METABOLIC CONTROLS ON NITROGEN FIXATION IN ANAEROBIC METHANE-OXIDIZING CONSORTIA & NOVEL EXPLORATIONS OF MICROBIAL RATE LAW IDENTIFICATION

by

KYLE PATRICK CRAIG

(Under the Direction of Christof Meile)

ABSTRACT

Most elemental cycles in Earth's near-surface environment, including marine sediments, are mediated by microbial reactions. To quantify these transformations and predict the distribution of chemical species, biogeochemical models need to capture the transport and reaction rates of these processes. Reactions of this nature often depend on environmental conditions, including the availability of substrates, as well as the microbial community present. Rate expressions then often involve numerous model parameters that may be site-specific or poorly constrained. To address this, we describe a novel approach to derive rate expressions and their parameterization using symbolic regression. In addition, we present model developments at the microscale to characterize nitrogen utilization and growth efficiency calculation in a microbial aggregate, comprised of methanotrophic archaea (ANME) in syntrophic association with sulfate-reducing bacteria (SRB). We demonstrate model agreement with observed growth efficiencies when using ammonium and investigate potential triggers for N_2 fixation activity in these consortia.

INDEX WORDS: Microbially-mediated, reaction rates, anaerobic methane oxidation, nitrogen fixation, reactive-transport model, flux balance analysis, growth efficiency, symbolic regression, biogeochemistry

METABOLIC CONTROLS ON NITROGEN FIXATION IN ANAEROBIC METHANE-OXIDIZING CONSORTIA & NOVEL EXPLORATIONS OF MICROBIAL RATE LAW IDENTIFICATION

by

KYLE PATRICK CRAIG

BS, University of Georgia, 2019

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2023

METABOLIC CONTROLS ON NITROGEN FIXATION IN ANAEROBIC METHANE-OXIDIZING CONSORTIA & NOVEL EXPLORATIONS OF MICROBIAL RATE LAW IDENTIFICATION

by

KYLE PATRICK CRAIG

Major Professor:	Christof Meile
Committee:	Adrienne Hoarfrost
	Sergei Katsev

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

ACKNOWLEDGEMENTS

This body of work would not be possible without the guidance of my advisor, Christof Meile, as well as my committee members Adrienne Hoarfrost and Sergei Katsev. I would also like to thank Ranjani Murali at Caltech for her always helpful discussions on the biochemistry of anaerobic methane oxidizers and sulfate reducers. I need to further thank Chris Henry and his group at Argonne National Laboratory for providing and assisting with the flux balance analysis models. In addition, I thank my lab mates, Kadir Biçe and Shahram Asgari, for their help and friendship throughout my graduate studies. Finally, a big thanks to my friends, family, and my partner Emma for their continuous love and support while conducting this research.

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Sciences Program under award number DE-SC0016469 and DE-SC0020373 to Christof Meile.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Our global environment is facing challenges yet unseen by much of history, causing major impacts on biotic and abiotic factors for the ecosystems that make up the planet. These include, for example, the availability of freshwater for agriculture, distribution of pollutants, changes to sea level rise, and storm intensity – all of which will vastly alter the shape of human society in the generations to come. Addressing these factors requires knowledge of the underlying systems and processes that control the behavior of our planet. Using Earth System Models (ESM) is one such way that we can better understand the interactions between our physical climate and biogeochemical processes, which can inform the biological processes and elemental transformations occurring therein (Flato 2011; Steffen et al. 2020).

Given that these models resolve spatial scales in the km range, they make for poor estimates when dealing with problems occurring in natural systems at the micro- and meso-scales. Many processes occurring at important interfaces are not resolved in these models (Ward et al. 2020), and are typically dependent on biological and chemical data that are lacking in quality (Flato 2011). Furthermore, microbial activity has also been difficult to resolve in ESMs (Wieder et al. 2015). Thus, broadening our understanding of the relationships between microbial processes and biogeochemical elemental cycles will have profound effects on our predictive capabilities for a rapidly changing climate.

Microbial processes play a vital role in altering environmental conditions, especially in the subsurface where geochemical substrate turnover is often dependent on mediation by microorganisms (Lovley & Chapelle 1995). The fate of organic matter deposited at the sediment surface is controlled primarily through a series of microbial reactions, causing remineralization of many C, N, and P compounds (Blackburn & Blackburn 1993). Within marine sediments specifically, the chemical composition of the overlying water and the burial rate of organic C compounds, in addition to the availability of dissolved oxidative species, dictate which microbial regime participates in the breakdown (Aller 2014). The sequential consumption of electron acceptors (e.g., O_2, NO_3^-, MnO_2, $Fe(OH)_3$, SO_4^{2-}, CO_2) is colloquially referred to as the redox ladder, which reflects the differences in free energy changes resulting from organic matter decomposition (Froelich et al. 1979; Stumm & Morgan 2012).

It is clear then that the reaction rates for microbially-mediated biogeochemical processes are strongly influenced by the local spatial and temporal distribution of chemical substrates and metabolic capabilities. Capturing the dynamic process rates in a model requires deep insight into chemical and biological parameters and their relationships. Rate expressions often encompass effective rate constants, for example, a parameter k' that contains a measure of the abundance of functional microbial populations (Meile and Scheibe 2019). Given that environmental conditions tend to change over time and space, however, parameters like k' are site-specific and often require intensive fine-tuning to match observed behaviors. To explore the challenges faced when describing reaction rates in models containing macroscopic biogeochemical processes, I employ the machine learning approach of symbolic regression (SR; Cranmer 2023) to infer the

relationships between chemical conditions and volumetric rates to build and parameterize expressions. This is the topic of Chapter 2, in which I investigate the use of SR to identify the nature and parameterization of rate expressions from co-registered concentration and rate measurements.

While it is true that the microbial processes we observe in subsurface environments often control macroscopic biogeochemical cycling, the physiology of microorganisms should also be considered as a potential major component of their activity. This is especially the case for organisms that live at the cusp of energetic availability, where intricate cellular processes such as electron bifurcation, syntrophic partnerships, and reverse electron transport aid in keeping the cell functioning (Peters et al. 2016; Schink 1997; Wheelis 1984). Investigations on functional microorganisms surviving under low-energy conditions will enhance our understanding of the fate of carbon after burial, including in the form of methane. This potent greenhouse gas is responsible for 20% of the warming in our climate, making it the second highest contributor behind CO_2 (Lyu et al. 2018). Microbial production of methane in the deep marine biosphere is estimated to be 4-25 Tg C yr^{-1} (Wallmann et al. 2012); however, this is almost entirely offset by the anaerobic oxidation of methane in marine sediments (Reeburgh 2007).

Anaerobic methanotrophic archaea (ANME) catalyze the oxidation of methane using, for example, ferric minerals independently (Sivan et al. 2011) or, most commonly, in a syntrophic consortium with partner-bacteria reducing sulfate, nitrate, or metal oxides (Boetius et al. 2000; Raghoebarsing et al. 2006; Leu et al. 2020). The mechanism by which the separate species share electrons was long thought to be through diffusible

intermediates (Stams & Plugge 2009); although, modeling efforts demonstrated that diffusible electron carriers are likely to accumulate, leading to conditions unsuitable for methane oxidation at the observed rates (Sørensen et al. 2001; Orcutt & Meile 2008). More recent evidence has confirmed that aggregates containing ANME and sulfate-reducing bacteria (SRB) transfer electrons directly using conductive nanowires (McGlynn et al. 2015; Wegener et al. 2015). This direct interspecies electron transfer (DIET) mechanism was later added to a reactive-transport model of an ANME-SRB consortium and was able to sustain similar rates and thermodynamic yields as seen in experiments (He et al. 2019 & 2021). Despite coupling their metabolic reactions to achieve thermodynamic favorability, the growth rates of these aggregates are very slow, with doubling times on the order of several months (Nauhaus et al. 2007; Orphan et al. 2009).

Even with their extremely constrained growth rates, ANME have been observed to fix nitrogen under certain conditions (Dekas et al. 2009; Dekas et al. 2018; Metcalfe et al. 2021). The motivation for using this energy-intensive reaction pathway is still unknown. Using a spatially resolved reactive-transport model informed by thermodynamic reactions, in Chapter 3 I explore the growth dynamics and N usage strategies of an ANME-SRB consortium at the aggregate scale. Then, I also investigate the metabolic pathways of an ANME-2b cell in more detail using a flux balance analysis (FBA) model. Using these two modeling approaches, my work aims to shed light on the potential conditions promoting diazotrophic activity in a globally important, slow-growing benthic microbial group.

References

Aller, R. C. (2014). Sedimentary diagenesis, depositional environments, and benthic fluxes.

Blackburn, T. H., & Blackburn, N. D. (1993). Rates of microbial processes in sediments. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, *344*(1670), 49-58.

Boetius, A., Ravenschlag, K., Schubert, C. J., Rickert, D., Widdel, F., Gieseke, A., ... & Pfannkuche, O. (2000). A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature*, *407*(6804), 623-626.

Cranmer, M. (2023). Interpretable machine learning for science with PySR and SymbolicRegression.jl. *arXiv preprint arXiv:2305.01582*.

Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-sea archaea fix and share nitrogen in methane consuming microbial consortia. *Science*, *326*(5951), 422-426.

Dekas, A. E., Fike, D. A., Chadwick, G. L., Green-Saxena, A., Fortney, J., Connon, S. A., ... & Orphan, V. J. (2018). Widespread nitrogen fixation in sediments from diverse deep-sea sites of elevated carbon loading. *Environmental Microbiology*, *20*(12), 4281-4296.

Flato, G. M. (2011). Earth system models: an overview. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 783-800.

Froelich, P., Klinkhammer, G. P., Bender, M. A. A., Luedtke, N. A., Heath, G. R., Cullen, D., ... &
Maynard, V. (1979). Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: suboxic diagenesis. *Geochimica et Cosmochimica Acta*, *43*(7), 1075-1090.

He, X., Chadwick, G., Kempes, C., Shi, Y., McGlynn, S., Orphan, V., & Meile, C. (2019). Microbial interactions in the anaerobic oxidation of methane: model simulations constrained by process rates and activity patterns. *Environmental Microbiology*, *21*(2), 631-647.

He, X., Chadwick, G. L., Kempes, C. P., Orphan, V. J., & Meile, C. (2021). Controls on interspecies electron transport and size limitation of anaerobically methane-oxidizing microbial consortia. *Mbio*, *12*(3), 10-1128.

Leu, A. O., Cai, C., McIlroy, S. J., Southam, G., Orphan, V. J., Yuan, Z., ... & Tyson, G. W. (2020). Anaerobic methane oxidation coupled to manganese reduction by members of the Methanoperedenaceae. *The ISME Journal*, *14*(4), 1030-1041.

Lovley, D. R., & Chapelle, F. H. (1995). Deep subsurface microbial processes. *Reviews of Geophysics*, *33*(3), 365- 381.

Lyu, Z., Shao, N., Akinyemi, T., & Whitman, W. B. (2018). Methanogenesis. *Current Biology*, *28*(13), R727-R732.

McGlynn, S. E., Chadwick, G. L., Kempes, C. P., & Orphan, V. J. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature*, *526*(7574), 531-535.

Meile, C., & Scheibe, T. D. (2019). Reactive transport modeling of microbial dynamics. *Elements: An International Magazine of Mineralogy, Geochemistry, and Petrology*, *15*(2), 111-116.

Metcalfe, K. S., Murali, R., Mullin, S. W., Connon, S. A., & Orphan, V. J. (2021). Experimentally-validated correlation analysis reveals new anaerobic methane oxidation partnerships with consortium-level heterogeneity in diazotrophy. *The ISME Journal*, *15*(2), 377-396.

Nauhaus, K., Albrecht, M., Elvert, M., Boetius, A., & Widdel, F. (2007). In vitro cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate. *Environmental Microbiology*, *9*(1), 187-196.

Orcutt, B., & Meile, C. (2008). Constraints on mechanisms and rates of anaerobic oxidation of methane by microbial consortia: process-based modeling of ANME-2 archaea and sulfate reducing bacteria interactions. *Biogeosciences*, *5*(6), 1587-1599.

Orphan, V. J., Turk, K. A., Green, A. M., & House, C. H. (2009). Patterns of 15N assimilation and growth of methanotrophic ANME-2 archaea and sulfate-reducing bacteria within structured syntrophic consortia revealed by FISH-SIMS. *Environmental Microbiology*, *11*(7), 1777-1791.

Peters, J. W., Miller, A. F., Jones, A. K., King, P. W., & Adams, M. W. (2016). Electron bifurcation. *Current Opinion in Chemical Biology*, *31*, 146-152.

Raghoebarsing, A. A., Pol, A., Van de Pas-Schoonen, K. T., Smolders, A. J., Ettwig, K. F., Rijpstra, W. I. C., ... & Strous, M. (2006). A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature*, *440*(7086), 918-921.

Reeburgh, W. S. (2007). Oceanic methane biogeochemistry. *Chemical reviews*, *107*(2), 486-513.
Schink, B. (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews*, *61*(2), 262-280.
Sivan, O., Adler, M., Pearson, A., Gelman, F., Bar-Or, I., John, S. G., & Eckert, W. (2011). Geochemical evidence for iron-mediated anaerobic oxidation of methane. *Limnology and Oceanography*, *56*(4), 1536-1544.
Sørensen, K. B., Finster, K., & Ramsing, N. B. (2001). Thermodynamic and kinetic requirements in anaerobic methane oxidizing consortia exclude hydrogen, acetate, and methanol as possible electron shuttles. *Microbial Ecology*, *42*, 1-10.
Stams, A. J., & Plugge, C. M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Reviews Microbiology*, *7*(8), 568-577.
Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H. J., Dube, O. P., Dutreuil, S., ... & Lubchenco, J. (2020). The emergence and evolution of Earth System Science. *Nature Reviews Earth & Environment*, *1*(1), 54-63.
Stumm, W., & Morgan, J. J. (2012). *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*. John Wiley & Sons.
Wallmann, K., Pinero, E., Burwicz, E., Haeckel, M., Hensen, C., Dale, A., & Ruepke, L. (2012). The global inventory of methane hydrate in marine sediments: A theoretical approach. *Energies*, *5*(7), 2449-2498.
Ward, N. D., Megonigal, J. P., Bond-Lamberty, B., Bailey, V. L., Butman, D., Canuel, E. A., ... & Windham-Myers, L. (2020). Representing the function and sensitivity of coastal interfaces in Earth system models. *Nature Communications*, *11*(1), 2458.
Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E., & Boetius, A. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature*, *526*(7574), 587-590.
Wheelis, M. (1984). Energy conservation and pyridine nucleotide reduction in chemoautotrophic bacteria: a thermodynamic analysis. *Archives of Microbiology*, *138*, 166-169.
Wieder, W. R., Allison, S. D., Davidson, E. A., Georgiou, K., Hararuk, O., He, Y., ... & Xu, X. (2015). Explicitly representing soil microbial processes in Earth system models. *Global Biogeochemical Cycles*, *29*(10), 1782-1800.

