A Genome Scale Metabolic Reconstruction of *Methanococcus maripaludis S2* that Accurately Depicts Hydrogenotrophic Methanogenesis

**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 are a crucial part of the global carbon cycle, producing about 1 billion tons of methane each year. We have constructed a genome-scale metabolic model for the model archaeon *Methanococcus maripaludis S2* that is the first model to accurately portray hydrogenotrophic methanogenesis*.* Our model contains the complete Wolfe cycle, the central catabolic pathway of our organism, including the crucial electron bifurcation step that completes the cycle. This model serves as a knowledgebase of *M. maripaludis* metabolism and provides a platform for generating hypotheses for strain designs.

# Introduction

Methane is a vital part of the global carbon cycle; it functions both as a greenhouse gas an order of magnitude more potent that carbon dioxide [1] and as a fuel source. Methanogenic archaea, or “methanogens”, naturally produce about 1 GT of methane per year [2] by undergoing methanogenesis, the process by which simple carbon substrates are reduced to methane gas. Studying the unique metabolism of these organisms gives us a window through which we can better understand the biochemistry of methanogenesis such that we can metabolically engineer ways to enhance methane production or manipulate metabolism to produce other chemicals.

*Methanococcus maripaludis S2* is an anaerobic hydrogenotrophic methanogen originally isolated from a salt marsh in 1983 [3]. Its genome is comprised of only 1722 protein coding genes [4] and its spectrum of major growth substrates is restricted to either hydrogen and carbon dioxide or formate [3], resulting in a relatively simple metabolism where electrons from formate or hydrogen reduce carbon dioxide to methane and build an ionic motive force that drives ATP synthesis {SOURCE NEEDED}. In addition to its uncomplicated metabolism, *M. maripaludis* possesses a well-developed set of genetic tools {SOURCE NEEDED}, and grows readily on solid media [5], in liquid batch cultures [3], and in liquid chemostat cultures [6]. Thus, it represents an ideal candidate for studying methanogenesis and for creating novel strain designs that produce industrially relevant products.

Genome scale metabolic reconstructions are powerful tools that can aid in mapping metabolic pathways and serve as platforms for generating hypothetical strain designs. Additionally, they serve as organism knowledge bases and can be made into models that predict growth phenotypes for potential wet lab experiments by simulating steady-state growth via flux balance analysis (FBA) {SOURCE NEEDED}. Their valuable ability to model metabolism has helped guide metabolic engineering efforts for production of industrial biochemical in multiple organisms [7]. 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 our hydrogenotrophic methanogen.

*M. maripaludis* has already been metabolically reconstructed as part of a mutualistic community model with *D. vulgaris* [8]and as an isolate [9]. 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 [8]. The latter case was the first genome-scale metabolic reconstruction of *M. maripaludis S2*[9]and although it was an important step towards understanding *M. maripaludis* metabolism, our examination of this model revealed a number of areas we could improve upon by incorporating additional features from published biochemical literature.

Here we present iMR533, the first metabolic reconstruction and model of *M. maripaludis S2* to accurately represent hydrogenotrophic methanogenesis. This reconstruction is the first to include the full Wolfe Cycle, the cylical catabolic pathway responsible for generating methane and ATP in our organism [10]. Most critically, our reconstruction includes the vital energy-conserving electron bifurcation step that couples the first and last steps of the cycle [11]. Other improvements include eliminating methanophenazine utilization, which is known not to occur in hydrogenotrophic methanogens [2], replacing sulfate with sulfide, the primary sulfur source in *M. maripaludis* {SOURCE NEEDED}, and adding biosynthesis pathways for all of the unique coenzymes involved in methanogenesis [12]. We further increased genome coverage and homology by employing likelihood-based gapfilling, a technique that fills reaction gaps based on probability rather than on parsimony [13]. Our reconstruction is the first manually-curated genome scale reconstruction to employ likelihood based gap filling.

