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

**Matthew A. Richards1,2, Juan Zhang3, Thomas J. Lie4, John A. Leigh4\* , 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

3Jiangnan University, China

4Department of Microbiology, University of Washington, Seattle, WA, USA

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

# Abstract

Methanogenic archaea naturally produce methane, a potent greenhouse gas and energy source, and hold promise as possible metabolic engineering targets for producing liquid fuels. *Methanococcus maripaludis* is a hydrogenotrophic methanogen that is notable for its genetic tractability, fast growth rate, and ability to grow in anaerobic chemostats. To better understand its metabolism and predict metabolic engineering strategies, we have built iMR544, a metabolic reconstruction of *M. maripaludis* that accounts for 544 of its 1722 protein-coding genes. Constructed using likelihood based gap filling and heavily rooted in biochemical literature, iMR544 incorporates flavin-based electron bifurcation into central carbon metabolism to accurately portray energy conservation in hydrogenotrophic methanogenesis. Our reconstruction incorporates vital biosynthetic pathways, including methanogenic coenzymes and archaeal lipids, and corrects inaccurate generic pathways, including sulfur assimilation and electron carrier usage. For model simulation, we have added thermodynamic considerations in a novel approach that calculates overall free energy without rigorously assessing individual reaction reversibility. This distribution also includes a toolbox for examining and modifying our reconstruction, running model simulations, and investigating flux distributions. Taken together, our reconstruction serves as a knowledgebase of *M. maripaludis* and provides a flexible platform to generate thermodynamically feasible strain design hypotheses.

# Introduction

Methane is a greenhouse gas that plays a critical role in the global carbon cycle and is 21 times more potent than carbon dioxide (1). However, it can also be used as a “green” energy fuel source that burns relatively cleanly compared to conventional fuels such as coal (2). Furthermore, there is growing interest in biologically converting of methane to high energy density liquid fuels (3). The greatest biological contributor of methane gas is a specialized group of anaerobic microorganisms known as methanogenic Archaea or methanogens, which produce about 1 Gt of methane gas per year through their metabolic activity (4). Methanogens employ unusual metabolic pathways with biologically unique cofactors and enzymes to produce methane from 1-C or 2-C carbon sources (5, 6). Studying methanogenesis gives us a window through which we can better understand these organisms and learn ways to metabolically engineer them to enhance methane production or produce commodity chemicals.

Though diverse, methanogens can be separated into two main groups based on the presence or absence of cytochromes (4). Metabolically, the cytochrome-lacking methanogens have a narrow substrate range limited to only H2 or formate as electron donors that reduce CO2 to methane; hence they are known as hydrogenotrophic methanogens. By contrast, methanogens with cytochromes utilize acetate and methylated compounds for methanogenic growth and a minority can also grow using H2 and CO2; they are known as methylotrophic methanogens. Although both groups have similar central pathways of CO2 reduction, they vary in their energy conservation approaches due to disparities in the number of coupling sites for generating ion gradients. For example, in the last step of methylotrophic methanogenesis, H2 reduces a heterodisulfide as part of a membrane bound hydrogenase-heterodisulfide complex that conserves energy via generation of an ion gradient (4, 7). The same complex in hydrogenotrophs, however, is cytoplasmic and thus does not provide a coupling site for the generation of an ion gradient (8). Instead, these methanogens conserve energy via flavin linked electron bifurcation, a non-traditional mechanism emerging in a variety of organisms (9) that couples the final heterodisulfide reduction step to the first step of CO2 reduction by using the latter as an electron sink for the former (10). The resulting circular pathway is known as the Wolfe Cycle (11) and is the key central catabolic pathway for understanding hydrogenotrophic methanogenesis.

*Methanococcus maripaludis* is an anaerobic hydrogenotrophic methanogen originally isolated from a salt marsh in 1983 (12). Its genome is comprised of only 1722 protein coding genes (13) and it has a relatively simple metabolism, using electrons from formate or hydrogen to reduce carbon dioxide to methane and build an ion motive force that drives ATP synthesis (4). *M. maripaludis* grows rapidly with a doubling time of 2 hours (12) and is genetically tractable (14), making it an ideal candidate for studying methanogenesis and for creating novel strain designs that produce industrially relevant products. Additionally, its established ability to thrive in anaerobic chemostats (15) allows for large scale steady state studies or for production of heterologous proteins, further solidifying its promise as a cell factory.