CHAPTER 2

IDENTIFYING MACROSCOPIC RATE EXPRESSIONS USING SYMBOLIC REGRESSION

Introduction

Biogeochemical cycling is driven in large part by the microbial groups catalyzing reactions using available substrates (Lovley & Chapelle 1995). Their impact on the global carbon cycle cannot be overstated; in marine sediments, recently deposited organic matter is remineralized, releasing C, N, and P sources and stimulating primary productivity in the water column (Van Cappellen and Ingall 1994). In aerobic mineralization, the breakdown of organic carbon sources also affects the pH within the porewater, dissolving carbonates and releasing other inorganic C sources that feed the carbon cycle (Hales & Emerson 1996). Microbial groups also oxidize this organic matter to generate energy for maintenance and biomass synthesis (Blackburn & Blackburn 1993). Modeling these microbially-mediated biogeochemical processes is an important step in furthering our insight into natural systems.

A typical formulation for microbially-mediated reaction rates (R) in reactive transport problems is given by,

$$R_i = k_i B F_{k,i} F_{T,i} \qquad \textit{Eq. 2.1}$$

where k_i is a cell-specific rate constant, B is the cell density of the microbial population conducting reaction i, $F_{k,i}$ represents the kinetic dependence of reaction i on substrate concentrations (e.g., $\frac{C}{K_{m,i} + C}$), and the thermodynamic factor $F_{T,i}$ ($0 \leq F_{T,i} \leq 1$) describes

the need for sufficient availability of free energy from reaction *i* to fuel ATP synthesis and maintain cell homeostasis (Jin & Bethke 2005; 2007). Calculating these rates requires insights into cell-specific reaction rates (k_i) and the abundance of active microbial functional groups, which in many instances is expressed by an effective parameter, $k' = k*B$, that captures a maximum volumetric rate, as well as the kinetic rate expression (F_k) often described using Monod kinetics with a half-saturation constant (Jin & Bethke 2007). Model parameters are often poorly constrained, which poses a central challenge to the application of biogeochemical models. The parameters of (semi-) empirical rate laws are often either derived from laboratory experiments (Eppley et al. 1969; Daebel et al. 2007) or determined through optimization approaches minimizing the difference between observed and modeled concentrations (Fennel et al. 2001). Such efforts rely on knowledge of the functional relationship between environmental conditions, including substrate concentrations, and the reaction rates.

Furthermore, parameterization difficulties often lie is in the quantification of the active microbial functional groups *B*. This can be costly and labor-intensive (Harris & Kell 1984), making it often not a part of many biogeochemical data sets involving microbially-mediated reactions. The lack of quality data on active microbial functional groups makes constraining rate expressions that are dependent on biomass difficult to model. In addition, such parameters are often assumed to constant, despite dynamic changes in the abundance and activity of the functional microbial groups. This raises the question if reaction rates can be related to accessible entities, such as measured and/or modeled concentrations, and how functional relationships between concentrations and rates, mediated via a dynamic microbial community, can be expressed.

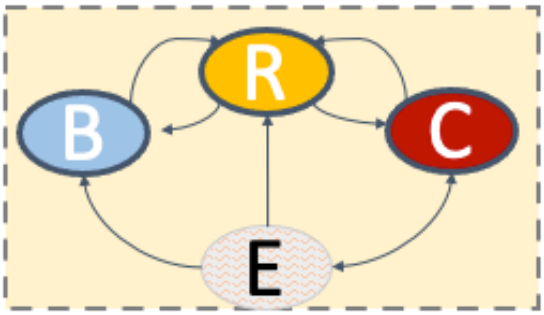

Fig 2.1. Conceptual framework relating concentrations (C), reaction rates (R), cell abundance of functional microbial groups (B) and Gibbs free energies (E).

Here, we present the use of symbolic regression (SR) to infer not only the parameterization but also the functional kinetic relationship between chemical conditions and rates. In addition to this, data sets are rarely collected that encompass concentrations, rates, and microbial dynamics. Hence, we explore approaches that approximate microbial dynamics from the chemical conditions, which are typically represented in biogeochemical models. We do so using a conceptual framework (Figure 2.1) that connects rates with the chemical (C) and biological (B) conditions, modulated via the reaction energetics (E). There is first the rate-concentration relationship (C $\rightarrow$ R), where substrate concentrations drive reaction rates (as reflected in F_k) and the rates affect the concentrations of products and reactants. In the rate-biomass relationship (B $\rightarrow$ R), metabolic rates catalyzed by microbes fuel cell growth, while increased cell numbers elevate the volumetric reaction rates. Furthermore, there is the interplay between energy, biomass, concentrations and rates (E, B, C $\rightarrow$ R), where the Gibbs Free energy of reaction depends on the activities (and hence concentrations) of substrates and products, while the availability of useable energy can impact the reaction rate (as reflected in F_T). In addition, a large gain of energy at the cell level allows for the generation and allocation of resources, stimulating cell-specific rates (Wu et al. 2022). It has further been

shown that the Gibbs free energy change for catabolic (ΔG_{cat}) and anabolic (ΔG_{ana}) reactions, together with the energy dissipated as heat during anabolism (ΔG_{dis}), can be used to calculate the growth yield for a microbial organism (Heijnen et al. 1992; Heijnen & Kleerebezem 2010; González-Cabaleiro et al. 2015). Thus, we hypothesize that volumetric (macroscopic) rates can be linked not only to the substrate availability but also to the Gibbs free energy of reaction (ΔG_{rxn}), through the growth yield that leads to a microbial community dominated by fast-growing organisms.

To address the relationships described above, we explored and analyzed reaction rates in the context of concentration fields and reaction energetics. Concentration and reaction rate data were selected from characteristic marine sedimentary environments that allow for calculation of the ΔG_{rxn}. The data was generated using reaction transport models that provide in situ chemical conditions and estimates of reaction rates. With these data, patterns were then explored through hypothesized relationships between the reaction rates and ΔG_{rxn}.

Methods

Symbolic regression

We opted to explore genetic programming (GP) algorithms that perform SR on our modeled datasets to reconstruct rate expressions. GP works by taking an initial population of models that are selected for fitness in a tournament, applying random mutations and crossovers on the fittest individual, and then creating a new generation using the mutated individual, a mechanism analogous to a Darwinian evolution (Augusto & Barbosa 2000). SR offers an interpretable machine learning approach that applies

mathematical operators to a set of data to find the best fit for the target. This results in functional relationships between input and target variables rather than establishing black-box models like those obtained with, for example, neural networks. Here, we use the symbolic regression package PySR (Cranmer 2023) to derive rate expressions from known rates and environmental concentrations of numerous chemicals.

PySR uses an iterative algorithm that performs genetic programming to optimize a loss function and find a solution. While the model itself is built in Python (3.11.0), the actual computation is offloaded into the programming language Julia. This allows for high-performance search power that substantially decreases overall calculation time for symbolic expressions (Cranmer 2023). To build the model, the population size and number of iterations (i.e., generations) that are run must first be set. The algorithm also requires a set of operators to be used in the construction of functions and are split into two categories by default within the PySR regressor: binary, which includes simple mathematical operators such as addition, subtraction, multiplication, and division, and unary, which encompass complex functions such as trigonometric and logarithmic operators with a single input argument. This gives a wide range of options to consider when selecting which data to input into the regressor. This function set can be expanded upon with the use of custom functions, requiring them to be written in Julia syntax. Through applying specifically defined functions to the operator set, we can more effectively steer the algorithm in a desired direction using expert knowledge, rather than allowing it to wander and consume computational resources. In the application here, for example, we define operators for the Monod-like terms that dictate substrate dependence on the rate expression.

PySR will apply constants between a range of -1 and 1 by default. While this range can be changed, one may also opt to use the appropriate units so that the numerical parameter values fall into the default range. Another useful tool are the constraints that can be set to only allow a defined amount of complexity to be used. One application of this functionality is setting constraints on the nesting of operators, as the algorithm has been observed to often apply operators to the same feature multiple times. Model parameters used for the results shown below are given in the Appendix. The number of input features that can be selected is limited to 10. Trying to add in more will slow down computation and may require dimensionality reduction to address which features ultimately matter.

Generation of test data

1D early diagenetic model. Initial explorations with SR were aimed at assessing whether the model could return a known rate law using only concentration fields as the input data. This data was generated by STEADYSED1, a steady-state early diagenesis reactive-transport model that simulates redox chemical conditions in surficial marine sediments (Wang & Van Cappellen 1996). The sites used for model parameterization are located in the eastern Skagerrak at water depths between 190 to 695 m. Sediment porewater profiles used for the model are described in prior field studies (Canfield et al. 1993). Physical parameters, bottom water chemistry, and other model parameters for the results used here are described in the Appendix.

To test the capabilities of PySR on reconstructing the correct rate expression as well as finding appropriate parameter values, we provided the SR with a target rate profile, and a range of concentration profiles and operators. We focused on the secondary

redox reactions, such as the oxidation of reduced iron using dissolved O_2 to produce iron oxide, whose reaction rate is proportional to both Fe^{2+} and O_2:

$$Fe^{2+} + \frac{1}{4}O_2 + 2HCO_3^- + \frac{1}{2}H_2O \rightarrow Fe(OH)_3 + 2CO_2 \quad \textit{Rxn. 2.1}$$

$$R_{FeOx} = k_{FeOx}[Fe^{2+}][O_2] \quad \textit{Eq. 2.2}$$

Chemostat. To discern whether SR can further identify the correct dependencies of rate law parameters for microbially-mediated reactions that explicitly depend on biomass, data generated in a chemostat system was used as the input for training the model. The chemostat model featured a simple one-population system where biomass and chemical concentrations and biomass-dependent reaction rates were calculated and used as data to fit the PySR regressor. The model is characterized by the consumption of C and N sources by a population of denitrifying bacteria:

$$CH_3COO^- + 1.6\,NO_3^- + 0.6\,H^+ \rightarrow 0.8\,N_2 + 2\,HCO_3^- + 0.8\,H_2O \quad \textit{Rxn. 2.2}$$

We used four different reaction rate formulations of increasing complexity (R_1-R_4; Eqns. 2.3-2.6), ranging from linear dependencies on substrate and biomass concentrations to Monod kinetics with thermodynamic constraints, to investigate what the algorithm was capable of reproducing:

$$R_1 = kB[Ac][NO_3^-] \quad \textit{Eq. 2.3}$$

$$R_2 = kB\frac{[Ac]}{K_m^{Ac} + [Ac]}\frac{[NO_3^-]}{K_m^{NO_3^-} + [NO_3^-]} \quad \textit{Eq.2.4}$$

$$R_3 = kB[Ac][NO_3^-]F_T \quad \textit{Eq. 2.5}$$

$$R_4 = kB\frac{[Ac]}{K_m^{Ac} + [Ac]}\frac{[NO_3^-]}{K_m^{NO_3^-} + [NO_3^-]}F_T \quad \textit{Eq. 2.6}$$

where k is the cell-specific rate constant imposed at a value of $1x10^{-5}$ for each rate, B is the microbial cell concentration, K_m^x is the half-saturation constant for substrate x. The units of k depend on the reaction; for R_1 and R_3 (Eqs. 2.3 and 2.5 respectively), the units are in m^6 mol^{-1} $cell^{-1}$ s^{-1} while for R_2 and R_4 (Eqs. 2.4 and 2.6), they become simply mol $cell^{-1}$ s^{-1}. F_T is the thermodynamic drive for the reaction ($0 \leq F_T \leq 1$; Jin & Bethke 2005; 2007):

$$F_T = \max\left(0, 1 - exp\left(\frac{\Delta G_{rxn} - m\Delta G_{ATP}}{\chi R_{gas} T}\right)\right) \quad \textit{Eq. 2.7}$$

where ΔG_{rxn} is the Gibbs free energy change of the catabolic reaction, ΔG_{ATP} is the energy required to form ATP (50 kJ mol^{-1}_{ATP} (Schink 1997)), m is the proportion of protons translocated per reaction to the number required for ATP synthesis (e.g., 2/3 for denitrification; Jin & Bethke 2005), and χ is the number of electrons transferred per reaction. The governing equations for the model take the form of:

$$\frac{dAc}{dt} = s_{Ac} R_x + \frac{Q}{V} in_{Ac} - \frac{Q}{V} Ac \quad \textit{Eq. 2.8}$$

$$\frac{dNO_3^-}{dt} = s_{NO_3^-} R_x + \frac{Q}{V} in_{NO_3^-} - \frac{Q}{V} NO_3^- \quad \textit{Eq. 2.9}$$

$$\frac{dB_{dnf}}{dt} = \gamma_{dnf} R_x + \frac{Q}{V} in_B - \frac{Q}{V} B_{dnf} \quad \textit{Eq. 2.10}$$

where s_{Ac} and $s_{NO_3^-}$ refer to the stoichiometry of the consumption for acetate and nitrate respectively, γ_{dnf} is the growth yield for denitrifying bacteria in the system in mol_{BM} mol_{Ac}^{-1}, R_x is the reaction rate as formulated in Eqs. 2.3-2.6 in units of mol_x m^{-3} s^{-1}, Q is the flow rate in m^3 s^{-1}, V is the volume in m^3, in_j is the inflow concentration of species j, and B_{dnf} is the cell concentration of denitrifying bacteria in cell m^{-3}. The model was run at increasing dilution rates until wash-out of the microbial population occurred, and rates,

biomass and concentrations were recorded over time for analysis using SR. R_x, B_{dnf} and the substrate concentrations represent a simple dynamic system where the microbial population achieves maximal reaction rates prior to the flow removing them from the system. The chemostat model was implemented in MATLAB R2022a (The MathWorks, Inc., Natick, MA, USA), and the code as well as the initial conditions for the model are attached in Appendix A. Simulation results over time for each of the four rate laws described above were then exported as a dataset containing the substrate concentrations (e.g., Ac and NO_3^-), the microbial cell density B, the thermodynamic factor F_T, and the flow rate Q.

