth on H2 and CO2 supplemented with acetate (Supplementary Materials)—using a 1-L chemostat under anaerobic conditions as described previously (18). The chemostat was operated in steady state continuous mode under H2-limiting conditions to match model simulation conditions, with gas flows of 10-20 mL/min H2, 40 mL/min CO2, 15 mL/min of a H2S:Ar mixture (1:99 v/v), and a balance of N2 up to a total 200 mL/min. We altered our growth rate of *M. maripaludis* during steady state by varying pump speeds to achieve dilution rates of approximately 0.045-0.090 h-1, checking OD660 periodically to ensure steady state at each data point. For each sample point, we measured growth rate based on dilution rate and methane evolution rate via a combination of a bubble flow meter to assess total gas outflow and a Buck Scientific model 910 gas chromatograph equipped with a flame ionization detector to quantify methane fraction.

We recalculated calibration curves for dry cell weight versus optical density by measuring dry cell weight via cell filtering and OD660 via a UV/Vis spectrophotometer {Spectronic 20D+} blanked with water. After measuring chemostat optical density, we sampled 50 mL aliquots of cells in suspension directly from chemostat culture and centrifuged samples at 7000 RPM for 15 minutes. 40 mL of supernatant was removed by pipette, then cells were re-suspended in the remaining 10 mL of media. These concentrated aliquots were vacuum filtered through 0.45 µM pore filters to remove all non-cellular components, then dried at room temperature and weighed daily until their weights stabilized.

Growth yields were calculated based on doubling time (td, equal to ln(2) x (dilution rate x 60)-1) as described previously (53), but with our measured conversion between OD660 and cell density:

## ATP Maintenance and Predicted Growth Yields

As described by Thiele and Palsson, the optimal way to obtain accurate ATP maintenance values is to plot ATP production versus growth data from chemostat growth experiments (35). In practice, this requires measuring steady state growth rate in concert with an uptake rate or, in our case, a product secretion rate, as described above.

To calculate ATP maintenance values in our model, we constrained our model to our measured growth rate and methane secretion rate at each sampling point and set the model objective to maximize ATP hydrolysis (rxn00062[c0]). We plotted each resulting value of ATP production as a function of growth rate and obtained the growth-associated (slope) and non-growth associated (y-intercept) ATP maintenance values using a linear model, as described by Thiele and Palsson (35). The resulting plot can be found in Supplementary Materials.

Our growth data points comprised a set of 9 measurements and we used them as both training and test data by employing leave one out cross validation (LOOCV). In the LOOCV approach, a set of N samples is divided into a training dataset of N-1 points and a test sample of 1 point. The model developed on the training set is then tested on the remaining point that was left out of the training data. In employing this method, we iteratively removed one point from our full dataset and determined ATP maintenance values for that N-1 dataset as described above to create a trained model. We then constrained our model’s methane secretion flux to the measured rate in the remaining test point and predicted maximum growth rate within that constraint using our trained model. Using these values, we calculated predicted growth yields for each point using the above formula and compared them to our measured values for each point. All simulations were performed using the default H2 + CO2 media formulation supplemented with acetate (McNA medium).

# Results

## Reconstruction Statistics

The basic statistics for iMR540 are displayed in Table 1. Notably, reactions are categorized as 1) internal reactions, occurring entirely within the cytoplasm; 2) transport reactions, involving translocation of at least one chemical species across the cell membrane; 3) exchange reactions, which supply metabolites to or remove metabolites from the model. Of the 586 internal reactions in our network, over 85% of the internal reactions are associated with at least one gene. We suspect that a major reason for this high percentage of gene-associated reactions was our use of likelihood based gap filling, which resulted in the automated addition of 66 genes to our reconstruction before manual curation. Furthermore, we relied heavily on biochemical knowledge from literature sources, particularly regarding recently-elucidated biosynthesis pathways that were not initially available in annotation databases. Our combined use of maximum likelihood gap filling and reliance on published literature sources are the likely causes for our consistent ties to gene homology.

Another salient detail of our reconstruction is that it includes many “dead-end” metabolites and reactions that cannot be synthesized or consumed. Thus, these metabolites and reactions are not part of our simulatable model, but we have included them in our reconstruction because they are all gene-associated; all dead end internal reactions in our reconstruction have at least one gene association. This indicates that there is genetic evidence supporting the presence of each dead end reaction and metabolite, thus they should be involved in metabolism even though we have not yet elucidated full synthesis or consumption pathways. They represent excellent candidates for further exploration of *M. maripaludis* metabolism, particularly as iMR540 is updated and expanded in the future.

Conversely, our reconstruction contains 86 internal reactions that lack genes, many of which were added during automated gap filling but some of which were added manually. All of our reactions are annotated with subsystems, allowing us to assess where each reaction fits into metabolism, including those without genes. Figure 1 shows the breakdown of reactions without genes, where the subsystems have been manually grouped into broader categories (e.g. “Amino Acid Biosynthesis” instead of “Glycine Biosynthesis”). The largest group of these reactions is the “Unique Coenzyme Syntheses”, which includes reactions that synthesize coenzyme M, coenzyme B, tetrahydromethanopterin (H4MPT), methanofuran, coenzyme F420, and coenzyme F430. Although these 24 reactions lack genes, all of them were added manually as hypothetical steps to complete essential biosynthetic pathways and are based on information from biochemical literature. These are distinct from, for example, the 11 reactions encompassed by “Vitamin and Cofactor Synthesis” that were added to fill biosynthesis gaps but have no supporting literature evidence. We expect that as experimental research groups uncover more biochemical phenomena, they will determine genes that tie to the reactions in the former group. The gap filling reactions, much like dead end reactions and metabolites, point us toward poorly-understood areas of metabolism in our organism and require more investigation into both the reaction pathways and their associated genes.

As an additional feature of our reconstruction, our use of likelihood based gap filling also assigned likelihood scores for many of the reactions in the reconstruction. These confidence scores quantify the probability of a given reaction being part of the metabolic reconstruction on a scale of 0-1 and provide a new metric of evaluating our confidence in the reconstruction. We can then use the scores to quickly hone in on both reactions that lack genes and gene-associated reactions with low gene homology as possible targets for more experimental investigation. They also provide a logical starting point for future users looking to expand upon and improve the existing reconstruction.

