Outline for MM model Paper

* Introduction/Background
  + Methane is a vital part of the global carbon cycle; it functions both as a potent greenhouse gas and as a fuel source. Methanogens make about 1 GT of methane per year, providing us with a biological gateway to methane production.
  + Among methanogens, Methanococcus maripaludis S2 is a model organism because it possesses a fast doubling time, grows readily in a chemostat, and has a well-developed set of genetic tools. These qualities make it an ideal candidate for studying and engineering methanogenesis.
  + Metabolic models are powerful tools that serve as organism knowledge bases and can be simulated to predict growth phenotypes for potential wet lab experiments. They have promise for guiding metabolic engineering efforts such as harnessing the unique energy metabolism of our hydrogenotrophic methanogen.
  + There is an existing model of M. maripaludis published in 2014, but that model does not accurately reflect the Wolfe cycle, the central catabolic pathway that is vital to methanogen metabolism. We have constructed the first metabolic model for M. maripaludis to accurately represent this central pathway, thus providing a solid computational platform to generate metabolic engineering designs. We employed likelihood-based gapfilling to build our model, increasing the gene homology in our model and making it the first manually-curated model constructed with our likelihood-based gapfilling approach.
* Methods
  + 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. 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)
  + 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 in Matlab.
  + We added free energies of formation to all exchanges for which these values could be calculated via group contribution using the Equilibrator tool (citation). Free energy values in the model are for a standard 1 mM concentration, temperature of 25 C, pressure of 1 bar, pH of 7, and ionic strength of 0.1 M. Given concentrations of exchange metabolites, we estimated t overall free energy generated during growth by multiplying each metabolite’s free energy of formation by its exchange flux.
  + 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 (marichards/methanococcus)
    - maxGrowthOn\_\_ codes
    - simulateKOPanel code
    - switchToFormate code
    - switchToSpecificFerredoxins code
    - Others available on GitHub
  + We conducted MM growth experiments using a 1-L anaerobic chemostat. Standard media recipe was McN (reference recipe resource), a chemically-defined medium. Chemostat was operated in continuous mode with dilution rate of 1 h.
  + 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.
  + For gene knockouts, we evaluated our model’s performance using the Matthews correlation coefficient (defined at source). The formula for the MCC is given as follows:

(FORMULA)

* Results
  + 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.
  + A common way of evaluating a metabolic model is comparing growth yield predictions to experimentally-determined values. Due to the narrow range of possible substrates for our system, our comparison was limited to two conditions: H2-limiting and formate-limiting. 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.
  + Growth yield validation is great, but we also want to know about how well our model can predict phenotype based on genotypes. The standard way to measure this is by comparison with experimental gene knockout data. As in the case of growth data, there is not an abundance of knockout data for MM, but 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.
  + 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.

Figure 1. Comparison of predicted and experimental growth yields. At the moment, our growth predictions don’t really match, but we think that’s more a function of poor growth measurements than of our model’s prediction capabilities. We’re still waiting on the controllers to show up, but I anticipate being able to re-measure these numbers next month using the a microfiltering system to measure optical density vs. dry cell weight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| KO Genes | 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** |
| **Total Correct:** | **10 of 10** | **14 of 16** | **2 of 2** | **1 of 2** | | **27 of 30** |
|  |  |  |  |  |  | |
| Figure 2. Knockout lethality predictions from FBA and agreement with experimental results. Our model achieves 90% agreement with experimental results for central catabolic knockouts, corresponding to a Matthew’s Correlation Coefficient of 0.67. | | | | | | |

|  |  |  |
| --- | --- | --- |
| ***Methanococcus maripaludis S2* model comparison** | | |
| Model | iMM518 | iMR529 |
| Protein Coding Genes | 518 | 529 |
| % ORF Coverage | 30 | 31 |
| Intra/Extracellular Metabolites | 556/49 | 650/52 |
| Dead End Metabolites | 163 | 268 |
| Internal Reactions | 570 | 571 |
| Exchange Reactions | 49 | 57 |
| Gene-Associated Reactions | 464 | 571 |
| % Reactions Associated with Genes (non-exchange) | 75 | 91 |
| Table 1A. A comparison between iMR533 and iMM518 indicates that our model covers slightly more of the genome, including over 100 more gene-associated reactions. Both models include approximately the same number of reactions, but our model has approximately 100 more internal metabolites and dead end metabolites. Though this 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. | | |
|
|
|
|
|