Results and Discussion

1D early diagenetic model. Reaction rates in the source model take the form of first-order rate equations, where the rate law is defined by a rate constant in units of M^{-1} yr^{-1} (Wang & Van Cappellen 1996). SR investigations for identifying the rate law for the iron oxidation reaction shown in Eq. 2 were conducted using the depth profiles of the chemical species involved in the target reaction (*Rxn 1*), i.e. O_2, Fe^{2+}, $Fe(OH)_3$ and HCO_3^-. In our initial test, multiplication was the only operator considered, in line with the rate law in the reactive transport model used to generate the dataset – a restriction loosened below. Using these input variables and some tuning of the parameters in the model, SR was able to correctly reproduce the target rate data (Figure 2.2).

The model not only selected the correct dependent variables from the data set provided, but it also found the correct constant ($k^{PySR} = 0.112$ µmol cm_{total}^{-3} yr^{-1} $(\mu M)^{-2}$) to the data which matches the rate constant used in the STEADYSED model ($k_{FeOx} = 1.4 \times 10^8$ M^{-1} yr^{-1}; see Appendix A for the unit conversion illustrating the equivalence of the two constants).

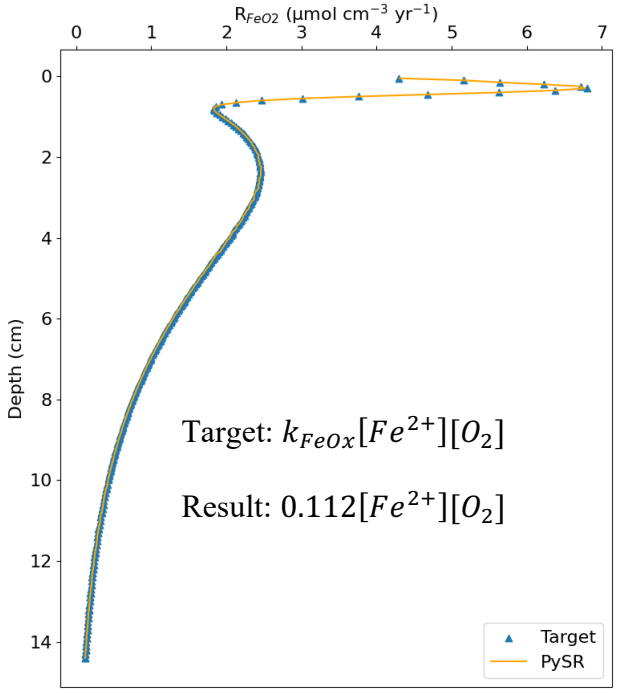

Fig 2.2. PySR simulation targeting the iron oxidation rate from STEADYSED using multiplication as the operator and the chemical species involved in *Rxn 2.1.*

While PySR tends to perform better when given a constrained set of operators and lower feature dimensionality, further experimentation revealed it is rather robust at selecting the correct variables when presented with extraneous choices of operators and features. In a model run where the target was again the rate of iron oxidation, the input variables were expanded to include the depth profiles for all 21 simulated chemicals in the STEADYSED model as well as each of the binary operators (e.g., addition, subtraction, multiplication, and division). The resulting expression came out to be exactly the same as that shown in Figure 2.2, requiring no extra parameterization of the SR

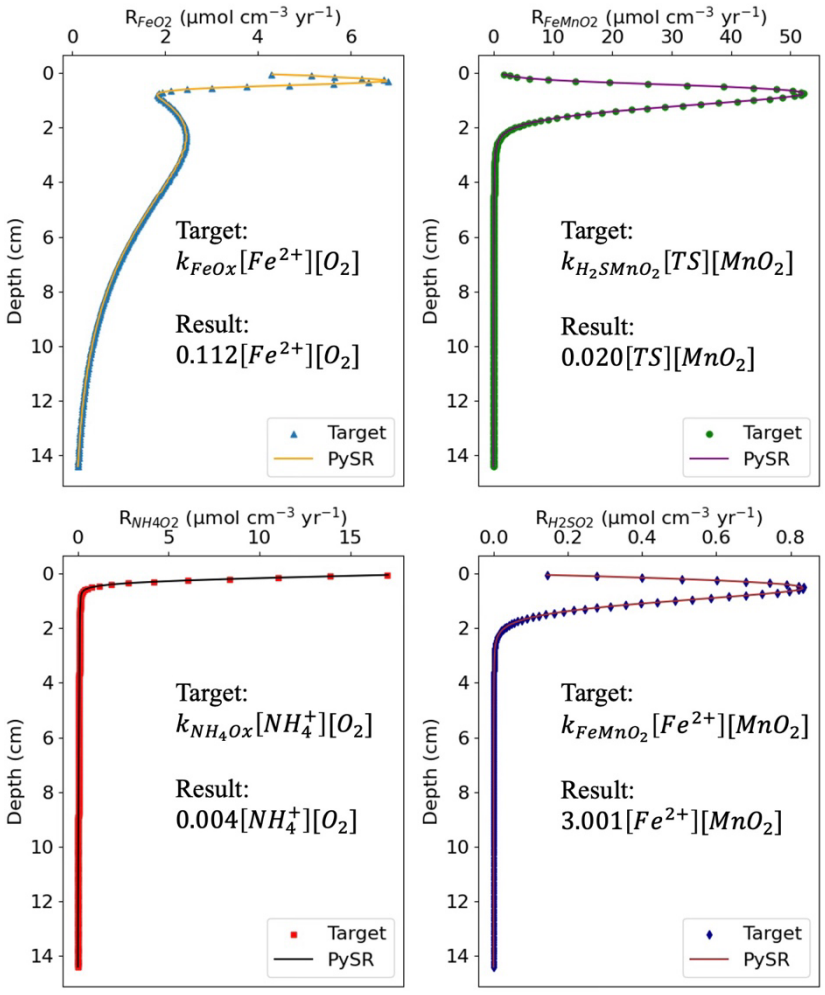

Fig 2.3. PySR simulation targeting rates of iron oxidation with O_2 and MnO_2, sulfide oxidation with MnO_2, and nitrification. Feature variables included all chemical profiles modeled in STEADYSED. Operator used was multiplication. [TS] = total dissolved sulfide concentration (in units of µmol/L)

algorithm. The application of SR to multiple processes was also successful (Figure 2.3), even when using all of the chemical concentrations as features and the binary operators. This simulation returned the correct rate expressions for each of the four targeted reactions (Figure 2.3), performing nearly identically to a simulation where we constrained the features to only the products and reactants of each reaction.

Chemostat. The initial SR model runs were conducted using each of the necessary variables for the rate law as shown in Eqs. 2.3-2.6, leaving the rate constant k and the rate law itself for prediction. These features were the concentrations of Ac and NO_3^-, the biomass concentration B, and the thermodynamic factor F_T. Rate laws R_1 and R_3 (Eqs. 2.3 and 2.5 respectively) were the first to be targeted from the chemostat simulations as they have a concentration-dependence similar to the rate expressions in the 1D early diagenetic model but with the additional dependency of biomass B and the

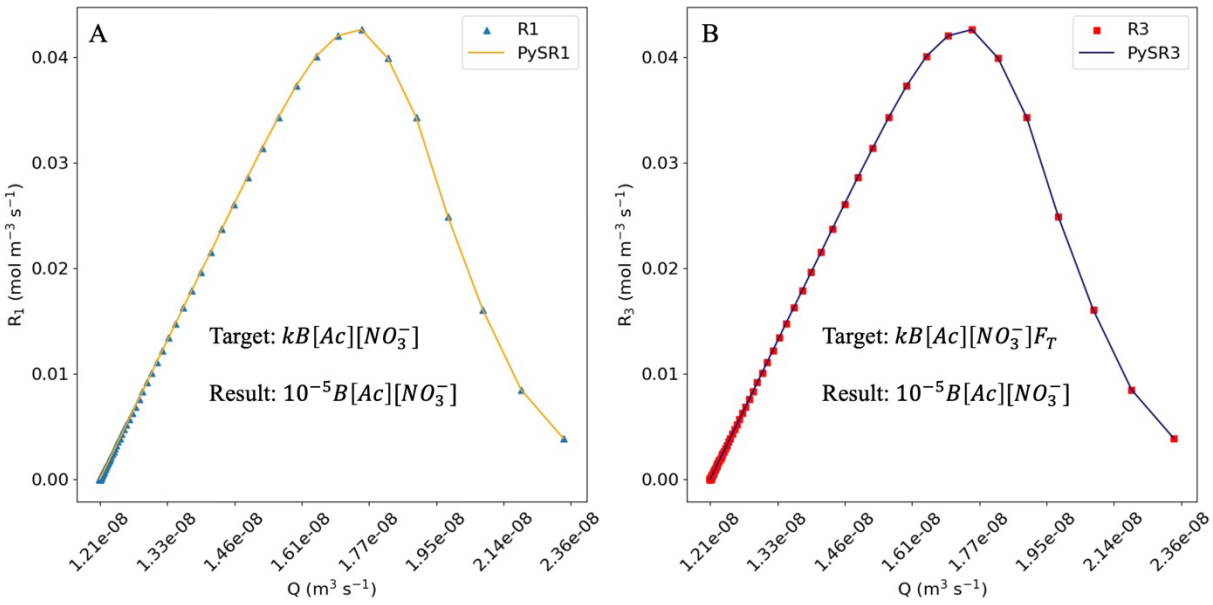

Fig 2.4. Results for rate law reconstruction of concentration-dependent rate expressions from our denitrifier chemostat model. R_1 and R_3 are the source rate expressions (e.g., Target) and PySR1 and PySR3 correspond to the regression expressions generated by the algorithm (e.g., Result).

thermodynamic factor F_T. Even with the addition of these new parameters, PySR was able to reconstruct the rate expressions for R_1 and R_3 including the correct rate constant (Figure 2.4). The rate law generated by PySR for R_1 is identical to the source expression used to calculate the rates in the chemostat (Figure 2.4A). The expression generated for R_3, however, leaves out the thermodynamic dependency of the reaction (Figure 2.4B). This is because the value of F_T is 1 for the entire simulation period, and the PySR algorithm chose not to include it to reduce the complexity of the expression.

For the remaining two rate expressions in the chemostat model R_2 and R_4, the concentration dependency of the reaction rate comes in the form of a Monod relationship, where a half-saturation constant K_m controls the rate of nutrient uptake (Levenspiel 1980). This relationship requires the use of both the division and addition operators,

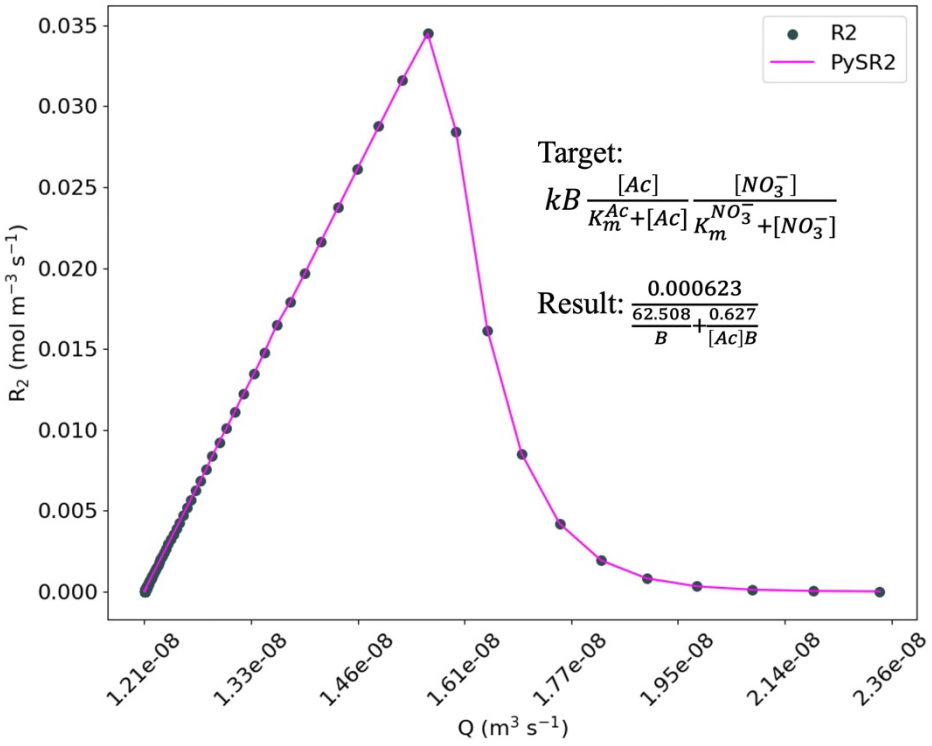

Fig 2.5. PySR results for targeting the R_2 rate law. Input variables used were the Ac and NO_3^- concentrations and biomass density B. Operators used were multiplication, addition, and division.

meaning the algorithm needs to apply these to the data, input a constant, and multiply the other parameters to reconstruct the desired rate expression. To test its capabilities, we gave PySR the rate law for R_2 (Eq. 2.4) as the target and used the Ac and NO_3^- concentrations and B as input variables, letting the algorithm apply three binary operators (e.g., multiplication, division, and addition), with the goal to see it recreate the Monod co-dependencies of the substrates and the rate constant k. The expression generated by SR produced a good fit on the data, but at first glance, the rate law had little resemblance to the target function (Figure 2.5; Eq. 2.11) and is difficult to interpret, compared to the original expression for R_2. However, the expression derived from SR can be recast as a Monod dependency on acetate:

$$R_2 = \frac{0.000623}{\frac{62.508}{B}+\frac{0.627}{[Ac]B}} = \frac{0.000623}{62.508} B \frac{[Ac]}{[Ac]+\frac{0.627}{62.508}} = 10^{-5} B \frac{[Ac]}{[Ac]+0.01} \qquad \textit{Eq. 2.11}$$

In doing this algebraic conversion, the values for the rate constant k and the half-saturation constant K_m are revealed in *Eq. 2.*11, demonstrating that SR can reconstruct these relationships. The Monod relationship for NO_3^- is not picked up by the regressor due to the fact that it is equal to 1 and thus the rate behavior is captured by the Ac dependency. To test further whether PySR could reconstruct the rate expression using only simple operators, we also performed simulations where the Monod expression for Ac in R_2 (*Eq. 2.4*) was used as a template (using the 'warm_start' parameter) for the next model fit targeting the full rate law, including the multiplication operator, as well as the NO_3^- and B features. Similar to the previous simulation, the SR-generated expression bore little resemblance to the R_2 rate law. In an effort to make the generated expressions more interpretable and have a greater resemblance to the source functions, we introduced custom functions informed by our expert knowledge to help guide the SR algorithm in a

direction that improves the readability of the end result. When operators for each of the Monod expressions, including the necessary K_m constants, are used alongside the basic binary operators, the algorithm will select the custom functions in place of the other operators for its output equation (Figure 2.6).