## Electron Bifurcation and Acetate Metabolism

Methanogenesis from H2 and CO2 has often been represented as a linear pathway with heterodisulfide reduction as the final step. This was demonstrated to be mediated by methanophenazine dependent membrane bound heterodisulfide (HdrDE) (10, 11) for the cytochrome containing methanogens. However, the non-cytochrome containing obligate hydrogenotrophs do not contain the typical membrane associated heterodisulfide reductase but instead one that is most likely associated with the cytoplasm (54, 55). Additionally, it is a three subunit complex (HdrABC) with a flavin adenine dinucleotide (FAD) containing co-factor in the HdrA subunit (56).HdrA and other FAD-containing enzymes have been increasingly recognized as sites for electron bifurcation, coupling an exergonic reaction with an endergonic reaction in a two-step transfer of one electron (57, 58). Recently, it has been demonstrated (12, 59) that this heterodisulfide reductase mediates the coupling of exergonic heterodisulfide reduction with endergonic ferredoxin reduction. As shown in Figure 2, this ferredoxin is used for reduction of the CO2 via Fwd thereby linking the last step of methanogenesis with the first step in a cyclical fashion (60)

The assumption of a linear pathway in *M. maripaludis* without accounting for electron bifurcation can affect the downstream predictions in the metabolic model. The default mechanism of energy conservation in our initial, uncurated model matched methylotrophic methanogens and utilized methanophenazine, an electron carrier known to be absent from *M. maripaludis* and other hydrogenotrophic methanogens. We removed methanophenazine-based electron flow and added the correct electron bifurcation pathway, linking heterodisulfide reduction with electrons from H2 to carbon dioxide reduction via reduced ferredoxin instead. This commonly-encountered reconstruction pitfall, in which information available in annotation databases does not sufficiently represent recently elucidated metabolic pathways, emphasizes the need to keep abreast of updated academic literature in spite of the improvement of automatic reconstruction methods.

To demonstrate that the linear pathway cannot support growth of *M. maripaludis*, we altered the native electron bifurcating heterodisulfide reductase (HdrABC) reaction:

by removing ferredoxin, balancing mass and charge to yield:

This scenario represented a hypothetical case where *M. maripaludis* does not contain a membrane-bound HdrDE complex but cannot perform electron bifurcation. We optimized this altered model for growth on CO2 + H2 and were unable to predict *in silico* growth, supporting the observation that the ferredoxin reduction via electron bifurcation is an essential part of our network. Lack of model growth can clearly be attributed to disruption of the central energy coupling mechanism in *M. maripaludis*, in which electron bifurcation must necessarily reduce ferredoxin for reducing CO2. The alternative source of reduced ferredoxin is the energy-converting Eha hydrogenase, which utilizes a sodium ion gradient to reduce ferredoxin with H2 on a 1:1 basis. CO2 reduction to methane requires reduced ferredoxin and pumps out sodium ions, also on a 1:1 basis. Thus, each cycle of methanogenesis in this scenario effectively produces no sodium ion gradient for synthesizing ATP, the central component necessary for biomass formation. Additionally, methanogenesis loses small amounts of carbon for biosynthesis; hence, reducing one ferredoxin effectively pumps less than one sodium ion across the cell membrane and creates an overall energy deficit. Overall, this simulation illustrates the essentiality of ferredoxin reduction via electron bifurcation and reinforces the idea that Eha hydrogenase can play only an anaplerotic role in methanogenesis (45).

Taking this analysis one step further, we used our reconstruction to probe into acetate assimilation, a pathway in *M. maripaludis* that can enhance growth but cannot replace H2 and CO2 as an energy source (61). This is in contrast to multiple methylotrophic methanogens such as *Methanosarcina barkeri* that can subsist using solely the aceticlastic pathway (62). It is unknown why *M. maripaludis* cannot be grown on acetate alone, and our reconstruction did not reveal any strictly stoichiometric obstacle to growth. However, much like the pathway in *M. barkeri,* an aceticlastic pathway in *M. maripaludis* would require energy-converting hydrogenases (Eha and Ehb) to produce H2 using reduced ferredoxin, pumping out sodium ions, and thrusting this reaction into a central stoichiometric role rather than an anaplerotic one. As shown in Figure 4, when we simulated our model and allowed Eha/Ehb unlimited flux, we could predict aceticlastic growth with Eha/Ehb oxidizing approximately two moles of ferredoxin per methane produced. We then constrained our model to enforce a solely anaplerotic or biosynthetic role of energy-converting hydrogenase by limiting flux through the Eha/Ehb reaction to 10% that of methane secretion rate. Doing so prevented our model from predicting growth from acetate alone, but did not restrict hydrogenotrophic growth or supplementary acetate uptake. This simulation supports the hypothesis that *M. maripaludis* cannot achieve aceticlastic growth because Eha or Ehb cannot assume a central role in methanogenesis. In keeping with this hypothesis, we have restricted flux through Eha/Ehb in our model to ≤ 10% of methane secretion as a default constraint.

Interestingly, there is evidence that *M. maripaludis* uses multiple forms of ferredoxin as electron carriers and may link certain steps, particularly those involved in electron bifurcation, reduction of CO2 to formylmethanofuran, and certain biosynthetic reactions, using specific ferredoxins (63). Presently, the full extent of this phenomenon is not well understood and requires more experimental investigation. However, in an effort to represent ferredoxin specificity in our model, we have included a function (see Supplemental Materials) that replaces promiscuous ferredoxins with two types of specific ferredoxins. One type is used for the Eha hydrogenase, Hdr, and formylmethanfuran dehydrogenase (Fwd) and the other type for Ehb hydrogenase and biosynthetic carboxylating oxidoreductases, as suggested by (44). Using this function tightens the coupling between the aforementioned reactions by restricting each set to one pool of electron carriers and allows us to predict how ferredoxin specificity could change possible model flux distributions. In wild type simulations, this change has minimal effects on predicted growth yields and fluxes but could have notable impact on gene knockout predictions, particularly those involving reactions that utilize ferredoxin. Moreover, electron movement through different ferredoxin species could have important implications for hypothesizing strain designs, thus including multiple ferredoxins could be vital for effective metabolic engineering.

## Other Biochemistry Improvements

A major part of our manual curation was adding biosynthesis pathways for the methanogenic coenzymes, sugars, and lipids. *M. maripaludis* utilizes a number of unusual coenzymes (methanofuran, H4MPT, coenzyme F420, coenzyme B, coenzyme M, coenzyme F430) as carbon and electron carriers during methanogenesis (64). It also contains recently characterized pathways for synthesizing a tetrasaccharide for N-linked glycosylation of archaellin (archeal flagellin) (65) and multiple forms of archaeol, an archaeal membrane ether lipid (66). None of these pathways were included in our draft reconstruction and few were completely present in the Model SEED database, thus the bulk of these reactions were added manually. These synthesis pathways, particularly for coenzymes, are vital pieces of *M. maripaludis* metabolism that set it apart from the vast majority of known biochemistry. Hence, including these synthesis pathways and adding these metabolites as required biomass components was crucial for distinguishing our reconstruction from existing networks.