Genome scale metabolic reconstructions are powerful tools that map metabolic pathways and aid in cell factory development by serving as platforms for generating hypothetical strain designs (16). They are organism knowledge bases and can be made into constraint-based models that predict growth phenotypes for potential wet lab experiments by simulating steady-state growth via flux balance analysis (FBA) (17). Their valuable ability to represent metabolism has helped guide metabolic engineering efforts for production of industrial biochemicals in multiple organisms (16, 18). Constructing a genome scale metabolic reconstruction for *M. maripaludis* would therefore have promise for better understanding methanogenesis and for guiding metabolic engineering efforts that harness the unique metabolism of this hydrogenotrophic methanogen.

The genome of the *M. maripaludis S2* has been sequenced (13) and a transcriptome (19) as well as its proteome (20) has been published. Together with these previous systems-based efforts, a metabolic model would complete a multifaceted approach to understanding its unique biology as we could study selected pathways through biochemical or genetic approaches. *M. maripaludis* has already been metabolically reconstructed as part of a mutualistic community model with *D. vulgaris* (21)and as an isolate (22). 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 one organism’s metabolism (21). The latter case was the first genome-scale metabolic reconstruction of *M. maripaludis* (22), an important step towards understanding *M. maripaludis* metabolism.

Here we present iMR544, an updated genome scale metabolic reconstruction of *M. maripaludis* based on biochemical literature*.* With iMR544, we have sought to accurately represent the crucial Wolfe cycle, particularly the electron bifurcation step, in our metabolic reconstruction to help explain the ability for this organism to grow despite the lack of membrane related energy coupling sites. Other improvements include eliminating methanophenazine utilization, which is known not to occur in hydrogenotrophic methanogens (4), correctly sulfur assimilation by replacing sulfate with sulfide (23), and adding biosynthesis pathways for all of the unique coenzymes involved in methanogenesis (24). We increased genome coverage and homology by employing likelihood-based gap filling, a technique that fills reaction gaps based on probability rather than on parsimony (25). 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 (26). 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 (27, 28). 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, iMR544 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 (29) 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 (30–32); we chose the Department of Energy Systems Biology Knowledgebase (Kbase; www.kbase.us), a suite of tools that includes the Model SEED reaction database (32). 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 available through 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 must 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 (33). 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 (34). 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 (35) in MATLAB [7.14.0.739] (The MathWorks Inc., Natick, MA).

To encourage model transparency (36) 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 (37) 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 here in their current versions (see Supplementary Materials) with the full up-to-date set of 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 (35). Our experimental test set included 18 knockout genotypes across 4 different growth conditions, with 30 total wet lab experiments across these conditions (38–43). 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)(44):

|  |  |
| --- | --- |
|  | [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 (45) to all exchanges reactions for which these values could be calculated via the group contribution method (46). To incorporate these values into our reconstruction, we expanded the standard 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 (35). Our code accepts effective concentrations (mM) for specified exchange metabolites, assumes standard concentrations of 1 mM for the remaining 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 are model constructs that 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.

## Experimental 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 (15). Chemostats were in steady state continuous mode were operated with gas flows of 110 L/h H2, 15 L/h CO2, 15 L/h N2, and 15 L/h H2S, with a dilution rate of 0.0833 h-1.

We recalculated calibration curves for dry cell weight versus optical density by measuring dry cell weight via cell filtering and optical density via a UV/Vis spectrophotometer {model number?}. Small ~5 mL aliquots of cells sampled directly from chemostat culture were measured for optical density to determine the overall chemostat optical density. 50 mL aliquots of cells in media were then sampled directly from chemostat culture into 50 mL Falcon tubes and filtered through 25 nM pore filters to remove all non-cellular components. Wet cells and their filters were dried in a 50 degree oven and their weight was measured daily until it stabilized to give the final dry cell weight.