Another example where user-defined building block functions are particularly useful is when we consider rate expressions that depend on the thermodynamic factor F_T. As described in Eq. 2.7, this function requires several operators to calculate, including an exponential, whose arguments depend on the activity of chemical constituents. Multiple constants are also needed for this expression, which would either have to be listed as inputs (e.g., the energy required for ATP generation, the universal gas constant, the temperature) or allow the algorithm a wider range of values to select from than the default (e.g., -1 < 0 < 1). For this reason, a similar approach to the Monod building block was implemented for the F_T, for which the ΔG_{rxn} was computed from the energies of

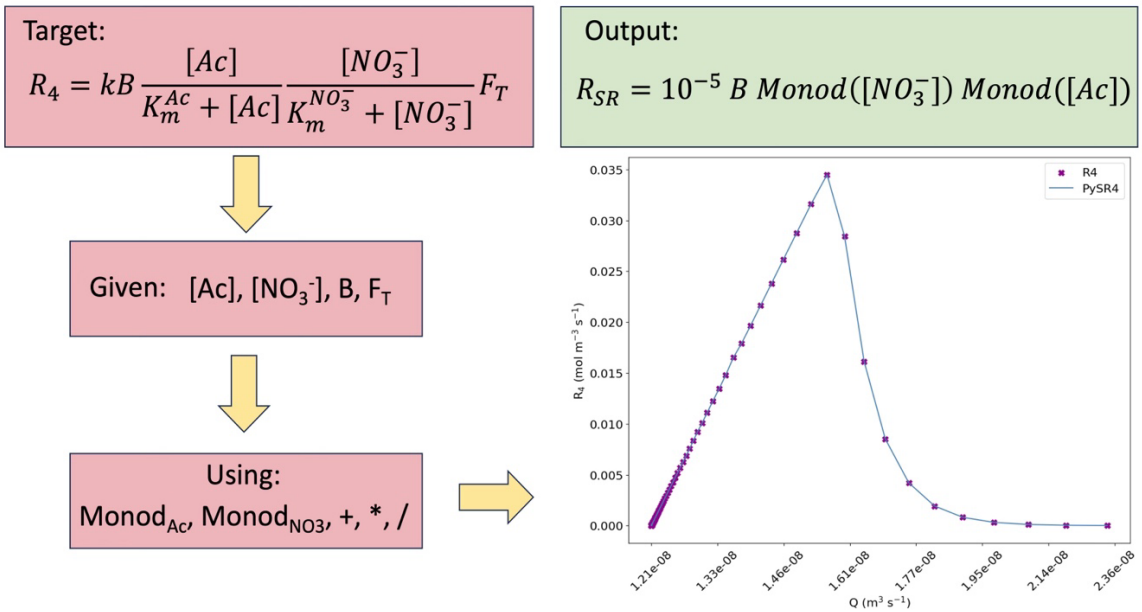

Fig 2.6. Model setup demonstrating why operator choice matters. Use of correct custom functions helps to steer the algorithm in the desired direction. $\mathrm{Monod}(\mathrm{C}) = \frac{C}{K_m + C}$

formation and environmental concentrations (Rxn. 2.2; pH and N_2 are set to 8.2 and 0.01 M respectively).

With the custom Monod and F_T operators implemented, the R_2 and R_4 rate formulations from the chemostat model were targeted (Figure 2.7). PySR was able to reconstruct each of the four rate expressions using the necessary features and building blocks, including the correct value of k for each (Figures 2.4 & 2.7). For R_4, where the rate depends on the impact of the thermodynamic factor F_T, PySR did not make use of it either as an input feature or when using the ΔG_{rxn} as a feature and implementing the F_T building block. This reflects that $F_T = 1$ for our chemostat simulations in which the $\Delta G_{rxn} \approx -740\ kJ\ mol_{Ac}^{-1}$, giving the denitrification rate full thermodynamic drive as a result. Just as was the case in Fig. 2.4B, multiplying the expressions by an integer of 1 would add unnecessary complexity to the expression and thus the SR algorithm did not include it in the rate expression.

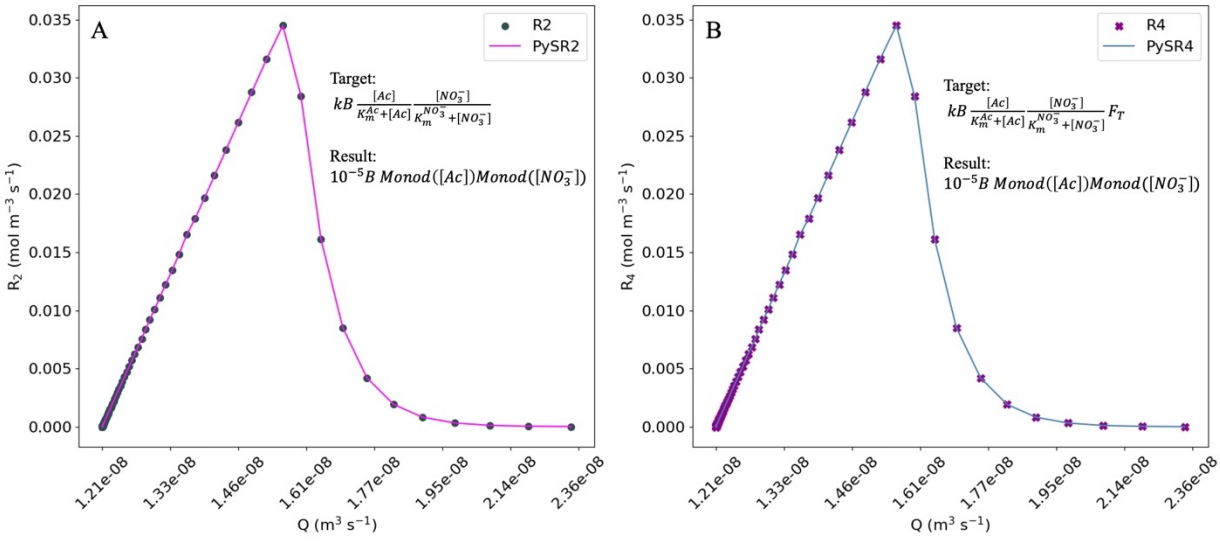

Fig 2.7. PySR results of rate law reconstructions for A) R_2 and B) R_4 from the chemostat model. Input variables used were the Ac and NO_3^- concentrations for both simulations and the microbial cell population B. Custom operators for Monod and F_T expressions were used along with multiplication. $\text{Monod}([C]) = \frac{[C]}{K_m+[C]}$

Application of PySR to the chemostat model dataset shows that the SR algorithm is efficient at recreating the rate laws governing substrate turnover for both chemical and microbially-mediated reactions. However, cell densities of the microbial community associated with a particular process (as shown in *Eq. 2.1*) are difficult to quantify in experimental and environmental systems. Thus, we attempt to reconstruct the chemostat model rates using expressions that utilize only the substrate concentrations present in the medium, but not on the microbial community. The results of these simulations are shown in Figure 2.8. In these example model simulations here, PySR can create a rate expression that closely matches the observed behavior of the denitrification rates despite

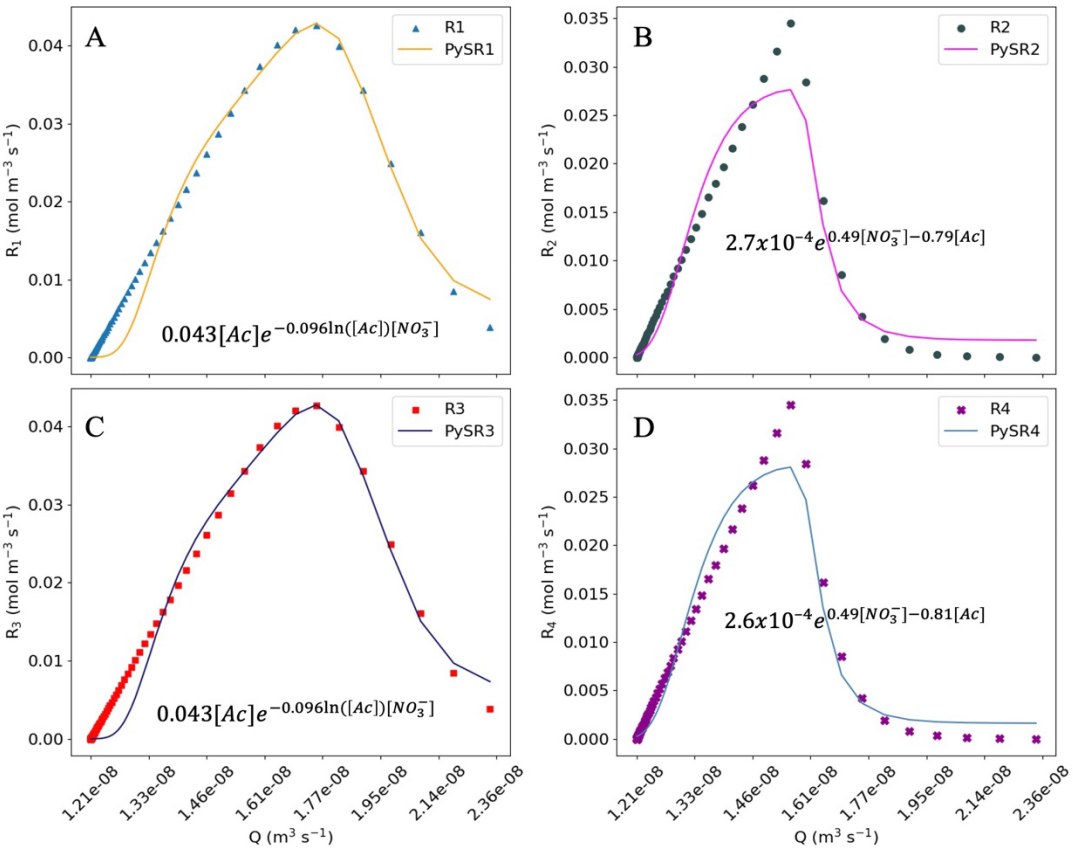

Fig 2.8. PySR results of rate law reconstructions for R_1 to R_4 (A-D) from the chemostat model. Feature variables used were only the Ac and NO_3^- concentrations for each simulation. Operators selected for regression were multiplication, exponential, and natural log. R^2 for the expressions generated are A) 0.97, B) 0.95, C) 0.97, D) 0.95

not having the microbial population density *B* used in the rate expression. Since the features to choose from are limited, the choice of operators is critical. The expressions here were fit with multiplication, an exponential function, and a natural log function as these seemed to match closely with the behavior of the data. These may differ depending on the kind of rates being targeted and thus operator choice will require careful consideration before finding an expression that makes sense to describe the kinetics of microbial reaction rates.

Conclusions

We demonstrated the capabilities of using symbolic regression to accurately reconstruct rate expressions dependent on co-registered observations of rates and concentrations. Application of symbolic regression to profiles generated using a reactive transport model illustrated the ability to select the corresponding concentrations required to describe the rate, even when given a choice of unrelated chemical species. Not only is the structure of the rate expression found, but SR can also quantify the rate constants required to describe rate laws for both microbially-mediated and secondary redox reactions. Symbolic regression thus is a promising approach for establishing rate expressions from measured concentration and rate data. Our work indicates that including operators based on expert knowledge, like the custom operators for Monod expressions, may significantly enhance the interpretability of the regression results. Future explorations applying this approach should also focus on experimental or in situ data of chemical concentrations and their corresponding rates, and explore the sensitivity towards noisy biogeochemical data. In addition, investigating alternative relationships

between process rates and concentrations, such as using reaction energetics derived from the availability of substrates, which may provide a simple framework to describe microbially-mediated reaction rate expressions. Having a reliable and practical method for rate law construction such as symbolic regression would broaden our understanding of the field of biogeochemistry by further uncovering the patterns of microbial processes occurring in environments like coastal marine sediments.

References

Augusto, D. A., & Barbosa, H. J. (2000, November). Symbolic regression via genetic programming. In *Proceedings. Vol. 1. Sixth Brazilian symposium on neural networks* (pp. 173-178). IEEE.