In a similar vein, we sought to accurately represent sulfur assimilation, a pathway not yet fully understood in *M. maripaludis*. Sulfate is known not to be the sulfur source for *M. maripaludis*; moreover,sulfate reduction would produce sulfite, a methanogenesis inhibitor (67). However, because sulfate is the default sulfur source for most microorganisms, our first draft reconstruction included a sulfate transporter and sulfate reduction pathway. We removed this default pathway and instead added a pathway to utilize hydrogen sulfide gas, the primary sulfur source for *M. maripaludis.* Our updated sulfur assimilation pathway includes sulfide oxidation to sulfite—an essential metabolite for multiple biosynthetic pathways—via a hypothesized dissimilatory sulfite reductase-like protein (29). Taken together with aforementioned syntheses, these modifications demonstrated the need for rigorous manual curation to add known biochemical pathways that were not part of the automated reconstruction and remove pathways that are known not to function in the organism. By employing these methods and by working collaboratively with *M. maripaludis* experts, we have created a reconstruction that maximizes consistency with biochemical literature of our organism.

## Growth Yield Validation and ATP Maintenance

Evaluating a metabolic network reconstruction by qualitatively comparing it to known biochemical phenomena is a valuable way to gauge how close the network can represent actual biochemistry. To make more quantitative comparisons, we must convert the reconstruction to a metabolic model by imposing flux constraints on the network, enforcing mass balance on all metabolites, and optimizing to an objective function (Methods). A common way of quantitatively evaluating the resulting model is to simulate maximum cell growth under steady-state conditions and compare growth yield predictions to experimentally-determined values. Due to the narrow range of possible substrates for our hydrogenotrophic system and scarcity of growth yield data for our organism, we generated our own experimental measurements of growth yield. We conducted chemostat growth experiments under H2-limiting conditions and measured growth yields as described previously (53), but varied our dilution rate to gather a range of different yield measurements. Cell density was assessed using optical density (OD) and was previously reported as OD600=1 corresponding to 0.34 mg(dry weight)∙ml-1 (46). We were unsure of the efficacy of this value, in part because we measured at 660 nm rather than 600 nm. We re-measured this conversion factor using a combination of centrifugation and vacuum filtering (Methods) and plotted a new calibration curve (Supplementary Materials), determining that OD660=1 corresponded to 0.462 ± 0.015 mg(dry weight)∙ml-1. Using this value, we calculated measured growth yields based on growth rates (equal to dilution rates) and measured methane evolution rates (Methods). Measured growth yields are plotted in Figure 4 for 9 independent steady state time points.

We then tested our model by generating growth yield predictions and comparing them to measured growth yields. Growth yield predictions depend not only on metabolic steps where ATP is generated or hydrolyzed, but more heavily on ATP maintenance energies (68). From a modeling perspective, maintenance energies are regarded as the moles of ATP needed to support cellular processes not otherwise depicted in metabolism, including DNA replication, RNA transcription, protein synthesis, and other requirements. We recognized that our model was essentially untrained in terms of ATP maintenance and contained automated values from our first draft reconstruction. Thus it was crucial to train our model by fitting to our experimental dataset. However, we were also wary of overfitting our model by training and testing on the same set of samples. We addressed both concerns by performing leave one out cross validation (LOOCV) on our full dataset. Thus, for each of our nine growth rate values, we used the remaining eight growth rates and their associated measured methane evolution rates to derive ATP maintenance values. We then used that ATP maintenance value in our calculation of predicted growth yield for the given growth rate. Using this method allowed us to essentially test our model’s growth yield predictions on each separate test point while training on the remaining 8 measurements. The resulting predicted growth yields are plotted in Figure 4 along with our measured growth yields. As illustrated by this plot, our model was able to consistently predict growth yield within the 95% confidence interval of a measured test sample after being trained on a separate dataset. Though growth yield validation is not an absolute measure of model performance, our model’s ability to closely reproduce experimental results in a LOOCV setting that mitigated overfitting suggested a high propensity for generating viable growth predictions. Moreover, the relative consistency between measured and predicted values indicated our model’s robustness for predicting growth yields across a range of different dilution and methane secretion rates.

We also used the full dataset of growth rates and methane evolution rates to set final values for growth associated maintenance (GAM) and non-growth associated maintenance (NGAM). The GAM represents ATP hydrolysis required to support growth-related processes and NGAM represents ATP hydrolysis required for non-growth associated cellular upkeep. GAM was originally set as 40.11 (mmol per grams [cell mass]), a relatively low value when compared with that of a fast-growing bacterial species, such as the GAM of 59.81 in *E. coli* (69). NGAM, represented by simple ATP hydrolysis, was unbounded in our first draft reconstruction and took on a value of 0 during all model simulations. After training on our full dataset, we set our GAM and NGAM values to 169.9 mmol ATP per gram [cell mass] and 5.0 mmol ATP per gram [cell mass] h-1, respectively (see Supplemental Material). Notably, these maintenance values are much higher than those in other methanogen models; for example, fellow methanogen *Methanosarcina barkeri* was reported to have a GAM of 65.00 (mmol per grams [cell mass]) (68), about 38% of our calculated value. This difference is reflective of the observed differences in growth yield for these organisms during growth on H2 and CO2. Using the same formula for growth yield in each case at nearly identical doubling times of 12 h, *M. maripaludis* grew at a yield of about 33% of that reported for *M. barkeri* (68)*.* Thus, though we calculated unusually high ATP maintenance requirements for growth, these high values reflect observed differences in growth data when comparing to a methylotrophic methanogen growing on the same substrates.

## Gene Knockout Validation

Gene knockout experiments present a different method for validating a metabolic reconstruction based on its model. At its core, a constraint-based model is built around gene-protein-reaction relationships that connect genotype to growth phenotype. Thus, comparing model predictions of gene knockout lethality provides an excellent way to quantitatively measure the qualitative content of the model. This process hinges on the availability of gene knockout data for the organism being modeled, ideally with the abundance of data found for a traditional model organism such as *Escherichia coli* (70). *M. maripaludis* lacks this abundance of *in vivo* gene knockout data, but was used for transposon mutagenesis to calculate an essentiality index of all of its genes (71)*.* Although this dataset does not contain the same quality of knockout data as actual knockout experiments, it provides a valuable “first pass” test set for gene essentiality of our model. However, essentiality index is itself a model for predicting gene knockout lethality, thus although we compared our model’s predictions to this dataset (see Supplemental Material) it did not provide the same clear picture as targeted knockout experiments.

Because much of methanogenesis revolves around the function of different hydrogenases, the bulk of available gene knockout data involves hydrogenase knockouts on different media. For our test set, we were able to assemble a knockout panel of 30 binary growth phenotypes based on previous publications (44–49). Though the breadth of these knockout genotypes is limited, they are all vital pieces of central carbon metabolism and therefore, they give us a good idea of how well our model can predict knockouts in central catabolism. In comparing with these data, as shown in Figure 5, our model achieved 90% prediction accuracy and a Matthew’s correlation coefficient of 0.67. These high values suggested that our model is an excellent predictor of growth phenotype based on genotype changes in central carbon metabolism. This result was particularly encouraging because we avoided training our model on this dataset in the interest of preventing overfitting our model to the validation set.

