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 rumen of multicellular organisms to marine environments and hot springs. In this work we focus on *Methanococcus maripaludis*, a well-studied hydrogenotrophic marine methanogen.Energy conservation in this organism had long been a mystery; only recently, it was reported that it gains energy for growth via electron bifurcation, thriving despite the lack of sufficient coupling sites. To better understand its metabolism and compare it with other methanogens that utilize a traditional chemiosmotic mode of energy conversation, we have built iMR540, a genome scale metabolic reconstruction that accounts for 539 of its 1722 protein-coding genes. 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 distinct methanogenic pathway allows us to accurately replicate experimental growth and gene knockout data. Furthermore, we use our reconstruction to probe into the implications of bifurcation by showing its essentiality in the absence of membrane-bound coupling sites 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 to developing new bio-energy technologies around methane gas. Although a significant portion 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 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) (ref 1) because it burns comparatively cleaner than traditional fossil fuels. Advancing technology is also enabling 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 (ref) and is produced in the environment by biological (refs) and non-biological (refs) sources.. Methanogens are the largest biological contributors of methane, producing about 1 Gt of methane gas per year (4). This group of microorganisms are from the domain Archaea and they grow on carbon dioxide or one or two carbon compounds (5, 6) using enzymes containing unique biological co-factors (Shima ref).

Though phylogenetically and metabolically diverse, methanogens are separated into two groups based on the presence or absence of cytochromes (4). The cytochrome-lacking methanogens have a narrow substrate range of only H2 or formate as electron donors that reduce CO2 to methane (Kyle’s ref); they are sometimes referred to as hydrogenotrophic methanogens. In contrast, cytochrome-containing (or methylotropic) methanogens utilize acetate and methylated compounds for methanogenic growth with a minority also being able to grow hydrogenotrophically. Both groups have similar central pathways of CO2 reduction as well as one universally conserved membrane bound energy conservation site (Mtr), but they also possess additional differing modes of energy conversation.

In the last step of methylotrophic methanogenesis, a traditional membrane-bound electron transport via cytochromes serves as the mechanism of energy conservation. Reducing equivalents are transferred to a membrane bound heterodisulfide complex (HdrDE) via cytochromes or methanophenazine resulting in the generation of an ion gradient for ATP synthesis (4, 7). In hydrogenotrophic methanogenesis however, energy is conserved via flavin-linked electron-bifurcation (ref) in lieu of chemiosmosis. Without cytochromes, these organisms use the energy of an energetically exergonic heterodisulfide reduction to drive the endergonic reduction of the ferredoxin associated with the first step of carbon dioxide reduction (Fwd). Donor electrons from hydrogen or formate oxidation bifurcate at the flavin containing subunit A of the heterodisulfide reductase to the high and low potential paths of heterodisulfide and ferredoxin reduction respectively (ref). Thus, the last exergonic methanogenic step drives the first endergonic reaction step in a cycle (reference to the Wolfe cycle). This type of energy conservation is not membrane associated (ref).

*Methanococcus maripaludis* (8) belongs to this group of hydrogenotrophic methanogens. Compared to the larger genome of methylotrophic methanogens, its genome is relatively small and contains for only 1722 protein coding genes (9). However, it grows robustly with a doubling time of 2 hours (8) and is genetically tractable (10), and thus, has been an ideal candidate for studying methanogenesis (ref), unique co-factors and their biosynthesis (Robert white ref), and novel systems of gene regulation (reference to selenocysteine biosynthesis).To avoid environmental fluctuations that can affect gene regulation, a system for continuous culture of *M. maripaludis* (11) has been established for steady state transcriptomic (ref), proteomic (ref) studies of *M. maripaludis* strains. Several groups have also employed larger systems biology approaches to perform predictive studies using this organism (refs). With these tools in place, and the ability for expression of heterologous genes into M. maripaludis (Lie refs and also Biswarup, for eg.), 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) (12) 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 for production of industrial biochemicals in multiple organisms (13, 14). Similarly, a genome scale metabolic reconstruction for *M. maripaludis* would not only provide for a better understanding methanogenesis but also for guiding 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* (15)and under axenic conditions (16). 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 (15). The latter case was the first genome-scale metabolic reconstruction of *M. maripaludis* (16), an important step towards understanding *M. maripaludis* metabolism.