Blackburn, T. H., & Blackburn, N. D. (1993). Rates of microbial processes in sediments. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, *344*(1670), 49-58.

Canfield, D. E., Thamdrup, B., & Hansen, J. W. (1993). The anaerobic degradation of organic matter in Danish coastal sediments: iron reduction, manganese reduction, and sulfate reduction. *Geochimica et Cosmochimica Acta*, *57*(16), 3867-3883.

Cranmer, M. (2023). Interpretable machine learning for science with PySR and SymbolicRegression.jl. *arXiv preprint arXiv:2305.01582*.

Daebel, H., Manser, R., & Gujer, W. (2007). Exploring temporal variations of oxygen saturation constants of nitrifying bacteria. *Water Research*, *41*(5), 1094-1102.

Eppley, R. W., Rogers, J. N., & McCarthy, J. J. (1969). Half-saturation constants for uptake of nitrate and ammonium by marine phytoplankton 1. *Limnology and Oceanography*, *14*(6), 912-920.

Fennel, K., Losch, M., Schröter, J., & Wenzel, M. (2001). Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, *28*(1-2), 45-63.

González-Cabaleiro, R., Ofiţeru, I. D., Lema, J. M., & Rodríguez, J. (2015). Microbial catabolic activities are naturally selected by metabolic energy harvest rate. *The ISME journal*, *9*(12), 2630-2641.

Hales, B., & Emerson, S. (1996). Calcite dissolution in sediments of the Ontong-Java Plateau: In situ measurements of pore water O2 and pH. *Global Biogeochemical Cycles*, *10*(3), 527-541.

Harris, C. M., & Kell, D. B. (1985). The estimation of microbial biomass. Biosensors, 1(1), 17-84.

Heijnen, J. J., & Van Dijken, J. P. (1992). In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnology and Bioengineering*, *39*(8), 833-858.

Heijnen, J. J., & Kleerebezem, R. (2010). Bioenergetics of microbial growth. Encyclopedia of Industrial Biotechnology: Bioprocess. *Bioseparation and Cell Technology2010*, 1-24.

Jin, Q., & Bethke, C. M. (2005). Predicting the rate of microbial respiration in geochemical environments. *Geochimica et Cosmochimica Acta*, *69*(5), 1133-1143.

Jin, Q., & Bethke, C. M. (2007). The thermodynamics and kinetics of microbial metabolism. *American Journal of Science*, *307*(4), 643-677.

Levenspiel, O. (1980). The Monod equation: a revisit and a generalization to product inhibition situations. *Biotechnology and Bioengineering*, *22*(8), 1671-1687.

Lovley, D. R., & Chapelle, F. H. (1995). Deep subsurface microbial processes. *Reviews of Geophysics*, *33*(3), 365-381.

Schink, B. (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and molecular biology reviews*, *61*(2), 262-280.

Van Cappellen, P., & Ingall, E. D. (1994). Benthic phosphorus regeneration, net primary production, and ocean anoxia: A model of the coupled marine biogeochemical cycles of carbon and phosphorus. *Paleoceanography*, *9*(5), 677-692.

Wang, Y., & Van Cappellen, P. (1996). A multicomponent reactive transport model of early diagenesis: Application to redox cycling in coastal marine sediments. *Geochimica et Cosmochimica Acta*, *60*(16), 2993-3014.

Wu, Q., Guthrie, M. J., & Jin, Q. (2022). Physiological Acclimation Extrapolates the Kinetics and Thermodynamics of Methanogenesis From Laboratory Experiments to Natural Environments. *Frontiers in Ecology and Evolution*, *10*, 838487.

CHAPTER 3

GROWTH EFFICIENCY AND NITROGEN USE IN ANAEROBIC METHANE OXIDATION: A MICROBIAL CONSORTIUM MODEL AND FLUX BALANCE ANALYSIS

Introduction

In marine sediments, anaerobic oxidation of methane (AOM) is often catalyzed by syntrophic consortia of methanotrophic archaea and sulfate-reducing bacteria (Boetius et al. 2000; Orphan et al. 2002). These microbial aggregates can consist of hundreds of individual cells that couple their metabolic activity together, allotting both groups enough energy to persist and grow. AOM serves a pivotal role in the breakdown of methane, preventing this potent greenhouse gas from being expelled into the atmosphere. An estimated 45-61 Tg CH_4 yr^{-1} is oxidized by assemblages of anaerobic methanotrophic archaea (ANME) and sulfate-reducing bacteria (SRB) in deep continental shelf sediments (Egger et al. 2018).

The mechanism by which syntrophic electron transfer is conducted in AOM consortia is still the subject of detailed studies (Chadwick et al. 2022; Meulepas et al. 2010). Commonly, diffusible electron carrier molecules are responsible for the transfer of reducing equivalents between cells, but previous model results indicate that diffusion alone is not enough to match observed rate values for AOM and SR respectively (He et al. 2019; Orcutt & Meile 2008). Direct interspecies electron transfer (DIET) can

overcome the diffusive limitations, with evidence suggesting the involvement of conductive structures (McGlynn et al. 2015; Wegener et al. 2015).

Despite only surpassing the energy threshold required for growth, these microbial consortia have been observed to fix inorganic dinitrogen (Dekas et al. 2009; 2014; 2018; Metcalfe et al. 2021). Due to the significant ATP requirement that comes with diazotrophy (Hubbell & Kidder 2009), one may expect differences in growth depending on the N source being used. Porewater ammonium concentrations are often high in sediment cores containing AOM consortia (Dekas et al. 2014 & 2018); however, it has been suggested that in sufficiently large aggregates (>50 μm diameter), ammonium may become limiting at the aggregate interior (Metcalfe et al. 2021), making N_2 fixation a favorable nitrogen source despite energetic constraints.

To explore drivers for nitrogenase activity, we investigate energetic limitation of reaction rates and growth yields of anaerobic methanotrophic archaea (ANME) and sulfate-reducing bacteria (SRB) using two different N sources as substrates, NH_4^+ and N_2. Expanding on the hybrid-DIET model implementation in He et al. (2021) by including explicit representations of anabolism using different N sources, we explore how the distribution of catabolic energy yields and spatial differences in nitrogen uptake patterns may affect where N_2 fixation is conducted in a methane-oxidizing microbial aggregate. The focus of this model is to observe how the incorporation of different nitrogen sources may affect the energetics and potential growth of an AOM aggregate. To quantify the assimilation of N sources into cell biomass, anabolic reaction terms are derived explicitly. Anabolic rates are dependent on the growth yield γ, which is calculated as a function of

the Gibbs free energy changes involved in catabolism, anabolism, and dissipation (González-Cabaleiro et al. 2015; Heijnen & Kleerebezem 2010; Heijnen et al. 1992).

While internal nitrogen limitation is a typical driver for N_2 fixation activity, there exists evidence that some diazotrophs may use nitrogenase enzymes to balance their redox state through increasing H_2 formation and changing post-translational enzyme regulation, even in the presence of ammonium (Wang et al. 2010; Luxem et al. 2020). We hypothesize here that biological nitrogen fixation may play an alternative role in the physiology of ANME-SRB aggregates by allowing excess reducing power to be shunted through the nitrogenase enzyme pathway to maintain internal cell redox homeostasis. This phenomenon has recently been discussed based on gene expression in *Geobacter sulfurreducens*, where a shift to oxidation of NADH by nitrogenase during energy-constrained periods is thought to occur (Ortiz-Medina et al. 2023). Using a metabolic flux balance model, we explore the factors that control growth efficiencies in ANME cells. We also investigate the effects of NH_4^+ limitation and increased redox potential through maximizing ferredoxin production to explore if this may stimulate N_2 fixation activity in ANME.

Methods

The model presented here builds on the three-dimensional model of an ANME-SRB aggregate presented in He et al. (2021), where methane is oxidized in regions occupied by the archaea and sulfate is reduced in those inhabited by the bacteria. Simulations feature syntrophic coupling of AOM and SR reactions through a mixed electron transfer mechanism. Electrons generated in the oxidation of methane are either

loaded onto a dissolved chemical transporter, expressed here as $HCOO^-$, for mediated interspecies electron transport (MIET) or are used on redox active molecules (Cyt: oxidized; CytH: reduced) that conductively transport them from one syntrophic partner to the other through DIET. This simplified model design serves to reduce complexity which reflects the limited knowledge for EET kinetics in AOM-associated syntrophic aggregates. The catabolic reactions for archaeal AOM and bacterial SR are shown in the reactions below:

$$CH_4 + (f_{MIET} - 1)HCO_3^- + f_{DIET}Cyt \rightarrow f_{MIET}HCOO^- + H^+ + f_{DIET}CytH + (f_{MIET} - 3)H_2O \quad \textit{Rxn 3.1}$$

$$SO_4^{2-} + H^+ + f_{MIET}HCOO^- + f_{DIET}CytH \rightarrow HS^- + f_{DIET}Cyt + f_{MIET}HCO_3^- + (4 - f_{MIET})H_2O \quad \textit{Rxn 3.2}$$

where f_{DIET} and f_{MIET} are stoichiometric coefficients representing the fraction of electrons transferred in DIET and MIET respectively. A total of 8 electrons are produced per methane oxidized to CO_2 and are loaded onto transfer molecules such that f_{DIET} = (0,8) and f_{MIET} = (8- f_{DIET})/2. f_{DIET} = 8 and f_{MIET} = 0 represents DIET only.

Reaction Rates

The catabolic reaction rates were calculated using the following form (Jin & Bethke 2003; Jin & Bethke 2007):

$$R^x = k_x B F_k^x F_T^x \quad \textit{Eq. 3.1}$$

where k_x is the cell-specific rate constant for reaction x, B is the cell density, and F_k^x represents the reaction kinetics using a Monod expression for the concentration of substrate S and a half-saturation constant K_m^s:

$$F_k^x = \frac{S}{K_m^S + S} \qquad \textit{Eq. 3.2}$$

The cell density B represents the local cell concentration and is calculated from the total number of ANME and SRB cells per aggregate, divided by the volume of the aggregate occupied by the respective microorganisms (see below).

The thermodynamic factor ($0 \leq F_T^x \leq 1$), which states that there must be sufficient availability of free energy generated by reaction *x* for use in processes such as ATP synthesis and cell maintenance, is implemented as:

$$F_T^x = \max\left(0, 1 - \exp\left(\frac{f_x}{\chi R_{gas} T}\right)\right) \qquad \textit{Eq. 3.3}$$

where χ is defined as the number of times the rate limiting step occurs per reaction, which we assign a value of 1 ATP synthesized per reaction of both AOM and SR (Orcutt & Meile 2008), R_{gas} is the ideal gas constant (8.314 J K^{-1} mol^{-1}), and *T* is the temperature (277.15 K). The thermodynamic driving force f_x then describes the relationship between the free energy yield of reaction *x* and the energy necessary for ATP formation (Jin & Bethke 2003; 2007). It is represented by:

$$f_x = -\Delta G_{net}^x - m\Delta G_{ATP} \qquad \textit{Eq. 3.4}$$

where ΔG_{ATP} is the minimum free energy requirement for the synthesis of ATP and *m* is the number of ATP produced per electron transferred. The parameter *m* reflects the impact of the rate limiting step often associated with energy conservation, such as the formation of a proton gradient (Jin & Bethke 2005). Using the reverse methanogenesis model as a metabolic blueprint for AOM (McGlynn 2017; Deppenmeier 2002), proton translocation is shown to occur in a ratio of $4H^+/2e^-$. This demonstrates that approximately 1 ATP molecules are generated per 2 electrons transported, giving us *m* =

$\frac{1}{2}$ (Orcutt & Meile 2008). Assuming the energy required for ATP synthesis in an actively growing cell is ~60 kJ per mol of ATP (Schink 1997), and that three protons are translocated per ATP molecule produced, this makes the energetic threshold 20 kJ mol^{-1} H^+. The product of $m\Delta G_{ATP}$ then outlines the lower bound of thermodynamic favorability for the reaction, set here to 10 kJ mol_{ATP}^{-1}. ΔG_{net}^{x} represents the net free energy available from catabolism, which is given by:

$$\Delta G_{net}^{x} = \Delta G_{rxn}^{x} - \Delta G_{act} - \Delta G_{om} \qquad \textit{Eq. 3.5}$$

where ΔG_{rxn}^{x} is the Gibbs free energy of reaction x is calculated as:

$$\Delta G_{rxn}^{x} = \Delta G_{rxn}^{x,o} + R_{gas}T\ lnQ_{rxn}^{x} \qquad \textit{Eq. 3.6}$$

The reaction quotient is given by $Q_{rxn}^{x} = \Pi a_i^{v_i}$, where a_i refers to the activity of substrate i and v_i is the corresponding stoichiometric coefficient. Activities are calculated as the product of the substrate concentrations and the activity coefficients, which are set to be constant assuming seawater conditions. The energetic loss terms ΔG_{act} and ΔG_{om} in Eq. 3.5 refer to the activation potential energy and ohmic resistance energy, respectively. These terms are in place to describe the energetic impact on metabolism, due to the loading and transport of electrons via DIET, for cells at some distance from the archaeal-bacterial interface. The differences in electric potential are converted to Gibbs free energy changes by $\Delta G = -nF\eta_x$, where η_x corresponds to the voltage of process x, F is the Faraday constant, and n is the number of electrons transferred per reaction.