It is also worth noting that all 3 incorrect predictions have similar bases in the model. In these cases, knockouts of 5 or 6 hydrogenases are experimentally found to be lethal in formate-grown cells, or in formate + CO-grown cells lacking carbon monoxide dehydrogenase (CODH), yet our model predicts these knockouts to be non-lethal. The reason for this disagreement lies in our innate assumption that every reaction performs at 100% efficiency, an ideal scenario that is not achievable in an actual organism. Methanogensis cannot be expected to operate at 100% enzyme efficiency, as some of substrates and electron carriers will not react, thus it can be considered as a “leaky” process where a portion of the metabolites are unused in every cycle. Specifically, in the Δ5H2ase and Δ6H2ase knockouts, small amounts of hydrogen are synthesized in biosynthetic reactions. Eha hydrogenase remains active in each mutant and can use this hydrogen to supply anaplerotic reduced ferredoxin for methanogenesis. However, in reality an additional non-stoichiometric amount of hydrogen is required. Thus, the actual mutants cannot grow on formate alone and require hydrogen. [Notably, most of our knockout predictions were made with glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) constrained to carry zero flux. The GAPOR reaction is ferredoxin-reducing and can serve as a supplemental source of reduced ferredoxin for growth on formate in the case of Eha knockout (47). However, in wild type strains the expression of GAPOR is not sufficient to support growth in the absence of other hydrogenases (e.g. the Δ5H2ase and Δ6H2ase mutants). As demonstrated previously, overexpression of the GAPOR operon allows for growth of these mutants (Δ6H2asesupp and Δ7H2asesup) on formate (47). To best reflect these genotypic differences, we altered the bounds of the GAPOR reaction (rxn07191[c0]) in our knockout simulation code, constraining the reaction to zero flux in all cases except those of the Δ6H2asesupp and Δ7H2asesup mutants.]

## Thermodynamic Calculations

Free energy plays a key role in biochemistry as all biological systems must have a sufficiently low overall free energy to support growth. When simulating optimal growth using a metabolic model we expect the same rules to apply to our system, hence we can apply thermodynamic constraints to the model based on metabolite free energies of formation. In a previous study, free energies of formation were used to constrain reversibility of all internal model reactions based on the second law of thermodynamics (33). This method, while rigorous, is highly dependent on concentration and can be overly restrictive with regard to predicted flux distributions; thus it is most effective when paired with metabolite effective concentration data (34). Lacking extensive effective concentration data for *M. maripaludis,* we chose to represent free energy constraints in a novel approach where we add free energies only to exchange reactions, the set of metabolites that can be taken up or produced by the model. These metabolites effectively represent the organism’s overall biochemical “reaction”; therefore it is reasonable to expect this overall reaction must produce a negative overall free energy to support growth. Indeed, applying this method to our default model growing on H2 + CO2 with methane evolution rate of 50 mmol/g(dry weight)∙h, overall free energy production is predicted as -5.59 kJ/g(dry weight). Optionally, this calculation can be used as an additional model constraint that restricts overall free energy to be negative, the equivalent of imposing the second law of thermodynamics on the organism itself.

We expect that this straightforward calculation (Methods) will be a useful addition to our model, particularly as we aim to use it as a platform for generating possible strain designs. With regard to free energy, methanogens are particularly notable in that they subsist close to the thermodynamic limit to support growth (32). It follows that for any potential strain design, we must pay particular attention to the overall free energy of our system, lest it dip below this vital threshold. It may also provide a metric for differentiating between multiple feasible strain designs by ranking them in order of thermodynamic feasibility. At the very least, it serves as an additional capability of our model and as a checkpoint to ensure that our overall stoichiometry matches up with overall free energy. We have included example functions for adding metabolite free energies to our model and performing FBA with an additional free energy calculation (Supplementary Materials).

## Reconstruction and Model Availability

Reconstructing a metabolic network is an iterative process and therefore, it is paramount that reconstructions be as clear as possible to encourage future updates and expansions (42). We have strived for clarity in both our nomenclature and in our decision making process for including each reaction present in our reconstruction. Reactions and metabolites in our network are based upon identifiers and names found in Kbase, but also include crosslinks to ChEBI (72) and KEGG identifiers (36), enzyme commission numbers, and reaction subsystems where available. Each reaction in the reconstruction is also connected to its literature and/or database source, plus its reaction confidence score when applicable (Supplementary Materials).

Additionally, we have sought to maximize usability of both our reconstruction and our model. The systems biology markup language (SBML) is a standard medium for distributing metabolic reconstructions (73); thus, we have included our reaction network in SBML level 2, the highest version currently supported by the COBRA Toolbox (41). In our experience using reconstructions from other groups, we have found a wide range of usability, from those that can easily be imported and simulated to those that are difficult to use and interpret. In the interest of making our simulations and results easy to reproduce, we have included our reconstruction in MATLAB data structure format and all of our codes for simulating model growth on different media and gene knockout phenotypes (Supplementary Materials).We have also made our codes and reconstruction available on Github (https://github.com/marichards/methanococcus).

# Discussion

Genome scale metabolic reconstructions provide a wide lens for studying the biochemical complexity in a computational setting. We used likelihood based gapfilling and meticulous manual curation to build iMR540, a comprehensive reconstruction of *M. maripaludis* that incorporates electron bifurcation to portray cyclical hydrogenotrophic methanogenesis. We incorporated many unique pathways that differentiate our network from those for other organisms, creating a novel tool for understanding and probing more deeply into hydrogenotrophic methanogenesis. The resulting network model compared favorably with measured growth yield and gene knockout data and provided a platform to develop a new method for estimating overall free energy generation during steady state growth.

Electron bifurcation is the central energy conservation mechanism in *M. maripaludis,* thus it is fitting that this process takes a central role in our reaction network. This mechanism is in stark contrast to existing methanogen models that contain linear methanogenesis based on oxidative (electron transport) phosphorylation (28, 68, 74). While the linear model is correct for methanogens with cytochromes, it is not correct for methanogens without cytochromes such as *M. maripaludis*. We have demonstrated that, in the absence of a membrane-bound HdrDE complex, ferredoxin reduction via electron bifurcation is essential for predicting growth in our network. Furthermore, constraining the energy-conserving Eha/Ehb reaction to a minor metabolic role provides a stoichiometric hypothesis for the inability of *M. maripaludis* to grow aceticlastically and will undoubtedly influence model predictions moving forward. Ferredoxin specificity for these and other reactions remains an open question that could profoundly affect electron carrier utilization and have implications in native and mutant genotypes, a possibility we have acknowledged by allowing either promiscuous or specific ferredoxins in our reconstruction.