In our model, iMR540, we made important refinements by updating various pathwaysbased on recent literature*.* The most critical was the critical electron bifurcation step that has been described above as it explains the ability for this organism to grow despite the lack of additional membrane related energy coupling sites (ref). This also includes eliminating methanophenazine utilization and synthesis, which is part of the membrane bound electron transport system of the methylotrophic methanogens and absent in hydrogenotrophic methanogens (4). Additional changes include a corrected sulfur assimilation pathway (17), and the addition of various biosynthesis pathways for all of the unique coenzymes involved in methanogenesis (18). We also increased genome coverage and homology by employing likelihood-based gap filling, a technique that fills reaction gaps based on probability rather than on parsimony (19). 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 stoichiometry 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 (20). 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 (21, 22). 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 (23) 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 (24–26); we chose the Department of Energy Systems Biology Knowledgebase (Kbase; www.kbase.us), a suite of tools that includes the Model SEED reaction database (26). 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 be converted to a simulatable 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 converted to a simulatable 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 converted to 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 (27). 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 (28). 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 (29) in MATLAB [7.14.0.739] (The MathWorks Inc., Natick, MA).

To encourage model transparency (30) 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 (see 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 (31) 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 (see 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” (see Supplementary Materials), which relies heavily on the “deleteModelGenes.m” function in the COBRA Toolbox 2.0 (29). Our experimental test set included 18 knockout genotypes across 4 different growth conditions, with 30 total wet lab experiments across these conditions (32–37). 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 (38):

|  |  |
| --- | --- |
|  | [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 (39) to all exchanges reactions for which these values could be calculated via the group contribution method (40). 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 (see Supplementary Materials) that is built around running the “optimizeCbModel.m” code in the COBRA Toolbox 2.0 (29). Our code accepts effective concentrations (mM) for specified exchange metabolites, assumes standard concentrations 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:

Measuring the total flux of the exchange reaction gives us 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 (see Supplementary Materials); thus by default, we calculate and print overall model free energy in every flux distribution.

## Dry Cell Weight Measurements

Wild type *M. maripaludis S2* cells were grown in a chemically defined medium (See Supplementary Materials) using a 1-L chemostat under anaerobic conditions as described previously (11). The chemostat was operated in steady state continuous mode under H2-limiting conditions to match model simulation conditions, with gas flows of 10-15 mL/min H2, 40 mL/min CO2, 15 mL/min H2S, and a balance of N2 up to a total 200 mL/min.

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 {model number?}. After measuring chemostat optical density, we sampled 50 mL aliquots of cells in suspension were directly from chemostat culture and centrifuged samples at 7000 RPM for 15 minutes. 40 mL of supernatant were removed by pipette, then the 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.

## 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 (23). In practice, this requires measuring steady state growth rate in concert with an uptake rate or, in our case, a product secretion rate. 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 to ensure steady state at each data point. For each time 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.

To calculate ATP maintenance values in our model, we constrained our model to our measured growth rate and secretion rate at each time 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 (23). The resulting plot can be found in Supplementary Materials.

These growth data points comprised a set of 8 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. Growth yields were calculated as describe previously, but with our measured conversion between OD660 and cell density:

All simulations were performed using the default H2 + CO2 media formulation.

# 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 in our reconstruction 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 we began to curate it manually. Furthermore, we relied heavily on biochemical knowledge from literature sources, particularly regarding recently-elucidated biosynthesis pathways that were not available in annotation databases. Combined, our use of maximum likelihood gap filling and reliance on published literature sources resulted in more 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 we have genetic evidence that each of these reactions and the involved metabolites should be involved in metabolism, but we have not yet elucidated full synthesis or consumption pathways. They represent excellent candidates for further exploration of *M. maripaludis* metabolism, particularly as this reconstruction 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, including those without genes, fits into metabolism. Figure 1 shows a breakdown of these reactions lacking 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.