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 (29). In *M. maripaludis,* the ratio of ATP to methane production can be closely approximated as 0.5 , hence we constructed our plot by measuring methane production versus growth rate. Growth rates were monitored by measuring dry cell weight via optical density values. Gas from the chemostat headspace was collected directly into 5-mL serum vials after flushing with at least 500 mL of chemostat gas outflow. Methane production rates were quantitatively assessed using a {model name/number?} gas chromatograph to periodically measure methane concentration in coordination with cell growth rates. The resulting plot can be found in Supplementary Materials.

# Results

## Reconstruction Statistics

The basic statistics for iMR544 are displayed in Table 1 compared to iMM518, the previously published genome-scale metabolic reconstruction for *M. maripaludis* (22). Although our reconstruction was generated independently of the iMM518 reconstruction, many of the differences in models represent an expansion of the reaction network with regard to gene coverage and biochemistry. We slightly increased gene coverage by 26 genes, but because iMR544 was not built directly from the previous reconstruction and shares only 434 of its genes, we added 110 novel genes to our reconstruction and excluded 84 genes for which we did not find sufficient literature evidence. Despite a relatively modest increase in gene coverage, our network contains over 100 more gene-associated reactions and, as a result, over 90% of the internal reactions in our reconstruction are associated with at least one gene. We suspect that a major reason for our increase in 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 primarily 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.

Notably, when compared to iMM518 our reconstruction has 75 more internal reactions and over 100 more of both internal metabolites and dead-end metabolites that cannot be synthesized or consumed. Thus, although our reconstruction contains more metabolites and reactions, the size of the mathematical model generated by removing these dead ends is comparable to the previous model. These metabolites and their reactions may not be 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. Thus, we have 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 85 internal reactions that lack genes, many of which were added during the automated gap filling phase 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 all added to fill biosynthesis gaps but have no supporting literature evidence. In total, of the 85 reactions lacking genes 46 (54%) are gap filling reactions and the remaining 39 (46%) are classified as “hypothetical”, with no known genes currently but with literature evidence pointing to their inclusion in the reconstruction. We expect that as experimental research groups uncover more biochemical phenomena, many reactions in this latter group will become gene-associated whereas the gap filling reactions, much like dead end reactions and metabolites, point us toward areas of metabolism that are poorly understood in our organism and require more in depth investigation.

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.

## Biochemistry Improvements

The most noteworthy addition to our reconstruction was including flavin based electron bifurcation to complete the Wolfe Cycle, resulting in what is, to our knowledge, the first accurate depiction of central carbon catabolism in hydrogenotrophic methanogenesis. 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 incident 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.

Interestingly, there is evidence that *M. maripaludis* uses multiple forms of ferredoxin as electron carriers and may link multiple steps, including electron bifurcation, using specific ferredoxins (47). 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 changes promiscuous ferredoxins to specific ferredoxins for the Eha hydrogenase, heterodisulfide reductase, and formylmethanfuran dehydrogenase (carbon dioxide reduction). Using this function, which also provides a specific ferredoxins with the ability to act as promiscuous ferredoxins, 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.

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 (48) and multiple forms of archaeol, an archaeal membrane ether lipid (49). These synthesis pathways, particularly those for the 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, many reactions were missing and nearly all of the reactions were added manually after automated reconstruction. Including these features in our reaction network and biomass definition distinguishes our model 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 (50). However, because sulfate is the default sulfur source for most microorganisms, our initial 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 (23). Taken together with additional coenzyme 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 in a team of experienced metabolic modelers and 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 useful 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, our comparison was limited to two conditions: H2-limiting and formate-limiting. Previous work yielded YCH4 (grams [cell mass] per mole of CH4 produced) values of 2.86 ± 0.58 for H2 limitation and 2.31 ± 0.58 for formate limitation (51). These experimental yields were based on a previously-reported conversion factor between optical density and dry cell weight with A600 of 1 corresponding to 0.34 mg(dry weight)∙ml-1 (40). We were unsure of the accuracy of this value and to mitigate our concerns, we re-measured this conversion factor (see Methods). We used our newly-measured value of #VALUE to calculate YCH4 values of 2.86 ± 0.58 for H2 limitation and 2.31 ± 0.58 for formate limitation.