Beyond bifurcation itself, we added numerous uncommon biosynthetic pathways to our network from literature sources that further separate it from models of other organisms. These pathways included syntheses for methanogenic coenzymes, archaellin sugars and archaeol lipids as well as a relatively novel sulfur assimilation pathway. Additionally, using likelihood-based gap filling helped us automatically identify 66 more genes, increasing the gene coverage of our reconstruction prior to the start of manual curation and assigning reaction likelihood scores for many reactions that lend a measure of confidence level to network. The efficacy of these methods is shown not only in the qualitative accuracy of our reconstruction, but also in the formidable quantitative capabilities of the resulting model. Our model performed well in a LOOCV analysis of growth yield data and compared favorably with experimental gene knockout data, suggesting a high propensity for generating predictions that are consistent with observed biology.

For a methanogen living close to the edge of thermodynamic feasibility, we also thought it salient to include some calculation of overall free energy when simulating our model. We have thus introduced a novel method of predicting overall model free energy generation based solely on standard free energies and concentrations of exchange metabolites. Though a relatively trivial calculation, our method gives a quick assessment of whether a predicted flux distribution is thermodynamically possible and could prove a particularly useful tool for guiding future metabolic engineering designs.

While considering our reconstruction’s consistency with existing literature and our model’s high performance on measured data, it is poignant that we acknowledge the limitations in our network. First, though we have attempted to address as many parts of metabolism as possible, many “dark areas” of *M. maripaludis* metabolism still exist in our reconstruction. For many of these cases, gene annotations from Kbase and likelihood based gap filling give us starting hypotheses for what may be occurring in these dark areas, but the accuracy of these predictions remains unknown until they have been biochemically characterized. We recognize that our reconstruction effort represents only an incremental step toward understanding *M. maripaludis* metabolism and that many other users may follow in our footsteps. With these considerations in mind, we strived for maximum transparency in our metabolic network to make our reconstruction decisions apparent to future users and to make our results easily reproducible. There is ample opportunity for improving our reconstruction in the future by elucidating the missing information for these dark areas and we hope that by providing information on the origins and likelihoods of our reactions, we can encourage exploration of these as-yet-unknown pathways.

Second, we recognize that even for the areas of metabolism that we understand well, our model is purely stoichiometric and therefore can only provide predictions from a metabolic perspective. This somewhat limits the scope of questions we can ask using our reconstruction because it does not explicitly include information for other cellular processes, e.g. transcriptional regulation. Given the wide expanse of unknown metabolism, we do not perceive this limitation as particularly crippling, as we can still ask a plentiful supply of questions just within the realm of stoichiometry. In the future, if we wish to address this limitation our stoichiometric predictions could be combined with those from other types of structures, thus providing the tools to probe questions that include other cellular processes.

Lastly, we stress that even within the metabolic space, our model’s power lies in predicting the scope of metabolic possibility, not absolute biological reality. Any particular flux distribution should be considered a hypothesis about the what our organism can theoretically achieve, not a precise prediction about all metabolic fluxes. These predictions provide valuable insight into the potential metabolic capabilities of our organism, but it would be folly to accept any single prediction as a facsimile of reality. Such a consideration is vital when considering our model or any other model as a tool for facilitating metabolic engineering designs because any model prediction should be considered as a starting point rather than a final product. By explicitly acknowledging this limitation, we hope to realistically portray the capabilities of our reconstruction as a tool to better understand the unique biochemistry of hydrogenotrophic methanogens, push forward biochemical discovery in these organisms, and unlock their potential as metabolic engineering targets.

# Acknowledgements

We would like to thank Eliora Gachelet for assisting enormously in chemostat growth experiments and Dr. Matthew Benedict for his expertise and advice concerning methanogenic archaea, likelihood-based gap filling, and metabolic model construction.

# References

1. **Haynes CA**, **Gonzalez R**. 2014. Rethinking biological activation of methane and conversion to liquid fuels. Nat Chem Biol **10**:331–339.

2. **Levi M**. 2013. Climate consequences of natural gas as a bridge fuel. Clim Change **118**:609–623.

3. **Mueller TJ**, **Grisewood MJ**, **Nazem-Bokaee H**, **Gopalakrishnan S**, **Ferry JG**, **Wood TK**, **Maranas CD**. 2014. Methane oxidation by anaerobic archaea for conversion to liquid fuels. J Ind Microbiol Biotechnol **42**:391–401.

4. **Montzka SA**, **Dlugokencky EJ**, **Butler JH**. 2011. Non-CO2 greenhouse gases and climate change. Nature **476**:43–50.

5. **Kirschke S**, **Bousquet P**, **Ciais P**, **Saunois M**, **Canadell JG**, **Dlugokencky EJ**, **Bergamaschi P**, **Bergmann D**, **Blake DR**, **Bruhwiler L**, **Cameron-Smith P**, **Castaldi S**, **Chevallier F**, **Feng L**, **Fraser A**, **Heimann M**, **Hodson EL**, **Houweling S**, **Josse B**, **Fraser PJ**, **Krummel PB**, **Lamarque J-F**, **Langenfelds RL**, **Le Quéré C**, **Naik V**, **O’Doherty S**, **Palmer PI**, **Pison I**, **Plummer D**, **Poulter B**, **Prinn RG**, **Rigby M**, **Ringeval B**, **Santini M**, **Schmidt M**, **Shindell DT**, **Simpson IJ**, **Spahni R**, **Steele LP**, **Strode SA**, **Sudo K**, **Szopa S**, **van der Werf GR**, **Voulgarakis A**, **van Weele M**, **Weiss RF**, **Williams JE**, **Zeng G**. 2013. Three decades of global methane sources and sinks. Nat Geosci **6**:813–823.

6. **Thauer RK**, **Kaster A-K**, **Seedorf H**, **Buckel W**, **Hedderich R**. 2008. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol **6**:579–591.

7. **DiMarco AA**, **Bobik TA**, **Wolfe RS**. 1990. Unusual coenzymes of methanogenesis. Annu Rev Biochem **59**:355–394.

8. structure of func of enzymes H2CO2 pathway 2002.pdf.

9. **Costa KC**, **Leigh JA**. 2014. Metabolic versatility in methanogens. Curr Opin Biotechnol **29**:70–75.

10. **Welte C**, **Deppenmeier U**. 2014. Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. Biochim Biophys Acta BBA - Bioenerg **1837**:1130–1147.

11. **Heiden S**, **Hedderich R**, **Setzke E**, **Thauer RK**. 1993. Purification of a cytochrome b containing H2:heterodisulfide oxidoreductase complex from membranes of *Methanosarcina barkeri*. Eur J Biochem **213**:529–535.