Our use of likelihood based gap filling not only directs us toward unknown portions of metabolism, but also lets us assign 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 novel metric of evaluating our confidence in the reconstruction. We can then use the scores allow us 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.

## Effects of Electron Bifurcation

It has been often assumed that the methanogenic pathway is linear with two main membrane complexes for generating membrane potential: the ion-translocating Mtr and a membrane-bound heterodisulfide complex (HdrDE). Without the membrane bound type heterodisulfide reductase of methylotrophic methanogens, *M. maripaludis* lacks one critical energetic coupling site to generate energy for growth. However, it has a cytoplasmic heterodisulfide reductase that conserves energy by an increasingly recognized form of energy conservation in strict anaerobes involving electron bifurcation (refs). As described above, this links the last step of methanogensis to the first step by exergonically driving the endergonic reduction of the 1st step resulting in a circular pathway (ref). This is noteworthy because the assumption of a linear pathway in *M. maripaludis* without accounting for electron bifurcation can affect the downstream predictions in the metabolic model.

Because it is a relatively novel mechanism of energy conservation, this heterodisulfide reduction step was not part of our annotation database and thus was not included in the automated first reconstruction draft. Rather, the default mechanism of energy conservation matched methylotrophic methanogens and utilized methanophenazine, an electron carrier known to be absent from *M. maripaludis* and other hydrogenotrophic methanogens. We replaced the methylotrophic pathway with the correct electron bifurcation pathway, linking heterodisulfide reduction with electrons from H2 to carbon dioxide reduction via reduced ferredoxin. This case serves as an example of a commonly-encountered reconstruction pitfall, in which information available in annotation databases does not sufficiently represent known biochemical phenomena. Therefore, it is paramount that even as automated reconstruction methods improve, we take the time to carefully evaluate biochemical literature maximize consistency of the reconstruction with experimentally-verified pathways.

To demonstrate the importance of ferredoxin reduction via bifurcation, we altered the existing bifurcating heterodisulfide reductase (HdrABC) reaction:

We removed ferredoxin from this reaction, balancing mass and charge to yield:

This scenario represented a case where *M. maripaludis* would not contain a membrane-bound HdrDE complex but could not perform electron bifurcation. We optimized our model for growth on CO2 + H2 using this model 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 be clearly attributed to disruption of the central energy conservation mechanism in *M. maripaludis*, in which electron bifurcation must necessarily reduce ferredoxin for reducing CO2. The only other complex capable of generating large amounts of reduced ferredoxin is the Eha hydrogenase, which utilizes a sodium ion gradient to reduce ferredoxin on a 1:1 basis. Similarly, CO2 reduction to methane requires reduced ferredoxin and pumps out sodium ions 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 also 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 (33).

Taking this analysis one step further, we used our reconstruction to probe into acetate assimilation, a supplemental pathway in *M. maripaludis* that can enhance but not fully support organism growth (41). This is in contrast to other methanogens such as *Methanosarcina barkeri*, which can subsist using solely the aceticlastic pathway (42). Though it is unknown exactly why *M. maripaludis* cannot be grown on acetate alone, our reconstruction did not reveal any strictly stoichiometric obstacle to growth. However, much like the pathway in *M. barkeri,* the aceticlastic pathway in *M. maripaludis* requires Eha/Ehb hydrogenase as a sink for reduced ferredoxin, thus thrusting this reaction into a central role rather than an anaplerotic one. Indeed, 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 the central energy-conserving role of bifurcation by limiting the magnitude of 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/Ehb cannot assume a central role in methanogenesis.

Interestingly, there is evidence that *M. maripaludis* uses multiple forms of ferredoxin as electron carriers and may link multiple steps, particularly those involved in electron bifurcation, using specific ferredoxins (43). 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 one type of specific ferredoxin for the Eha hydrogenase, Hdr, and formylmethanfuran dehydrogenase (Fwd) and a second type of specific ferredoxin for Ehb hydrogenase and biosynthetic carboxylating oxidoreductases, as suggested by {ref}. Using this function, tightens the coupling between the aforementioned reactions by restricting them all 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 virtually no effect on predicted fluxes but could have notable impact on gene knockout predictions, particularly those involving reactions that utilize ferredoxin.