Growth yield predictions can also vary considerably in response to model ATP maintenance energies (52). Our default biomass equation obtained through Kbase specified growth associated maintenance (GAM) as 40.11 (mmol per grams [cell mass]) and the first draft model specified no non-growth associated maintenance (NGAM). Though these figures gave what we considered to be reasonable growth yield predictions, we chose to measure GAM and NGAM ourselves with respect to methane production rate (see Methods). Following our growth experiments, we altered these values to #VALUE# and #VALUE# in H2-limiting conditions and #VALUE # and #VALUE# in formate-limiting conditions for GAM and NGAM, respectively.

Using our recalibrated growth yields and ATP maintenance values, we compared our model’s FBA predictions of maximum growth yields to the experimentally-derived measurements as shown in Figure 2. Both of our computational values agreed closely with the experimental values, falling within the uncertainty range in each case. Though growth yield validation is not an absolute measure of model performance, our model’s ability to closely reproduce experimental results without manual 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 provide 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* (53). In the case of *M. maripaludis* transposon mutagenesis has been used to calculate an essentiality index of all genes in *M. maripaludis* (54), but 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 (38–43). 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 had regulatory bases. In these cases, knockouts of 5 or 6 hydrogenases are lethal in formate-grown cells with downregulation of the glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) cycle. When the GAPOR cycle is upregulated (as in Δ6H2asesupp and Δ7H2asesupp), it provides anaplerotic electrons that allow cell growth in formate media, even when missing up to 7 hydrogenases. Upregulation of the GAPOR cycle is a regulatory difference, and we have chosen to leave the cycle upregulated as a default to allow for a richer spectrum of possible flux distributions. Thus, although our default model does not predict knockout growth phenotypes with 100% accuracy, it can easily be tuned to reflect regulatory differences.

## 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 (27). 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 (28). 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 (26). 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 Distribution

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 (36). 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 KEGG identifiers (30), 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 (55); thus, we have included our reaction network in SBML. 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 iMR544, 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.

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. **Zhang X**, **Myhrvold NP**, **Caldeira K**. 2014. Key factors for assessing climate benefits of natural gas versus coal electricity generation. Environ Res Lett **9**:114022.

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

9. **Buckel W**, **Thauer RK**. 2013. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. Biochim Biophys Acta BBA - Bioenerg **1827**:94–113.

10. **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.

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

12. **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.

13. **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.

14. **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.

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

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

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

18. **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.

19. **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.

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. **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.

22. **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.

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

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

25. **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.

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

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

28. **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.

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

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

31. **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.

32. **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.

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

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

35. **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.

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

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

38. **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.

39. **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.

40. **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.

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

42. **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.

43. **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.

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

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

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

47. **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.

48. **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.

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

50. **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.

51. **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.

52. **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.

53. **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.

54. **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.

55. **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 comparison** | | |
| Model | iMM518 | iMR544 |
| Protein Coding Genes | 518 | 544 |
| % ORF Coverage | 30 | 32 |
| Intra/Extracellular Metabolites | 556/49 | 665/53 |
| Dead End Metabolites | 163 | 265 |
| Internal Reactions | 570 | 645 |
| Exchange Reactions | 49 | 59 |
| Gene-Associated Reactions | 464 | 590 |
| % Reactions Associated with Genes (non-exchange) | 75 | 91 |
|  | | |

Table 1: A comparison between iMR544 and iMM518 indicates that our model covers slightly more of the genome, including over 125 more gene-associated reactions. Our model includes more internal reactions and metabolites, as well as approximately 100 more dead end metabolites. Though these metabolites represent the portion of metabolism that cannot carry flux, all of our model's dead end metabolites are part of gene-associated reactions and thus represent promising avenues for future model expansion

Figure 1: A chart showing broad subsystem groupings of the 85 reactions in iMR544 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 our models’ growth yield predictions on hydrogen and formate, respectively, to experimental data we found that our predictions fell within the experimental error of the measured values. This comparison demonstrates that our models can reproduce these values, suggesting that the models and the reconstruction they were derived from can reliably predict growth yields are quantitatively consistent with *in vivo* results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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.