12. **Kaster A-K**, **Moll J**, **Parey K**, **Thauer RK**. 2011. Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proc Natl Acad Sci **108**:2981–2986.

13. **Jones WJ**, **Paynter MJB**, **Gupta R**. 1983. Characterization of *Methanococcus maripaludis* sp. nov., a new methanogen isolated from salt marsh sediment. Arch Microbiol **135**:91–97.

14. **Hendrickson EL**, **Kaul R**, **Zhou Y**, **Bovee D**, **Chapman P**, **Chung J**, **Macario EC de**, **Dodsworth JA**, **Gillett W**, **Graham DE**, **Hackett M**, **Haydock AK**, **Kang A**, **Land ML**, **Levy R**, **Lie TJ**, **Major TA**, **Moore BC**, **Porat I**, **Palmeiri A**, **Rouse G**, **Saenphimmachak C**, **Söll D**, **Dien SV**, **Wang T**, **Whitman WB**, **Xia Q**, **Zhang Y**, **Larimer FW**, **Olson MV**, **Leigh JA**. 2004. Complete Genome Sequence of the Genetically Tractable Hydrogenotrophic Methanogen *Methanococcus maripaludis*. J Bacteriol **186**:6956–6969.

15. **Sarmiento FB**, **Leigh JA**, **Whitman WB**. 2011. Chapter three - Genetic Systems for Hydrogenotrophic Methanogens, p. 43–73. *In* Ragsdale, ACR and SW (ed.), Methods in Enzymology. Academic Press.

16. **Graham DE**, **White RH**. 2002. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nat Prod Rep **19**:133–147.

17. **Stock T**, **Selzer M**, **Connery S**, **Seyhan D**, **Resch A**, **Rother M**. 2011. Disruption and complementation of the selenocysteine biosynthesis pathway reveals a hierarchy of selenoprotein gene expression in the archaeon *Methanococcus maripaludis*. Mol Microbiol **82**:734–747.

18. **Haydock AK**, **Porat I**, **Whitman WB**, **Leigh JA**. 2004. Continuous culture of *Methanococcus maripaludis* under defined nutrient conditions. FEMS Microbiol Lett **238**:85–91.

19. **Hendrickson EL**, **Liu Y**, **Rosas-Sandoval G**, **Porat I**, **Soll D**, **Whitman WB**, **Leigh JA**. 2008. Global Responses of Methanococcus maripaludis to Specific Nutrient Limitations and Growth Rate. J Bacteriol **190**:2198–2205.

20. **Xia Q**, **Wang T**, **Hendrickson EL**, **Lie TJ**, **Hackett M**, **Leigh JA**. 2009. Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen Methanococcus maripaludis. BMC Microbiol **9**:149.

21. **Yoon SH**, **Turkarslan S**, **Reiss DJ**, **Pan M**, **Burn JA**, **Costa KC**, **Lie TJ**, **Slagel J**, **Moritz RL**, **Hackett M**, **Leigh JA**, **Baliga NS**. 2013. A systems level predictive model for global gene regulation of methanogenesis in a hydrogenotrophic methanogen. Genome Res **23**:1839–1851.

22. **Johnson EF**, **Mukhopadhyay B**. 2008. Coenzyme F420-Dependent Sulfite Reductase-Enabled Sulfite Detoxification and Use of Sulfite as a Sole Sulfur Source by Methanococcus maripaludis. Appl Environ Microbiol **74**:3591–3595.

23. **Lie TJ**, **Dodsworth JA**, **Nickle DC**, **Leigh JA**. 2007. Diverse homologues of the archaeal repressor NrpR function similarly in nitrogen regulation. FEMS Microbiol Lett **271**:281–288.

24. **Kauffman KJ**, **Prakash P**, **Edwards JS**. 2003. Advances in flux balance analysis. Curr Opin Biotechnol **14**:491–496.

25. **Simeonidis E**, **Price ND**. 2015. Genome-scale modeling for metabolic engineering. J Ind Microbiol Biotechnol **42**:327–338.

26. **Milne CB**, **Kim P-J**, **Eddy JA**, **Price ND**. 2009. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. Biotechnol J **4**:1653–1670.

27. **Stolyar S**, **Van Dien S**, **Hillesland KL**, **Pinel N**, **Lie TJ**, **Leigh JA**, **Stahl DA**. 2007. Metabolic modeling of a mutualistic microbial community. Mol Syst Biol **3**:92.

28. **Goyal N**, **Widiastuti H**, **Karimi IA**, **Zhou Z**. 2014. A genome-scale metabolic model of *Methanococcus maripaludis S2* for CO2 capture and conversion to methane. Mol Biosyst **10**:1043–1054.

29. **Susanti D**, **Mukhopadhyay B**. 2012. An Intertwined Evolutionary History of Methanogenic Archaea and Sulfate Reduction. PLoS ONE **7**:e45313.

30. **Graham DE**, **White RH**. 2002. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nat Prod Rep **19**:133–147.

31. **Benedict MN**, **Mundy MB**, **Henry CS**, **Chia N**, **Price ND**. 2014. Likelihood-Based Gene Annotations for Gap Filling and Quality Assessment in Genome-Scale Metabolic Models. PLoS Comput Biol **10**:e1003882.

32. **Jackson BE**, **McInerney MJ**. 2002. Anaerobic microbial metabolism can proceed close to thermodynamic limits. Nature **415**:454–456.

33. **Henry CS**, **Broadbelt LJ**, **Hatzimanikatis V**. 2007. Thermodynamics-Based Metabolic Flux Analysis. Biophys J **92**:1792–1805.

34. **Hoppe A**, **Hoffmann S**, **Holzhütter H-G**. 2007. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. BMC Syst Biol **1**:23.

35. **Thiele I**, **Palsson BØ**. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc **5**:93–121.

36. **Kanehisa M**, **Goto S**. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res **28**:27–30.

37. **Caspi R**, **Altman T**, **Dale JM**, **Dreher K**, **Fulcher CA**, **Gilham F**, **Kaipa P**, **Karthikeyan AS**, **Kothari A**, **Krummenacker M**, **Latendresse M**, **Mueller LA**, **Paley S**, **Popescu L**, **Pujar A**, **Shearer AG**, **Zhang P**, **Karp PD**. 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res **38**:D473–D479.

38. **Henry CS**, **DeJongh M**, **Best AA**, **Frybarger PM**, **Linsay B**, **Stevens RL**. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol **28**:977–982.

39. **Price ND**, **Reed JL**, **Palsson BØ**. 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol **2**:886–897.

40. **Feist AM**, **Palsson BO**. 2010. The biomass objective function. Curr Opin Microbiol **13**:344–349.

41. **Schellenberger J**, **Que R**, **Fleming RMT**, **Thiele I**, **Orth JD**, **Feist AM**, **Zielinski DC**, **Bordbar A**, **Lewis NE**, **Rahmanian S**, **Kang J**, **Hyduke DR**, **Palsson BØ**. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc **6**:1290–1307.