As an additional measure, we have included free energies of formation to our reconstruction. This is an especially salient consideration for methanogenic archaea, which can grow close to the thermodynamic limits that support life [14]. A well-established method of applying thermodynamic constraints involves applying the second law of thermodynamics to metabolic models to restrict reaction directionalities in the direction of negative free energy generation [15,16]. Rather than apply thermodynamic constraints to every metabolic reaction, we devised a novel method of adding free energies of formation to our model to predict the overall free energy generated during steady state growth based solely on standard free energies and concentrations of external metabolites. Our method allows us to evaluate and restrict our FBA solutions to only those that have a negative net free energy of formation, but does not bound the solution space by placing reaction directionality constraints on internal reactions where metabolite concentrations are difficult to measure. In combining these novel thermodynamic considerations with stoichiometric information, iMR533 provides a means to predict energetically feasible strain designs, enhancing our metabolic engineering capabilities with *M. maripaludis.*

# Methods

## Genome Scale Reconstruction

We built the first draft of our reconstruction using the Model SEED automated reconstruction tools in the Department of Energy Systems Biology Knowledgebase (Kbase; http://kbase.us). We built our reconstruction and gapfilled it to create a model in Kbase using their model-building tools. Chief among these is our likelihood-based gapfilling approach, which maximizes gene homology as it fills gaps in the model. We used the default Kbase biomass definition for gram-negative bacteria.

Then we expanded and refined the model by manually adding information from literature sources. Our final model has each reaction tagged for how it was added to the model and what evidence was used to justify its inclusion. It is crucial to include this information to uphold network and model transparency [17]. The model itself can be found in SBML and Matlab structure format in the supplementary materials. A current version of the model can be found on Github (marichards/methanococcus)

## Growth Simulations

To simulate growth, we use the steady state assumption (Sv=0) and defined bounds on our fluxes (vmin ≤ v ≤ vmax). All model simulations were performed using the COBRA toolbox 2.0 [18] in MATLAB [7.14.0.739] (The MathWorks Inc., Natick, MA). It is vital that metabolic reconstruction efforts remain transparent and that resulting models be easily usable by other groups [17]. An important part of building a model is ensuring that it is easy to simulate. To assist in this process, we have created functions specifically written to simulate model growth, make systematic changes to the model, and replicate all of the results reported in this manuscript. These codes are included in supplementary data and current versions can be found on Github (https://github.com/marichards/methanococcus)

maxGrowthOn\_\_ codes

simulateKOPanel code

switchToFormate code

switchToSpecificFerredoxins code

Others available on GitHub

## Gene Knockout Phenotype Simulations

We performed gene knockout simulations using our “simulateExpKOPanel.m” function (see Supplementary Materials), which utilizes the COBRA Toolbox 2.0 [18] for MATLAB. In this series of simulationsWe incorporated knockout predictions across six Experimental gene knockout data phenotypes were pFor gene knockouts, we evaluated our model’s performance using the Matthews correlation coefficient (MCC)[19]. The formula for the MCC is given as follows:

## Thermodynamic Calculations

In a novel approach, we added free energies of formation from the Equilibrator database [20] to all exchanges reactions for which these values could be calculated via the group contribution method [21]. We expanded the standard structure of the model to include free energy values for a standard 1 mM concentration, temperature of 25 C, pressure of 1 bar, pH of 7, and ionic strength of 0.1 M. Our “optimizeThermoModel.m” code (See Supplementary Materials) allows us to specify concentrations of exchange metabolites, alter these standard free energies to reflect differences in concentration, and estimate the overall free energy generated during growth by multiplying each metabolite’s free energy of formation by its exchange flux. The model can additionally be constrained by setting an upper bound on permissible values of overall free energy generation; this scenario would disallow FBA solutions that do not meet minimal energy required for steady state cell growth.

## Dry Cell Weight Measurements

We grew *M. maripaludis S2* cells in a chemically defined media (See Supplementary Materials) using a 1-L chemostat under anaerobic conditions. We operated the chemostat in continuous mode 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 measured dry cell weight via cell filtering. 100 mL aliquots of cells in media were filtered through 25 nM pore filters to remove all non-cellular components. The wet filters were then dried in a 50 degree oven and their weight was measured daily until it stabilized, giving the final dry cell weight.

# Results

## Model Statistics

Our model stats are displayed in Table 1A and, as shown, it compares favorably to the existing model. Our gene coverage is slightly better, but perhaps more importantly, over 90% of the non-exchange reactions in our model are gene associated. This suggests that our model has more consistent ties to gene homology as a direct result of using our likelihood-based gapfilling method and of maximizing our reliance on biochemical knowledge from literature.