## Other Biochemistry Improvements

A major part of our manual curation was adding biosynthesis pathways for the methanogenic coenzymes, sugars, and lipids. *M. maripaludis* utilizes various unusual coenzymes directly as electron carriers (methanofuran, H4MPT, coenzyme F420, coenzyme B, coenzyme M) and vital pieces of catabolic enzymes (coenzyme F430) during methanogenesis (5). It also synthesizes an archaellin tetrasaccharide as part of N-linked glycosylation (44) and multiple forms of archaeol, an archaeal membrane ether lipid (45). These synthesis pathways, particularly those for coenzymes, are vital pieces of *M. maripaludis* metabolism, hence we were adamant about including synthesis pathways for these metabolites and adding them to our biomass composition. Although several of these pathways were completely included in the Model SEED database, most were missing from the automated reconstruction draft and were added manually. Including these features in our reaction network and biomass definition distinguishes our reconstruction by incorporating multiple pathways that differentiate its metabolism from most other organisms.

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* because sulfate reduction produces sulfite, which inhibits methanogenesis (46). 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 the sulfate transporter and instead added a transporter for sulfide, the primary sulfur source for *M. maripaludis.* Our updated sulfur assimilation pathway includes a pathway for sulfide oxidation to sulfite, which is essential for multiple biosynthetic pathways, via a hypothesized dissimilatory sulfite reductase-like protein (17). 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 biochemical experts, we have created a reconstruction that is qualitatively consistent with accumulated biochemical knowledge of our organism.

## Model Validation

Evaluating the 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 (see 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 (47), but varied our dilution rate to gather a range of different yield measurements. Cell density was assessed using optical density (OD660) and was previously reported as OD660=1 corresponding to 0.34 mg(dry weight)∙ml-1 (34). We were unsure of the efficacy of this value, thus we re-measured the correlation using a combination of centrifugation and vacuum filtering (see Methods). Based upon our calibration curve (see Supplementary Information) we determined that OD660=1 corresponded to 0.49 mg(dry weight)∙ml-1.

Using our calibrated cell density value, we calculated 7 steady state methane evolution and growth rates during H2-limited growth. Prior to testing our model on this dataset, we also recognized that our model was essentially untrained in terms of ATP maintenance and contained automated values from our first draft reconstruction. Growth yield predictions can also vary considerably in response to model ATP maintenance energies (48), 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 (see Methods). As shown by Figure 2, using this method allowed us to essentially test our model’s growth yield predictions on 8 separate test points. Furthermore, our model consistently predicted growth yields within close range of measured values, particularly for lower growth rates (<0.08 h-1).

Following our LOOCV growth yield analysis, we used the full dataset to set growth associated maintenance (GAM) and non-growth associated maintenance (NGAM). 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* (49). 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 and 5.0 (mmol per grams [cell mass]), respectively. 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]), about 38% of our calculated value. This difference is reflective of the observed differences in growth yield for these organisms during hydrogenotrophic growth. Using the same formula for growth yield in each case at nearly identical doubling times of 12 h, *M. barkeri* reportedly grew at a yield of about 7.35 g(dry weight)∙mol CH4-1 and *M. maripaludis* grew at a yield of about 2.40 g(dry weight)∙mol CH4-1, or about 33%. Thus, our high GAM requirement reflects observed differences in growth data when comparing to a methylotrophic methanogen. 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. This result bodes well for our model’s utility as a predictive tool as we look to use it to generate quantitatively feasible growth hypotheses for novel strain designs.