42. **Heavner BD**, **Price ND**. 2015. Transparency in metabolic network reconstruction enables scalable biological discovery. Curr Opin Biotechnol **34**:105–109.

43. **Kostromins A**, **Stalidzans E**. 2012. Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism. Biosystems **109**:233–239.

44. **Porat I**, **Kim W**, **Hendrickson EL**, **Xia Q**, **Zhang Y**, **Wang T**, **Taub F**, **Moore BC**, **Anderson IJ**, **Hackett M**, **Leigh JA**, **Whitman WB**. 2006. Disruption of the Operon Encoding Ehb Hydrogenase Limits Anabolic CO2 Assimilation in the Archaeon *Methanococcus maripaludis*. J Bacteriol **188**:1373–1380.

45. **Lie TJ**, **Costa KC**, **Lupa B**, **Korpole S**, **Whitman WB**, **Leigh JA**. 2012. Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. Proc Natl Acad Sci **109**:15473–15478.

46. **Lupa B**, **Hendrickson EL**, **Leigh JA**, **Whitman WB**. 2008. Formate-Dependent H2 Production by the Mesophilic Methanogen *Methanococcus maripaludis*. Appl Environ Microbiol **74**:6584–6590.

47. **Costa KC**, **Lie TJ**, **Jacobs MA**, **Leigh JA**. 2013. H2-Independent Growth of the Hydrogenotrophic Methanogen *Methanococcus maripaludis*. mBio **4**:e00062–13.

48. **Costa KC**, **Wong PM**, **Wang T**, **Lie TJ**, **Dodsworth JA**, **Swanson I**, **Burn JA**, **Hackett M**, **Leigh JA**. 2010. Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. Proc Natl Acad Sci **107**:11050–11055.

49. **Hendrickson EL**, **Leigh JA**. 2008. Roles of Coenzyme F420-Reducing Hydrogenases and Hydrogen- and F420-Dependent Methylenetetrahydromethanopterin Dehydrogenases in Reduction of F420 and Production of Hydrogen during Methanogenesis. J Bacteriol **190**:4818–4821.

50. **Matthews BW**. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta BBA - Protein Struct **405**:442–451.

51. **Flamholz A**, **Noor E**, **Bar-Even A**, **Milo R**. 2011. eQuilibrator—the biochemical thermodynamics calculator. Nucleic Acids Res gkr874.

52. **Jankowski MD**, **Henry CS**, **Broadbelt LJ**, **Hatzimanikatis V**. 2008. Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks. Biophys J **95**:1487–1499.

53. **Costa KC**, **Yoon SH**, **Pan M**, **Burn JA**, **Baliga NS**, **Leigh JA**. 2013. Effects of H2 and Formate on Growth Yield and Regulation of Methanogenesis in *Methanococcus maripaludis*. J Bacteriol **195**:1456–1462.

54. **Setzke E**, **Hedderich R**, **Heiden S**, **Thauer RK**. 1994. H2: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Eur J Biochem **220**:139–148.

55. **Thauer RK**, **Kaster A-K**, **Seedorf H**, **Buckel W**, **Hedderich R**. 2008. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol **6**:579–591.

56. **Hedderich R**, **Thauer R k.** 1988. Methanobacterium thermoautotrophicum contains a soluble enzyme system that specifically catalyzes the reduction of the heterodisulfide of coenzyme M and 7-mercaptoheptanoylthreonine phosphate with H2. FEBS Lett **234**:223–227.

57. **Nitschke W**, **Russell MJ**. 2012. Redox bifurcations: Mechanisms and importance to life now, and at its origin: A widespread means of energy conversion in biology unfolds…. BioEssays **34**:106–109.

58. **Herrmann G**, **Jayamani E**, **Mai G**, **Buckel W**. 2008. Energy Conservation via Electron-Transferring Flavoprotein in Anaerobic Bacteria. J Bacteriol **190**:784–791.

59. **Costa KC**, **Wong PM**, **Wang T**, **Lie TJ**, **Dodsworth JA**, **Swanson I**, **Burn JA**, **Hackett M**, **Leigh JA**. 2010. Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. Proc Natl Acad Sci **107**:11050–11055.

60. **Thauer RK**. 2012. The Wolfe cycle comes full circle. Proc Natl Acad Sci **109**:15084–15085.

61. **Shieh JS**, **Whitman WB**. 1987. Pathway of acetate assimilation in autotrophic and heterotrophic methanococci. J Bacteriol **169**:5327–5329.

62. **Welander PV**, **Metcalf WW**. 2005. Loss of the *mtr* operon in *Methanosarcina* blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. Proc Natl Acad Sci U S A **102**:10664–10669.

63. **Kaster A-K**, **Goenrich M**, **Seedorf H**, **Liesegang H**, **Wollherr A**, **Gottschalk G**, **Thauer RK**. 2011. More Than 200 Genes Required for Methane Formation from H2 and CO2 and Energy Conservation Are Present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. Archaea **2011**:1–23.

64. **DiMarco AA**, **Bobik TA**, **Wolfe RS**. 1990. Unusual coenzymes of methanogenesis. Annu Rev Biochem **59**:355–394.

65. **Siu S**, **Robotham A**, **Logan SM**, **Kelly JF**, **Uchida K**, **Aizawa S-I**, **Jarrell KF**. 2015. Evidence that Biosynthesis of the Second and Third Sugars of the Archaellin Tetrasaccharide in the Archaeon *Methanococcus maripaludis* Occurs by the Same Pathway Used by *Pseudomonas aeruginosa* To Make a Di-N-Acetylated Sugar. J Bacteriol **197**:1668–1680.

66. **Jain S**, **Caforio A**, **Driessen AJM**. 2014. Biosynthesis of archaeal membrane ether lipids. Front Microbiol **5**.

67. **Balderston WL**, **Payne WJ**. 1976. Inhibition of methanogenesis in salt marsh sediments and whole-cell suspensions of methanogenic bacteria by nitrogen oxides. Appl Environ Microbiol **32**:264–269.

68. **Gonnerman MC**, **Benedict MN**, **Feist AM**, **Metcalf WW**, **Price ND**. 2013. Genomically and biochemically accurate metabolic reconstruction of *Methanosarcina barkeri* Fusaro, iMG746. Biotechnol J **8**:1070–1079.

69. **Feist AM**, **Henry CS**, **Reed JL**, **Krummenacker M**, **Joyce AR**, **Karp PD**, **Broadbelt LJ**, **Hatzimanikatis V**, **Palsson BØ**. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol **3**.

70. **Orth JD**, **Conrad TM**, **Na J**, **Lerman JA**, **Nam H**, **Feist AM**, **Palsson BO**. 2014. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism--2011. Mol Syst Biol **7**:535–535.