Notably, our model has nearly 100 more internal metabolites and over 100 more dead-end metabolites that cannot be synthesized or consumed by the model. Although these metabolites and their reactions are not part of our mathematical model, we have included them in our reconstruction because they are all gene-associated [should we have a “reconstruction” separate from the “model”?]. Thus, we have evidence that each of these metabolites should be involved in metabolism, but we have not yet elucidated their synthesis or consumption pathways. They represent excellent candidates for further exploration of MM metabolism, particularly as this model is updated and expanded in the future.

The most important distinction between our model and the existing model is that ours accurately depicts methanogenesis in the form of the Wolfe Cycle. Unlike the other model, we include the vital electron bifurcation step discovered in 2012 that completes the cycle by connecting methane production to the first step of the pathway via electron carriers. The other group also includes other errors that appear to be the result of basing their model primarily off general annotations from the KEGG database. Including these errors, such as the inclusion of sulfate as the primary sulfur source and of methanophenazine as a major electron carrier, demonstrates the need for rigorous manual curation and working directly with an expert in the organism’s biochemistry. By employing these methods, we have avoided these and other errors, resulting in a model that is more consistent with accumulated biochemical knowledge of our organism.

The likelihood based gapfilling resulted in the automated addition of 66 genes to our reconstruction before we began manually curating. The likelihood scores themselves also provide a novel metric of evaluating our confidence in the model because each gapfilled reaction is annotated with a confidence score ranging from 0-1. These scores allow us to quickly hone in on reactions with low gene homology as possible targets for more experimental investigation.

## Growth Yield Predictions

A major quantitative function of a metabolic model is predicting organism growth yield on a variety of substrates. Thus, comparing model predictions of growth yields to experimentally-determined values is a common way of evaluating model accuracy. *M. maripaludis* has a narrow range of possible substrates, restricting our system’s test cases to two conditions: H2-limiting and formate-limiting. For these scenarios, we used raw optical densities from a previous study [22] but chose to re-measure the relationship between dry cell weight and optical density (see Methods). These experimental yields were determined based on optical density and converted to dry cell weight, but we had reason to believe our conversion factor may have been incorrect. To guard against this possibility, we re-measured dry cell weight versus optical density as described in Methods. We recalculated the previously-reported values using our new conversion factor and determined that the experimental growth yields were # and # on H2 and formate, respectively. We compared these yields to our computational predictions, as shown in Figure 1, and found that our computational values fell (or didn’t fall) within close range of the experimental values. We didn’t hit the values on the nose, but we’re not worried because aiming to do so would lead to overfitting.

## Gene Knockout Validation

At its core, constraint-based modeling is concerned with taking genotype information and annotating those genes as metabolic reactions in order to predict growth phenotypes. Among the best ways to evaluate the predictive ability of a model is to compare model predictions of gene knockout growth phenotypes with experimental data. Though one group used transposon mutagenesis to assess gene function of all genes in *M. maripaludis* [23], there is relatively little data where gene knockout experiments have been systematically carried out *in vivo.* For our model, we were able to assemble a knockout panel of 30 genotype/media combinations across 6 previous publications. These genotypes consisted primarily of hydrogenase knockouts in central carbon metabolism and thus, they give us a good idea of how well our model can predict knockouts in central catabolism. In comparing with these data, we achieved 90% accuracy across all conditions and an overall Matthew’s correlation coefficient of 0.67. This high value suggested that our model is an excellent predictor of growth phenotype based on genotype changes in central carbon metabolism. It was particularly encouraging because we purely tested our model on these data; we did no fitting based on the knockout validation set.

# Discussion

We’ve created the highest-quality model of M. maripaludis currently available with emphasis on using manual curation and likelihood-based gapfilling to maximize gene homology and biochemical accuracy. This is the first metabolic model to accurately depict the Wolfe cycle, the vital central carbon pathway in hydrogentrophic methanogens.

The main conclusions of the paper are:

1. We made the best available model of M. maripaludis by leveraging biochemical literature and using likelihood-based gapfilling increased our gene homology
2. We’ve made all the information from the model and codes to do growth simulations available, which helps with model transparency
3. Our model includes a new way of estimating free energy generation that is vital when dealing with a methanogen, an organism that lives close to the limit of thermodynamic feasibility

This model represents the first manually curated model that was constructed with likelihood-based gapfilling, at least to our knowledge. The likelihood scores lend an element of accountability to our gapfilling, but we’ve also strived for accountability making our decisions explicit throughout the curation process (this is a tie-in with Ben’s paper).