Activation potential losses η_{act} are described by a Butler-Volmer relationship that assumes a one-step, single-electron transfer between the cell and the conductive pili (Storck et al. 2016). These costs are calculated as follows:

$$\frac{I}{N_{nw}} = FA_{act}k_{act}Cyt_{tot}(exp\left(\frac{(1-\beta)F}{R_{gas}T}\eta_{act}\right) - exp\left(\frac{-\beta F}{R_{gas}T}\eta_{act}\right)) \quad \textit{Eq. 3.7}$$

where the current generated during methane oxidation is $I = f_{DIET}R_{AOM}N_{ANME}F$ (R_{AOM} is the rate of methane oxidation in fmol cell^{-1} day^{-1}, N_{ANME} is the archaeal cell number, and F is the Faraday constant) and N_{nw} is the number of conductive connections within the consortium, which is described by $N_{nw} = Cyt_{tot}V_{agg}k_{nw}$, where V_{agg} is the consortia volume in m^3 and k_{nw} is the conductive network constant in mol^{-1}. A_{act} is the surface area of redox active molecules (m^2 cell^{-1}), k_{act} is the activation loss constant (m s^{-1}), β is the charge transfer coefficient, and Cyt_{tot} is the total concentration of redox electron carrier molecules.

Ohmic resistance losses occur as electrons flow through the conductive matrix between partner cells and is proportional to the current density:

$$\eta_{om} = \frac{R_{nw}I}{N_{nw}} = \frac{d}{\sigma A_{nw}}\frac{f_{DIET}R_{AOM}N_{ANME}F}{Cyt_{tot}V_{agg}k_{nw}} \quad \textit{Eq. 3.8}$$

with $R_{nw} = \frac{d}{\sigma A_{nw}}$ representing the electrical resistance (Ω), d is the distance from the syntrophic partner, approximated as the distance to the nearest ANME-SRB interface, σ is the pilus electrical conductivity (S m^{-1}), and A_{nw} is the cross-sectional area of a conductive pilus.

Growth Efficiencies & N Usage

The catabolic and anabolic reactions for a microorganism can be used to calculate growth efficiencies, while also accounting for the energy dissipated as heat due to inefficiencies during biomass production (González-Cabaleiro et al. 2015; Heijnen & Kleerebezem 2010; Heijnen et al. 1992). The energy a microorganism acquires through the conversion of a substrate acting as the electron donor (eD; here CH_4) in catabolism

(ΔG_{cat}^{x}) is partially invested in maintaining cellular processes, while the rest can be used to fuel growth via anabolism (ΔG_{ana}^{x}). Anabolic reaction stoichiometries for archaeal and bacterial cells were defined based on the approach outlined in Heijnen et al. (1992) using the cell biomass formula $C_x = CH_{2.0}O_{0.5}N_{0.27}$ which is based on a methanogen with similar cell composition to members of AOM consortia (Popovic 2019; van Dijken & Harder 1975):

$$0.524\ CH_4 + 0.476\ CO_2 + 0.27\ NH_4^+ \rightarrow C_x + 0.453\ H_2O + 0.27\ H^+ \qquad \textit{Rxn 3.3}$$

$$CO_2 + 0.524\ HS^- + 0.27\ NH_4^+ + 0.595\ H_2O \rightarrow C_x + 0.524\ SO_4^{2-} + 0.794\ H^+ \qquad \textit{Rxn 3.4}$$

With methane being the most reduced form of carbon, the ANME benefit from an additional oxidized C source, in this case CO_2, to achieve the 'biosynthetic net oxidation' necessary for building biomass (Nauhaus et al. 2007). The SRB build biomass autotrophically through CO_2 fixation (Murali et al. 2023).

Energy generated during catabolism cannot all be used for growth; thus, it is necessary to account for the portion of energy dissipated during anabolism ($\Delta G_{dis}, kJ\ mol_{Cx}^{-1}$). There are competing views for how to best represent this release of free energy (Heijnen & Kleerebezem 1993); here the dissipation of energy due to biomass formation is estimated using only the number of C atoms in the carbon source (*NoC)* and the degree of reduction (ϵ_s), as shown in Eq. 3.9 (Heijnen & Kleerebezem 1992):

$$\Delta G_{dis} = 200 + 18(6 - NoC)^{1.8} + \exp\left[((3.8 - \epsilon_s)^2)^{0.16} * (3.6 + 0.4C)\right] \qquad \textit{Eq. 3.9}$$

Relating the energies of catabolism and anabolism, as well as considering the amount of energy dissipated based on the C source used for growth, the maximum growth yield γ for a microbe can be estimated:

$$\gamma_{NH4} = \frac{\Delta G_{net}}{-(\Delta G_{ana} + \Delta G_{dis})} \quad \textit{Eq. 3.10}$$

where γ_{NH4} is in units of $mol_{Cx}\ mol_{CH4}^{-1}$ and, using NH_4^+ as the only N source for anabolism. Using this approach, rates of nitrogen uptake for anabolism can then be easily defined as:

$$R_{NH_4} = \nu_{N:C}\gamma_{NH4}R_X \quad \textit{Eq. 3.11}$$

where $\nu_{N:C}$ is the ratio of N to C in biomass and R_x is the reaction rate of catabolism (mol m^{-3} s^{-1}). This approach assumes that the uptake of ammonium is directly linked to the catabolic rate and thus the archaea and bacteria are only able to grow when catabolism is active. When considering growth using N_2 fixation, the energy required for nitrogenase activation must also be considered. The growth efficiency when using N_2 fixation can be formulated as:

$$\gamma_{N_2} = \frac{\Delta G_{net}}{-(\Delta G_{ana} + \Delta G_{dis} + \Delta G_{nfx})} \quad \textit{Eq. 3.12}$$

where ΔG_{nfx} is the Gibbs free energy change of reaction required for N_2 fixation to fulfill the N demand of the cell (expressed as the NH_4^+ requirement for use in building biomass that is captured in ΔG_{ana}). This reaction takes the form of:

$$0.135\ N_2 + 1.08\ Fdx_{red} + 3.51\ H^+ + 2.16\ ADP + 2.16\ HPO_4^{2-} \quad \textit{Rxn 3.5}$$
$$\rightarrow 0.27\ NH_4^+ + 1.08\ Fdx_{ox} + 0.135\ H_2 + 2.16\ ATP + 2.16\ H_2O$$

with $Fdx_{red/ox}$ representing the reduced and oxidized forms of ferredoxin respectively. This was balanced in terms of the N demand required for growth in anabolism for ANME and SRB (*Rxn 3.3 & 3.4* respectively). ATP hydrolysis is what ultimately drives the reaction energetically (Alberty 1994). This is captured in *Rxn 3.5* by formulating ATP as

a product, thereby accounting for the energetic cost of producing ATP used in N_2 fixation. As a consequence, ΔG_{nfx} becomes more positive and adds to the cost of biosynthesis in *Eq. 3.12*. The rate of N_2 fixation reaction is then given by:

$$R_{N_2} = \nu_{N:C}\gamma_{N_2}R_x \qquad \textit{Eq. 3.13}$$

Eqs. 3.11 & 3.13 give us two models that use either NH_4^+ or N_2 as the nitrogen source for calculation of the growth yields in an ANME-SRB aggregate.

Energy conservation mechanisms play a significant role in the physiology of slow growing chemotrophic anaerobes (Thauer et al. 1977; Jørgensen & Marshall 2016). Strategies like regeneration of reducing equivalents and proton gradient formation through intracellular mechanisms is key to their survival in low energy regimes. We implemented two methods for energy conservation where the H_2 formed during nitrogen fixation can either be used to regenerate ferredoxin (*Rxn 3.6*) or to build a proton gradient to aid ATP production (*Rxn 3.7*):

$$\alpha(H_2 + 2Fdx_{ox} \rightarrow 2Fdx_{red} + 2H^+) \qquad \textit{Rxn 3.6}$$

$$(1-\alpha)(H_2 + methanophenazine + 2H_{in}^+ \rightarrow H_2\text{-}methanophenazine + 2H_{out}^+) \qquad \textit{Rxn 3.7}$$

where α is a parameter switch set to 1 or 0 depending on the pathway selected for using nitrogenase-produced H_2.

Amino Acid Leakage & Redox Homeostasis

N_2 fixation was investigated primarily as a method for providing an N source in biomass synthesis (*Eq. 3.12*). However, nanoSIMS imaging of labeled ^{15}N incorporation has shown evidence that diazotrophic ANME-2 also share reduced N sources with their SRB partners (Dekas et al. 2009). While ammonium is often widely available in marine sediments known to contain ANME-SRB aggregates (Joye et al. 2004; Dekas et al.

2018), the ANME may be producing the excess fixed nitrogen through use of the nitrogenase enzyme to maintain intracellular redox conditions when under energy constrained conditions, similar to recent observations in *Geobacter sulfurreducens* (Ortiz-Medina et al. 2023). To further test this idea in our model, we simulated the aggregate to use NH_4^+ for growth while nitrogen fixation produced an amino acid (AA), in this case arginine (Arg), to be exported to the consortia and surrounding porewater. AA production is calculated as a fraction of the rate of catabolism, to mimic that faster metabolism leads to the production of more reducing power within the cell. As in this conceptual model, N_2 fixation does not contribute to biomass growth, its impact on growth efficiency estimates is accounted for as a reduction of the catabolic energy available for growth so that

$$\gamma_{N_2} = \frac{\Delta G_{cat} - n^* \Delta G_{nfx}}{-(\Delta G_{ana} - \Delta G_{dis})} \qquad \textit{Eq. 3.14}$$

where n^* represents the amount of N_2 that can be fixed per mol of methane oxidized. N_2 fixation has been reported to cause a 20-fold reduction in growth rates (Dekas et al. 2009). Hence, n^* can be estimated from the reaction energetics and γ_{N_2} to be on the order of 0.05 – 0.085 mol_{N2} mol_{CH4}^{-1}, using representative values of ΔG_{cat}, ΔG_{ana} and ΔG_{dis} that were determined from simulations in which the AA concentration was imposed as 0 at the model boundary away from the aggregate surface, assuming uptake occurs by other microorganisms in the surrounding porewater.

Table 3.1: Default parameters used in the 3D reactive-transport model.

Name	Value	Description	Sources
Geometry			
r_{agg}	20 µm	Radius of the ANME-SRB aggregate	Metcalfe et al. 2021
r_{ANME}	0.4 µm	Radius of archaeal cell	Boetius et al. 2000; Knittel et al. 2005

r_{SRB}	0.4 µm	Radius of bacterial cell	Boetius et al. 2000; Knittel et al. 2005
r_{env}	50 µm	Radius of aqueous environment around aggregate	He et al. 2021
V_{agg}	3.35 x 10^{-14} m^3	Volume of the aggregate	Calculated by $V_{agg} = \frac{4}{3}\pi r_{agg}^3$
Kinetics			
k_{AOM}	4.0 x 10^{-16} m^3 $cell^{-1}$ d^{-1}	Rate constant for methane oxidation	He et al. 2019; He et al. 2021
k_{SR}	4.0 x 10^{-16} m^3 $cell^{-1}$ d^{-1}	Rate constant for sulfate reduction	He et al. 2019; He et al. 2021
$\boldsymbol{K_m^{CH_4}}$	7 mM	Half-saturation constant for methane	Bowles et al. 2019
$\boldsymbol{K_m^{SO_4^{2-}}}$	5 mM	Half-saturation constant for sulfate	Roychoudhury 2004
f_{DIET}	7.4	Fraction of electrons transferred via DIET	He et al. 2019; He et al. 2021
f_{MIET}	0.4	Fraction of electrons transferred via MIET	He et al. 2019; He et al. 2021
Thermodynamics			
$\boldsymbol{\Delta G_{ATP}}$	-10 kJ mol^{-1}	Energy requirement for ATP synthesis	Orcutt & Meile 2008; He et al. 2019
$\boldsymbol{\chi}$	1	Number of ATP synthesized per reaction	Orcutt & Meile 2008; He et al. 2019
R_{gas}	8.314 J K^{-1} mol^{-1}	Ideal gas constant	
T	277.15 K	Temperature	Dekas et al. 2018; Metcalfe et al. 2021
n	8	Electrons transferred per reaction	From rxns 1 & 2
F	96485.3 C mol^{-1}	Faraday constant	
$\boldsymbol{NoC_{CH_4}}$	1	Number of C atoms per methane	
$\boldsymbol{\epsilon_{CH_4}}$	8	Degree of reduction for methane	Heijnen et al. 1992
$\boldsymbol{NoC_{CO_2}}$	1	Number of C atoms per CO_2	
$\boldsymbol{\epsilon_{CO_2}}$	0	Degree of reduction for CO_2	Heijnen et al. 1992
$\boldsymbol{n^*}$	0.05-0.085	mol N_2 fixed per mol CH_4 oxidized	Calculated
Electron Transfer			
Cyt_{tot}	10 mM	Concentration of cytochrome redox molecules	Korth et al. 2015; He et al. 2021
k_{ET}	1 x 10^5 m^4 mol^{-1} s^{-1}	Electron transfer rate constant on conductive matrix	He et al. 2021
k_{act}	2 x 10^{-9} m s^{-1}	Activation loss rate constant	He et al. 2021

A_{act}	2 x 10^{-13} m^2	Redox-active surface area of cell	Storck et al. 2016; He et al. 2021
$\boldsymbol{\delta}$	0.7 nm	Width between redox molecules	Strycharz-Glaven et al. 2011
$\boldsymbol{\sigma}$	0.01 S m^{-1}	Conductivity of matrix	He et al. 2021
$\boldsymbol{\beta}$	0.5	Charge transfer coefficient	Storck et al. 2016
k_{EF}	1 x 10^{-5} m^4 mol s^{-1}	Electric field rate constant	He et al. 2021
k_{nw}	1.2 x 10^{19} mol^{-1}	Conductive matrix constant	He et al. 2021
d_{nw}	4 nm	Diameter of a single conductive pilus	Malvankar et al. 2011
A_{nw}	1.26 x 10^{-17} m^2	Pilus cross-sectional area	He et al. 2019; He et al. 2021
N_{nw}	4 x 10^6	Total conductive connections in the aggregate	He et al. 2019; He et al. 2021
$N_{nw,cell}$	64	Conductive connections per cell	Storck et al. 2016; He et al. 2021

Model Domain & Spatial Transport

The model places a spherical aggregate in the center of a spherical domain representing the pore-space environment (Figure 3.1A), having a radius of 2.5x the radius of the aggregate (r_{agg}). The morphology of the ANME-SRB consortia was adopted from He et al. (2021) and fluorescence *in situ* hybridization (FISH) images of labeled aggregates to investigate how energetic impacts due to DIET may affect N_2 fixation potential. The archaeal and bacterial cells within the aggregate have radii of 0.4 μm (Boetius et al. 2000; Orphan et al. 2002) and are set to a 1:1 cell ratio. While growth efficiencies are calculated in the model, cell densities and cell sizes within the consortia, and hence the volumes of AOM and SR, are held constant. We recognize that AOM aggregates and the cells therein can vary in shape and size, where alterations of these physical characteristics would likely change the results shown here; however, this does not undermine our study as the steady state model represents a snapshot in time, justified by the slow change in aggregate morphology compared to the timescales of solute

distributions within the model domain. The model was implemented in COMSOL Multiphysics 6.1 (COMSOL Inc., Burlington, MA, USA).