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* (50). In the case of *M. maripaludis* transposon mutagenesis has been used to calculate an essentiality index of all genes in *M. maripaludis* (51) and provided a “first pass” test set for gene essentiality. This dataset contained 4 different libraries with gene essentiality predictions in defined media, thus genes were predicted to be essential in 0-4 cases. We tested against this set four times using different essentiality thresholds; e.g. “Case 4” counted a gene in the dataset as essential only if it was predicted as essential in all 4 libraries, “Case 1” counted a gene in the dataset as essential if it was predicted as essential in at least 1 library. The iMR540 reconstruction shared 542 genes with this dataset, thus we were able to compare gene essentiality predictions across nearly the entire model. As shown in Figure 3, our model’s predictive accuracy in the four cases ranged from 61.8-65.7% and our Matthew’s Correlation Coefficient (MCC) ranged from 0.289-0.331. Predictive accuracy was maximized in the “Case 3” dataset whereas MCC was highest for “Case 2”, a small discrepancy in metrics that reflects less frequent prediction of false negative outcomes when total negative outcomes are increased. Overall, these metrics show a slight positive correlation between essentiality index predictions and gene essentiality predictions from out model. However, essentiality index is disparate from gene knockout data, thus even though this dataset provided a broad assessment of our model’s ability to predict gene essentiality, it does not provide the same clear picture as actual knockout experiments.

For *M. maripaludis* there is relatively little data where gene knockout experiments have been systematically carried out *in vivo.* 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 (32–37). 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 3, 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, yet our model predicts these knockouts to be non-lethal. The reason for this disagreement lies in acetyl-CoA synthase (ACS), which ordinarily works with carbon monoxide dehydrogenase (CODH) to synthesize biosynthetic equivalents of acetyl-CoA from 5-Methyl-H4MPT using electrons from ferredoxin. During a 5- or 6-hydrogenase knockout, *M. maripaludis* loses its ability to generate anaplerotic ferredoxin from Eha/Ehb hydrogenase, but in iMR540 this is supplemented by the ACS-CODH complex working in reverse to reduce ferredoxin by oxidizing acetyl-coA. This reverse pathway is made possible by a predicted alternate acetyl-coA synthesis pathway from glycine, in which glycine can be converted to acetyl-coA through acetate. Notably, this alternate route cannot provide the same growth yield as the wild type strain but is predicted to be sufficient for preventing complete lethality. This alternate pathway, which was added from Kbase during the automated reconstruction process, provides an interesting target for further experiments and we are currently investigating its potential as an alternative acetyl-coA synthesis pathway.

## 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 (21). 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 (22). 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 overall biochemical “reaction” of a model, therefore it is reasonable to expect this overall reaction must produce a negative overall free energy to support growth. This additional constraint can be added by restricting 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 (see 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 (20). 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 (see 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 (30). 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 (52) and KEGG identifiers (24), enzyme commission numbers, and reaction subsystems where available. Each reaction and gene in the reconstruction is justified and connected to its literature and/or database source (see Supplementary Materials).

Additionally, we have sought to maximize usability of both our reconstruction and the resulting model. The systems biology markup language (SBML) is a standard medium for distributing metabolic reconstructions (53); thus, we have included our reaction network in SBML level 2, the highest version currently supported by the COBRA Toolbox (29). 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 (see Supplementary Materials). In the interest of transparency and to support the iterative process of reconstructing metabolic networks, we have also made our codes and reconstruction available on Github (https://github.com/marichards/methanococcus).

# Discussion

Metabolic reconstructions provide us with a computational approach to studying the complexity of genome-scale biochemistry. With iMR540, we have created the most comprehensive reconstruction of *M. maripaludis* currently available, the first manually-curated reconstruction built on top of likelihood based gap filling, and a straightforward new method to add thermodynamic constraints to metabolic models.

Our reconstruction process hinged first and foremost on literature sources, which provided high quality biochemical information. Through manual curation from these sources, we created a model with an accurate representation of the Wolfe Cycle, added pathways to synthesize methanogenic coenzymes, included archaellin and archaeol lipid biosyntheses, and corrected the sulfur assimilation pathway. The efficacy of rigorous manual curation is shown not only in the qualitative accuracy of our reconstruction, but also in the formidable quantitative capabilities of the resulting model.

This reconstruction is also, to our knowledge, the first genome scale metabolic network created using likelihood-based gap filling. Favoring this method over a maximum parsimony method helped us to automatically identify 66 more genes, increasing the gene coverage of our reconstruction prior to the start of manual curation. Simultaneously, we were also able to assign reaction likelihood scores for many of the reactions, lending a measure of confidence level for including each reaction in the network.