We expect to use our model as a tool to make predictions for how to metabolically engineer our organism and to generate hypotheses regarding unknown portions of M. maripaludis metabolism.

# Acknowledgements

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

# References

1. Milich L. The role of methane in global warming: where might mitigation strategies be focused? Glob Environ Change. 1999;9: 179–201. doi:10.1016/S0959-3780(98)00037-5

2. Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol. 2008;6: 579–591. doi:10.1038/nrmicro1931

3. Jones WJ, Paynter MJB, Gupta R. Characterization of Methanococcus maripaludis sp. nov., a new methanogen isolated from salt marsh sediment. Arch Microbiol. 1983;135: 91–97. doi:10.1007/BF00408015

4. Hendrickson EL, Kaul R, Zhou Y, Bovee D, Chapman P, Chung J, et al. Complete Genome Sequence of the Genetically Tractable Hydrogenotrophic Methanogen Methanococcus maripaludis. J Bacteriol. 2004;186: 6956–6969. doi:10.1128/JB.186.20.6956-6969.2004

5. Jones WJ, Whitman WB, Fields RD, Wolfe RS. Growth and Plating Efficiency of Methanococci on Agar Media. Appl Environ Microbiol. 1983;46: 220–226.

6. Haydock AK, Porat I, Whitman WB, Leigh JA. Continuous culture of Methanococcus maripaludis under defined nutrient conditions. FEMS Microbiol Lett. 2004;238: 85–91. doi:10.1111/j.1574-6968.2004.tb09741.x

7. Milne CB, Kim P-J, Eddy JA, Price ND. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. Biotechnol J. 2009;4: 1653–1670. doi:10.1002/biot.200900234

8. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. Mol Syst Biol. 2007;3: 92. doi:10.1038/msb4100131

9. Goyal N, Widiastuti H, Karimi IA, Zhou Z. A genome-scale metabolic model of Methanococcus maripaludis S2 for CO2 capture and conversion to methane. Mol Biosyst. 2014;10: 1043–1054. doi:10.1039/c3mb70421a

10. Thauer RK. The Wolfe cycle comes full circle. Proc Natl Acad Sci. 2012;109: 15084–15085. doi:10.1073/pnas.1213193109

11. Kaster A-K, Moll J, Parey K, Thauer RK. Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proc Natl Acad Sci. 2011;108: 2981–2986. doi:10.1073/pnas.1016761108

12. Graham DE, White RH. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nat Prod Rep. 2002;19: 133–147. doi:10.1039/B103714P

13. Benedict MN, Mundy MB, Henry CS, Chia N, Price ND. Likelihood-Based Gene Annotations for Gap Filling and Quality Assessment in Genome-Scale Metabolic Models. PLoS Comput Biol. 2014;10: e1003882. doi:10.1371/journal.pcbi.1003882

14. Jackson BE, McInerney MJ. Anaerobic microbial metabolism can proceed close to thermodynamic limits. Nature. 2002;415: 454–456. doi:10.1038/415454a

15. Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-Based Metabolic Flux Analysis. Biophys J. 2007;92: 1792–1805. doi:10.1529/biophysj.106.093138

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

17. Heavner BD, Price ND. Transparency in metabolic network reconstruction enables scalable biological discovery. Curr Opin Biotechnol. 2015;34: 105–109. doi:10.1016/j.copbio.2014.12.010

18. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6: 1290–1307. doi:10.1038/nprot.2011.308

19. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta BBA - Protein Struct. 1975;405: 442–451. doi:10.1016/0005-2795(75)90109-9

20. Flamholz A, Noor E, Bar-Even A, Milo R. eQuilibrator—the biochemical thermodynamics calculator. Nucleic Acids Res. 2011; gkr874. doi:10.1093/nar/gkr874

21. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V. Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks. Biophys J. 2008;95: 1487–1499. doi:10.1529/biophysj.107.124784

22. Costa KC, Yoon SH, Pan M, Burn JA, Baliga NS, Leigh JA. Effects of H2 and Formate on Growth Yield and Regulation of Methanogenesis in Methanococcus maripaludis. J Bacteriol. 2013;195: 1456–1462. doi:10.1128/JB.02141-12

23. Sarmiento F, Mrázek J, Whitman WB. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. Proc Natl Acad Sci. 2013;110: 4726–4731. doi:10.1073/pnas.1220225110