Model simulations resolve steady-state diffusive transport and reactions that are governed by equations of the form:

$$\frac{\partial \phi C_i}{\partial t} = \nabla \cdot (\phi D_i \nabla C_i) + \phi R_i \qquad \textit{Eq. 3.15}$$

where ϕ is the porosity – imposed at a value of 0.3 inside the aggregate and set to 1 in surrounding pore water, C_i represents the concentration of chemical species i, t is time, D_i is the effective diffusion coefficient for chemical species i accounting for tortuosity and

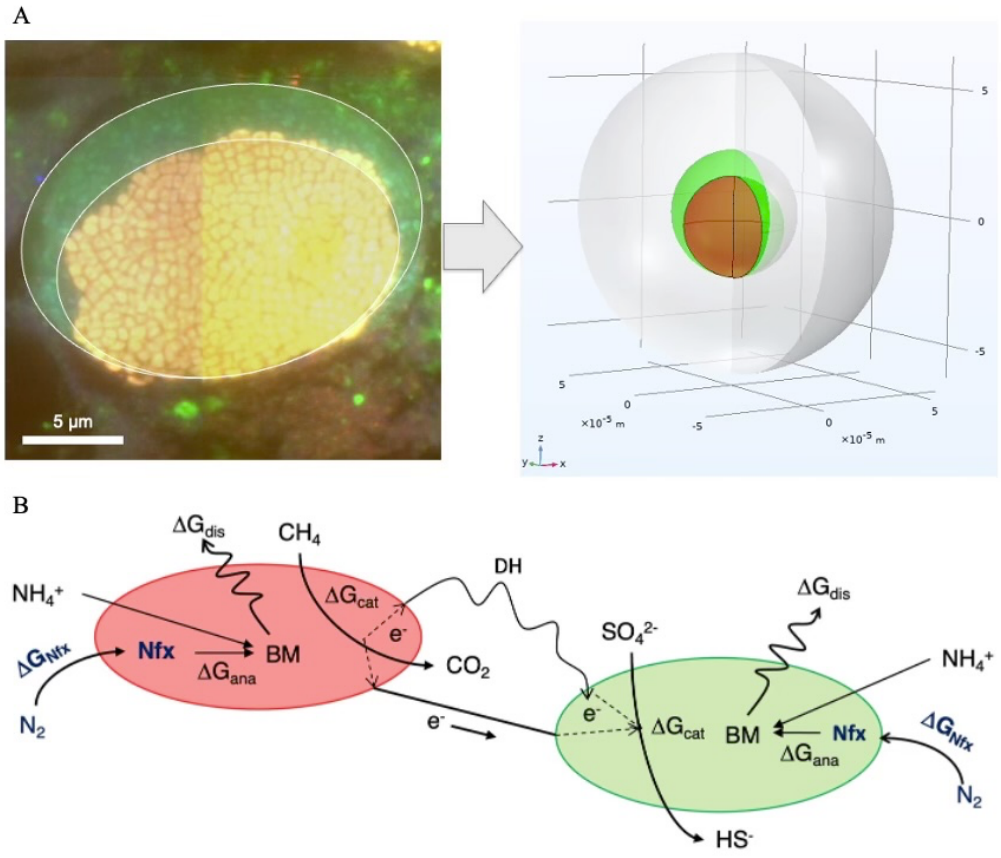

Fig 3.1. A) FISH image (Metcalfe et al. 2021) of representative AOM-aggregate that is then used to generate a 2D axisymmetric model used to describe AOM activity. Red sphere at the center represents ANME, green shell contains the SRB, and the white outer shell is the aqueous environment; B) schematic of the metabolic processes considered in the model; electron transfer is split between direct interspecies electron transfer (DIET) and dissolved intermediate carriers (D/DH; mediated interspecies transfer (MIET). BM: biomass, Nfx: N_2 fixation, ana: anabolism, cat: catabolism, dis: dissipation.

effects of extracellular polymeric substances (EPS), and R_i is the net reaction rate. The chemical species being solved for are $CH_{4\,(aq)}$, HCO_3^-, SO_4^{2-}, HS^-, $CO_{2\,(aq)}$, CO_3^{2-}, OH^-, H^+, $H_{2\,(aq)}$, $HCOO^-$, $HCOOH$, $B(OH)_4^-$, NH_4^+, N_2, and Arg. A schematic representation of the reactions and substrates involved in the model is shown in Figure 3.1B. The boundary conditions and aqueous diffusion coefficients used in the model are listed in Table 3.2. The outer boundary was set to fixed concentrations, using the same values as initial conditions.

Table 3.2: Boundary Conditions & Aqueous Diffusion Coefficients

Chemical Species	**Concentration (mM)**	**D_{aq} ($m^2\ s^{-1}$)**	**Source**
$CH_{4\,(aq)}$	4.5	8.97×10^{-10}	Schulz 2000
HCO_3^-	(*)	6.87×10^{-9}	Schulz 2000
SO_4^{2-}	28	5.72×10^{-9}	Schulz 2000
HS^-	0.1	1.11×10^{-9}	Schulz 2000
$CO_{2\,(aq)}$	(*)	1.03×10^{-9}	Schulz 2000
CO_3^{2-}	(*)	5.04×10^{-10}	Schulz 2000
OH^-	(*)	2.97×10^{-9}	Schulz 2000
H^+	6.31×10^{-6}	5.89×10^{-9}	Schulz 2000
$H_{2\,(aq)}$	1×10^{-5}	2.36×10^{-9}	Schulz 2000
$HCOO^-$	0.01	4.91×10^{-9}	Schulz 2000
$HCOOH$	0.01	1.52×10^{-9}	Schulz 2000
$B(OH)_4^-$	(**)	9.56×10^{-10}	Yongquan et al. 2013
NH_4^+	0.025	1.1×10^{-9}	Schulz 2000
$N_{2\,(aq)}$	0.95	1.07×10^{-9}	Schulz 2000
Arg	0	7.0×10^{-10}	Ma et al. 2005

(*) calculated for DIC = 2.36 mM & pH 8.2 (Zeebe & Wolf-Gladrow 2005; Goyet 1994)
(**) calculated for total borate = 0.427 mM & pH 8.2 (Zeebe & Wolf-Gladrow 2005; Goyet 1994)

Within the aggregate, the aqueous diffusion coefficients (D_{aq}) were adjusted for tortuosity and the effects of EPS that ANME-SRB consortia have been observed to secrete (Orphan et al. 2001; Knittel et al. 2005). This results in an expression for effective diffusivity D of the form:

$$D = f_{eps}\frac{D_{aq}}{\theta^2} \qquad \textit{Eq. 3.16}$$

where f_{eps} is the factor controlling effects of EPS, which is set here to 0.25 for organic ions and 0.6 for inorganic ions and gases (Stewart 2003), and θ^2 is the tortuosity factor of 2.5 based on a porosity on the order of 0.3 and the tortuosity-porosity relationships presented in Figures 4.9 & 4.10 of Boudreau (1997).

Implementation of DIET

There have been multiple attempts to model the DIET mechanism found in some anaerobic microbes; one such model is defined by electrons that hop along a concentration gradient of redox-active molecules creating a bridge for paired cells, while in another, the electrons follow an electric field that is generated between the syntrophic neighbors (Storck et al. 2016; Strycharz-Glaven et al. 2011). We represent both methods in our model implementation, acknowledging that in previous modeling efforts, the electric gradient method alone was not sufficient to demonstrate the overall electron flux generated in an AOM aggregate (He et al. 2021). Our DIET mechanism is further combined stoichiometrically with formate to act as a small diffusive molecule that results in a hybrid of electron transfer methods, using a fraction of the total number of electrons transferred per reaction split between the two pathways, as shown in *Rxn 3.1* and 3.*2*. While previous studies have shown that formate does not inhibit AOM in high concentrations, suggesting it may not serve a role in electron transfer (Meulepas et al. 2010), genomic evidence suggests the presence of formate dehydrogenases that may still be used in catabolic reactions (Chadwick et al. 2022). Thus, formate serves as a general representative for a dissolved diffusive electron carrier candidate in our model here.

To quantify the dynamics of DIET within the aggregate, we establish a mass balance equation similar to the other substrates. The concentration of reduced electron

carrier *CytH* is dependent on the rate of AOM and electron hopping along the conductive pili, which is defined in the model by:

$$\frac{\partial CytH}{\partial t} = \phi f_{DIET} R_{AOM} + \nabla \cdot \left(D_{Cyt} \nabla [CytH]\right) + \nabla \cdot J_{EF} \qquad \textit{Eq. 3.17}$$

where the effective diffusion coefficient, $D_{Cyt} = k_{ET}[Cyt_{tot}]\delta$, depends on the electron transfer rate constant (k_{ET}), the total concentration of redox molecules, and the spacing width between two active redox molecules (δ). $\nabla \cdot J_{EF}$ describes the flux of electrons through a local electric field (Strycharz-Glaven et al. 2011) and is defined by:

$$J_{EF} = k_{EF}[Cyt][CytH]\left(e^{\frac{\beta F E \delta}{R_{gas}T}} - e^{-\frac{(1-\beta)FE\delta}{R_{gas}T}}\right) \qquad \textit{Eq. 3.18}$$

where k_{EF} is the rate constant associated with the electric field, E is the electric field strength, and β is the charge transfer coefficient.

Implementation of acid base buffer

An acid-base reaction network involving dissolved inorganic carbon (DIC) and borate speciation based on the kinetic description detailed in Zeebe & Wolf-Gladrow (2005) and Dickson & Goyet (1994) was implemented to control protonation of cell surface-bound carboxy (R-COOH) and amino (R-NH_3) groups. The system further calculates dynamic concentrations of the carbonate system, including HCO_3^- and CO_2, using a total borate (B_{tot}) concentration of 0.427 mM and a total carbonate concentration (C_{tot}) of 2.36 mM.

FBA Model

In an effort to further investigate the growth dynamics and N usage in methane oxidizing archaea, we obtained an unpublished *in silico* flux balance analysis (FBA) model describing an ANME-2b cell from Chris Henry and colleagues at Argonne

National Laboratory (Faria et al. 2023). This model consists of 810 reactions and 814 metabolites that make up the metabolic network of an ANME-2b cell, informed by genomic data. While the mechanism for AOM discussed so far has used sulfate as the terminal electron acceptor via the syntrophic relationship with SRB, the FBA model here only considers the ANME archaea. Decoupling the syntrophic relationship between the two microbial groups has been shown to occur using the artificial electron acceptor 9,10-anthraquinone-2,6-disulfonate (AQDS) while still maintaining similar rates of AOM as when conducted with sulfate, serving as a suitable substitute for methane-derived electron capture (Scheller et al. 2016). Thus, we use AQDS as the electron sink, representing the role of SRB in the environment.

The core function of the model is to quantify the fluxes of substrates into the cell and the terminal products the cell secretes at the end. Fluxes of metabolites are established by a linear programming problem behind the metabolic reaction network (Eq. 3.19)

$$0 = S * F \qquad \textit{Eq. 3.19}$$
$$lb < F < ub$$

Where the fluxes *F* represent the fluxes or metabolic reactions and the coefficients of the stoichiometric matrix *S* capture the impact on the metabolites in these reactions. The system is considered at steady-state, i.e. the production and consumption of each metabolite is balanced. Fluxes are further constrained by imposing lower and upper bounds, *lb* and *ub* respectively, that control the capacity of metabolite flux through the cell. For example, the medium composition can impose upper limits on substrate uptake from the environment. Using the stoichiometric and flux capacity constraints to generate

a solution space, the model then calculates a balanced flux distribution to maximize or minimize a particular objective function (OF). OFs are typically biologically relevant phenotypes applicable to the topic of study, such as biomass production or the rate of formation for an intermediate metabolite (Orth et al. 2010). Given that this is a steady-state model, the elemental fluxes entering the cell should be matched by internal sinks and the effluxes (Orth et al. 2010). A simple diagram that demonstrates some of the inputs and outputs for our ANME model is shown in Figure 3.2, however, there are a total of 21 input and 84 output chemical species described in the model that will vary depending on the conditions the model is run under.

It should be noted that previous experimental evidence suggests organic C compounds such as acetate, formate, and methanethiol (MTTL) are likely not produced by ANME for syntrophic electron transport given the small inhibition response to AOM that one would expect a dissolved intermediate electron carrier to have (Nauhaus et al.