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.

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. In addition to distributing our model in SBML format, we have also included numerous functions for modifying the reconstruction, simulating growth, and predicting gene knockout lethality. By following this model of reconstruction distribution, we hope to encourage increased transparency and user-friendliness in all metabolic network reconstructions.

Along these lines, it is crucial to recognize that our reconstruction is by no means a finished product and must be expanded and improved in the future. It is a common fallacy to believe that any model, particularly one that focuses solely on one aspect of a biological system, can be relied upon to produce quantitatively accurate growth predictions, particularly when using a model as a platform for strain design. The iMR540 reconstruction contains many of the most crucial areas of metabolism and thus captures the majority of known metabolism, but there is far more about this organism that remains unknown and that therefore remains unpredictable through merely metabolic means. For example, *M. maripaludis* cannot grow on acetate alone, yet it can take up acetate and use it as a supplementary carbon source for growth if also given formate or H2+CO2. Through the metabolic lens of iMR540, we predict that *M. maripaludis* should be able to achieve acetotrophic growth, yet this plainly goes against empirical evidence to this point. This example illustrates two salient points as they relate to this and other reconstructions.

Firstly, we emphasize again that this reconstruction is a living network, meant not as a final product but as the first iteration of a malleable entity. The answer to biological quandaries such as the inability of *M. maripaludis* to grow on acetate alone may or may not lie in a purely metabolic model, but careful manual additions to our reconstruction can only enrich our understanding of metabolism and help us resolve how this system fits with transcriptional regulation and other systems within the cell. Secondly, we stress that quantitative model predictions should very rarely, if ever, be taken as absolutes in and of themselves. It is particularly tempting to examine predicted model flux distributions and to declare an individual distribution or range of distributions as what is expected out of the actual organism. In reality, a metabolic model’s power lies in predicting the scope of what we believe is metabolically possible, not what will actually happen in a living system. Thus, we expect that future modelers will our reconstruction and model as tools to generate testable experimental hypotheses and to catalyze elucidation of novel pathways in *M. maripaludis*. Above all, the goal of a metabolic network reconstruction is to create a realistic representation of organism metabolism that can then be used to push forward biological discovery and fuel innovations in strain design. With this reconstruction, we have contributed a tool to help understand the unique biochemistry of hydrogenotrophic methanogens and unlock their potential as metabolic engineering targets.

Taken together, this work demonstrates both the importance of doing extensive manual curation and the effectiveness of using automated gene homology based methods when reconstructing metabolic networks. Automated methods such as likelihood based gap filling are invaluable in their ability to quickly expand and enrich the reaction network, but literature from experimental studies still serves as the gold standard for adding new pieces to a reconstruction and verifying biochemical pathways. Above all, the goal of a metabolic network reconstruction is to create a realistic representation of organism metabolism that can then be used to push forward biological discovery and fuel innovations in strain design. With this reconstruction, we have contributed a tool to help understand the unique biochemistry of hydrogenotrophic methanogens 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. **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.

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

6. **Deppenmeier U**. 2002. The unique biochemistry of methanogenesis. Prog Nucleic Acid Res Mol Biol **71**:223–283.

7. **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.

8. **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.

9. **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.

10. **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.

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

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

13. **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.

14. **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.

15. **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.

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

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

18. **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.

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

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

21. **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.

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

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

24. **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.

25. **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.

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

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

28. **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.

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

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

31. **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.

32. **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.

33. **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.

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

35. **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.

36. **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.

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

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

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

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

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

42. **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.

43. **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.

44. **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.

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

46. **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.

47. **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.

48. **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.

49. **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**.

50. **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.

51. **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.

52. **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.

53. **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.

# 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: Comparing growth yield predictions on hydrogen to measured data using LOOCV (see Methods). All but two predicted growth rates fall within 5% of the measured values. The two outlying points are within 14% and 28% error, respectively, and may suggest a larger degree of unpredictability during more rapid growth.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 3: 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.