71. **Sarmiento F**, **Mrázek J**, **Whitman WB**. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. Proc Natl Acad Sci **110**:4726–4731.

72. **Degtyarenko K**, **Matos P de**, **Ennis M**, **Hastings J**, **Zbinden M**, **McNaught A**, **Alcántara R**, **Darsow M**, **Guedj M**, **Ashburner M**. 2008. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res **36**:D344–D350.

73. **Hucka M**, **Finney A**, **Sauro HM**, **Bolouri H**, **Doyle JC**, **Kitano H**, **Forum and the rest of the S**, **Arkin AP**, **Bornstein BJ**, **Bray D**, **Cornish-Bowden A**, **Cuellar AA**, **Dronov S**, **Gilles ED**, **Ginkel M**, **Gor V**, **Goryanin II**, **Hedley WJ**, **Hodgman TC**, **Hofmeyr J-H**, **Hunter PJ**, **Juty NS**, **Kasberger JL**, **Kremling A**, **Kummer U**, **Novère NL**, **Loew LM**, **Lucio D**, **Mendes P**, **Minch E**, **Mjolsness ED**, **Nakayama Y**, **Nelson MR**, **Nielsen PF**, **Sakurada T**, **Schaff JC**, **Shapiro BE**, **Shimizu TS**, **Spence HD**, **Stelling J**, **Takahashi K**, **Tomita M**, **Wagner J**, **Wang J**. 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics **19**:524–531.

74. **Benedict MN**, **Gonnerman MC**, **Metcalf WW**, **Price ND**. 2012. Genome-Scale Metabolic Reconstruction and Hypothesis Testing in the Methanogenic Archaeon *Methanosarcina acetivorans* C2A. J Bacteriol **194**:855–865.

# Tables and Figures

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | ***Methanococcus maripaludis S2* model statistics** | | | Protein Coding Genes | 540 | | % ORF Coverage | 31 | | Intra/Extracellular Metabolites | 658/53 | | Dead End Metabolites | 259 | | Internal Reactions | 586 | | Transport/Exchange Reactions | 49/59 | | Dead End Reactions | 206 | | Gene-Associated Reactions | 500 | | Table 1. General statistics for the iMR540 reconstruction. | | |
|  |

Figure 1: A chart showing broad subsystem groupings of the 85 reactions in iMR540 that are not associated with any genes. Reactions falling underneath the “None” subsystem grouping were present in the Model SEED database but had no subsystems listed there and no obvious membership in another subsystem. Reactions grouped within “Other” were dissimilar both from the other categories and from one another, thus we felt they did not merit creation of multiple additional categories.



Figure 2: The native pathway of hydrogenotrophic methanogenesis present in *M. maripaludis.* As shown, electrons from 2 moles of H2 are split between reducing ferredoxin and regenerating coenzymes B and M. Reduced ferredoxin from this reaction links it to CO2 reduction, the first step in the pathway. Enzyme names are shown in **blue**. Metabolites: Fdrd, reduced ferredoxin; Fdox, oxidized ferredoxin ; MFR, methanofuran; HSCoM, coenzyme M; HSCoB, coenzyme B; F420, coenzyme F420. Enzymes: Fwd, formylmethanofuran dehydrogenase; Ftr, formylmethanofuran/H4MPT formyl transferase; Mch, methenyl-H4MPT cyclohydrolase; Hmd, H2-dependent methylene-H4MPT dehydrogenase; Mtd, F420-dependent methylene-H4MPT dehydrogenase; Mer, methylene-H4MPT reductase; Mtr, methyl-H4MPT coenzyme M methyltransferase; Mcr, methyl coenzyme M reductase; Hdr, heterodisulfide reductase;Eha/Ehb, energy-conserving hydrogenases; ATPS, ATP-synthase; Fru, F420-reducing hydrogenase (selenocysteine-containing); Frc. F420-reducing hydrogenase (cysteine-containing).



Figure 3: Hypothetical pathway for aceticlastic methanogenesis in *M. maripaludis*. As demonstrated, this scheme would require 2 cycles of Eha/Ehb in order to oxidize ferredoxin reduced by the CODH/ACS and Hdr reactions. By constraining the Eha/Ehb reaction to only 10% of methane efflux, this pathway becomes infeasible. Enzyme names are shown in **blue.**  Metabolites: Fdrd, reduced ferredoxin; Fdox, oxidized ferredoxin ; MFR, methanofuran; HSCoM, coenzyme M; HSCoB, coenzyme B; F420, coenzyme F420. Enzymes: CODH/ACS, carbon monoxide dehydrogenase/acetyl-CoA synthase complex; Mtr, methyl-H4MPT coenzyme M methyltransferase; Mcr, methyl coenzyme M reductase; Hdr, heterodisulfide reductase;Eha/Ehb, energy-conserving hydrogenases; ATPS, ATP-synthase.



Figure 4: Comparing growth yield predictions on hydrogen to measured data using LOOCV (Methods). All but two predicted growth rates fall within the 95% confidence interval of the measured values. Each of the two outlying points are predicted to grow to higher than measured growth yields.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Genotype | H2 | Formate | H2 + Formate | Formate + CO |
| ∆hmd | N | N | N | N |
| ∆mtd | N | N | N | N |
| ∆frcA | N | N | N | N |
| ∆fruA | N | N | N | N |
| ∆frcA∆fruA | N | N | N | N |
| ∆vhcAU∆vhuA | N | N | N | N |
| ∆hdrB2 | N | N | N | N |
| ∆fdhA1 | N | N | N | N |
| ∆fdhA2 | N | N | N | N |
| ∆fdhA1∆fdhA2 | N | L | N | L |
| ∆fdhA2∆fdhB2 | N | N | N | N |
| ∆ehbF | N | N | N | N |
| ∆3H2ase | N | N | N | N |
| ∆5H2ase | L | N | N | N |
| ∆6H2ase | L | N | N | N |
| ∆6H2ase∆cdh | L | N | N | N |
| ∆6H2asesupp | L | N | N | N |
| ∆7H2asesupp | L | N | N | N |
| **Total Correct:** | **10 of 10** | **14 of 16** | **2 of 2** | **1 of 2** |

Figure 5: Knockout lethality predictions from running FBA on our models show close agreement with experimental results of hydrogenase knockouts. Green boxes indicate growth phenotypes where our models correctly replicated experimental results; red boxes indicate growth phenotypes where our models were incorrect; white boxes indicate growth phenotypes where we lacked experimental validation data. Across the full spectrum of conditions, our models correctly predicted 27 of 30 conditions (90%) accurately, resulting in a strong Matthews Correlation Coefficient of 0.67. This suggests that our reconstruction produces models that accurately depict the effects of genotype alterations on growth phenotypes. L = lethal, N = non-lethal.