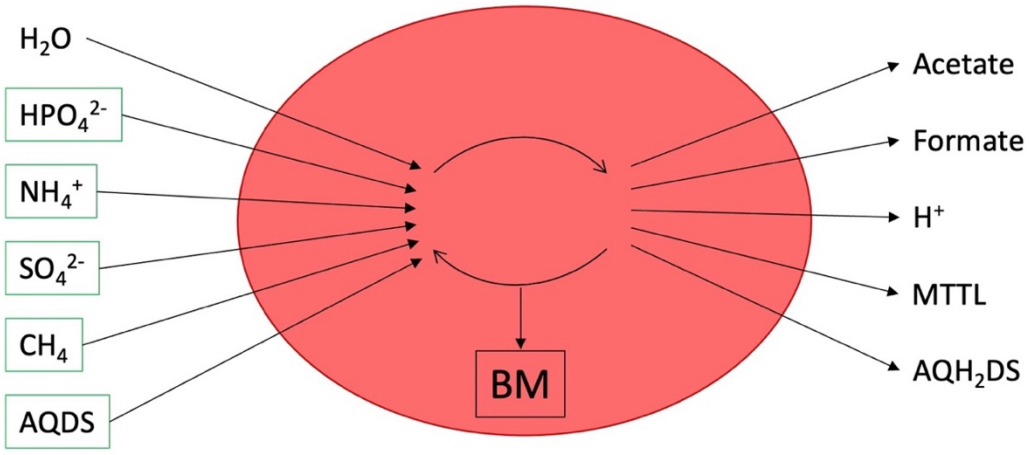

Fig 3.2. An example of the input substrates required for biomass production (green outline) in the ANME FBA model. These substrates are used internally to generate energy via ATP production and building biomass through N and C assimilation. Metabolites on the right side indicate those that are secreted from the cell, representing only a few possible molecules produced. BM: biomass, MTTL: methanethiol

2005; Meulapas et al. 2010). This highlights the preliminary nature of the model presented here and should be considered when interpreting results.

The FBA model uses a defined growth medium that sets limits on the exchange fluxes of substrates from the external environment to the internal metabolic network. Imposing reasonable maximum constraints on the growth medium is necessary to set an appropriate reference frame for the model. We elected to translate growth medium substrate concentrations into estimates of maximum diffusive uptake fluxes (see Appendix B) based on an experimental setup where accurate ANME growth efficiencies have been quantified (Nauhaus et al. 2007). A diffusive boundary layer of 0.1 mm was imposed for calculating the diffusive fluxes, assuming similar substrate uptake patterns as have been observed in the overlying water of sedimentary microbial mats (Jørgensen & Marais 1990).

For our simulations shown here, the OF was set to maximize biomass production. Biomass fluxes were then translated into growth efficiencies, which required knowledge on the chemical composition of biomass used in the model. Using the stoichiometry for the biomass production reaction, the chemical formula for biomass was estimated to be $C_{33.51}H_{72.54}O_{22.22}N_{9.338}P_{0.761}$, which has similar composition to the biomass formula used in *Rxns. 3.3 & 3.4* (Popovic 2019; van Dijken & Harder 1975). With this formulation, the growth efficiencies are calculated as:

$$\gamma = \frac{\nu_{C-BM} F_{BM}}{F_{CH4}} \qquad \textit{Eq. 3.20}$$

where γ is the growth efficiency in $mol_{BM}\, mol_{CH4}^{-1}$, ν_{C-BM} is the number of C per mol of biomass while F_{BM} and F_{CH4} are the fluxes of biomass and methane respectively.

Previous model results of ANME-SRB aggregates have shown that thermodynamic constraints begin to affect the reaction rate when electron transfer is at < 90% DIET (He et al. 2021). To ensure that the FBA model functions similarly, a constraint was supplied here to ensure that electron transfer was occurring at comparable rates by imposing the flux through AQDS be at least 90% of the flux of methane into the cell (e.g., *lb:* $0.9F_{CH4} < F_{AQDS}$). Maintenance energy requirements for the cell were also considered as a constraint, such that *lb: min. power req.* $< F_{ATP}$, where F_{ATP} is the rate of ATP consumption expressed in terms of the minimum power requirement associated with maintaining cell functions based on values for cells of similar volumes (Kempes et al. 2017).

Protein leakage was also explored as a possible method for reducing model growth efficiencies and stimulation of nitrogen fixation activity. Leakage was assumed to be proportional to growth rates. Thus, it was implemented by increasing the stoichiometric coefficient for protein required in building biomass (see Appendix B), assuming that protein consumption beyond what is required for growth is being exported from the cell.

Maintaining redox homeostasis is vital for cell survival (Le Gal et al. 2021); if the intracellular environment is too strongly reduced or oxidized, it can cause changes in phenotype expression or even cell death. Aerobic ANME species have been observed to balance redox conditions in the cell during hypoxia by using extracellular electron transfer to reduce ferrihydrite minerals (Zheng et al. 2020). We further explore using the nitrogenase enzyme pathway in ANME as a redox balancing mechanism here in our FBA model. To stress the cell towards more reducing conditions, we increased fluxes through

a TCA cycle reaction by using constraints to impose minimum ferredoxin production rates (*Rxn. 3.8*) larger than those established by the baseline simulations:

$$CoA + 2\text{–}oxoglutarate + 2Fdx_{ox} \rightarrow CO_2 + H^+ + Succinyl\text{–}CoA + 2Fdx_{red} \quad \textit{Rxn 3.8}$$

Creating a more reduced environment may then cause the cell to stimulate the N_2 fixation pathway, consuming the reduced ferredoxin (*Rxn. 3.5*). Hydrogenase knockout simulations were also conducted by shutting down fluxes through that reaction pathway, reducing the number of reactions where ferredoxin may enter, considering that ANME may not have the capability to use membrane-bound hydrogenases for ferredoxin cycling (Chadwick et al. 2022).

Results

3D Reactive-Transport Model

Ammonium uptake. Building on the model foundations set by He et al. (2021), ammonium uptake was incorporated to explicitly simulate anabolism. NH_4^+ assimilation is modeled via growth efficiencies that are computed from the energies of catabolism, anabolism, and energy dissipation due to growth, as shown in *Eq. 3.10*. The resulting growth efficiencies (Figure 3.3) agree with experimentally determined values of 1% of the methane oxidized ending up in cellular biomass (Nauhaus et al. 2007; Orphan et al. 2009).

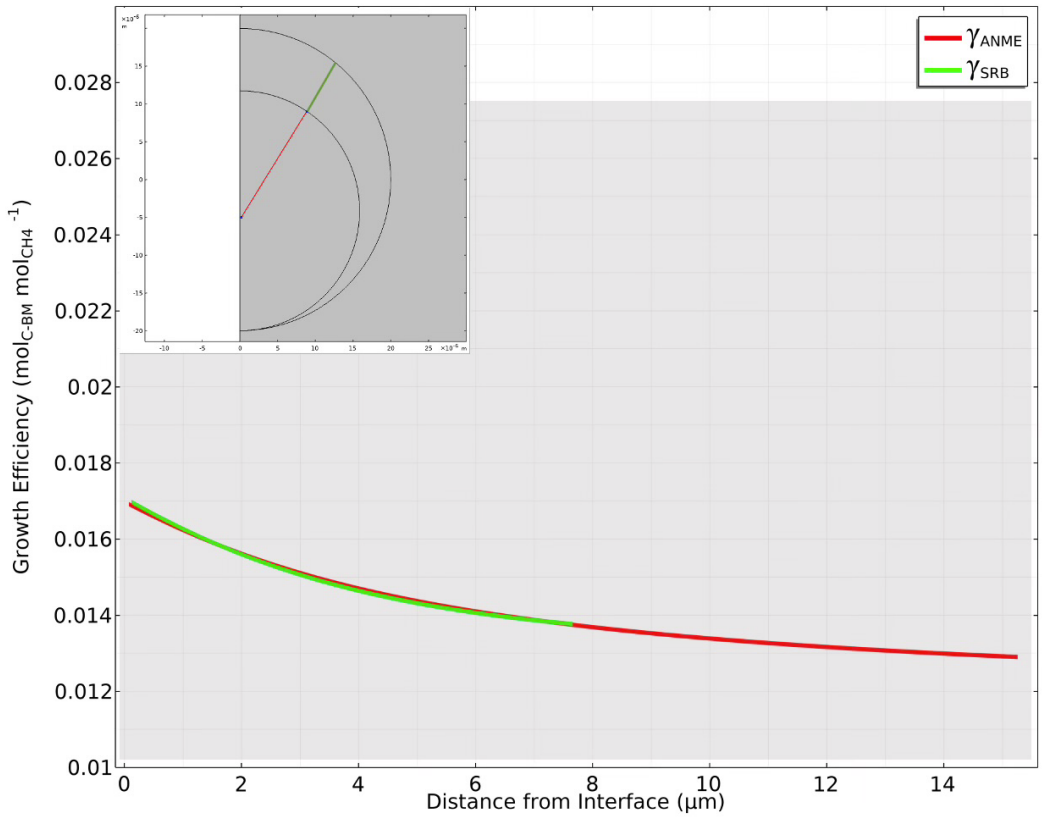

Fig 3.3. Growth efficiency across the ANME domain. Inset shows a cross-section of the model aggregate where the distance measurements were taken; inset axes depict length scales in units of 10^{-6} m. X-axis refers to the spatial distance from the interspecies interface between the ANME and SRB Shaded region represents the range of observed growth yields that have been published (0.0103-0.0275 $mol_{C\text{-}BM}$ mol_{CH4}^{-1}; Orphan et al. 2009; Nauhaus et al. 2007; Girguis et al. 2005). γ: growth efficiency

Spatial analysis of these model results shows little variance across the aggregate, supporting prior work in which ammonium uptake and thus growth efficiency was assumed to be constant (He et al. 2019; 2021). Because ANME-SRB consortia are slow growing (e.g., doubling times on the order of months as reported in Orphan et al. 2009), their N demands are very small. In agreement with back of the envelope calculations (Appendix B), the model shows that the amount of ammonium uptake due to cell anabolism results in a drop of 1.6 pM, demonstrating there is essentially no drawdown of NH_4^+ due to growth.

N_2 fixation to support growth. Growth efficiencies were calculated using N_2 fixation derived N for biosynthesis (*Eq. 3.12*), resulting in about a 16-17% reduction than when using NH_4^+ as the N source (Figure 3.4). The Gibbs free energy of reaction for N_2 fixation ΔG_{nfx} was also explored while considering different fates of reaction products

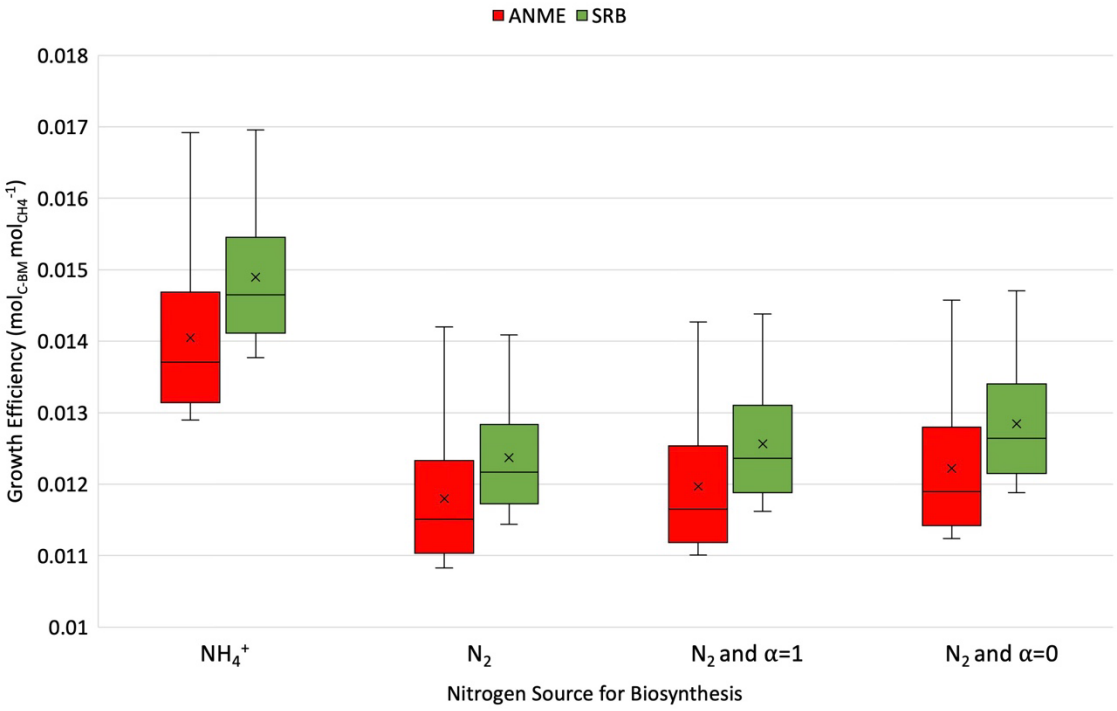

Fig 3.4. Comparison of growth efficiencies using NH_4^+ and N_2 as N sources for biosynthesis. N_2 use was also amended with energy conservation terms (*Rxns. 3.6 & 3.7*). Range of values corresponds to distance across aggregate from the interspecies interface (Figure 3.2 inset). α: coefficient for energy conservation strategy; when set to 1, *Rxn 3.6* is used and if set to 0, *Rxn 3.7* is used. x represents mean growth efficiencies; box lines represent median values.

(e.g., H_2 and ferredoxin; *Rxns 3.6 & 3.7*). Growth efficiencies were reduced when using N_2 fixation with either pathway $\alpha = 0$ or $\alpha = 1$ by 13% and 14% respectively (Figure 3.4).

N_2 fixation driving N compound leakage. The potential for N_2 fixation to support production of excess reduced nitrogen sources was also investigated in our 3D reactive transport model. Production of the AA arginine was quantified using Eq. 3.14 and then

allowed to leak out of the cells into the surrounding aggregate. This leads to a distribution of ^{15}N label that gets spatially distributed, as opposed to being localized to the cell that fixes the N_2 (Figure 3.5). Based on the Arg export model implemented here, the excess AA concentration due to leakage in the aggregate itself is only about 1 µM.

FBA Model

Growth efficiencies. Our initial flux balance model simulations using biomass production as the OF resulted in a growth efficiency of 0.523 $mol_{C\text{-}BM}$ mol_{CH4}^{-1}. To reach values more in line with observations, we introduced a maintenance power requirement that the cell must achieve. The rationale is that the cell requires a minimum energy quantum to maintain intracellular machinery aside from building new biomass. We selected a range of maintenance power requirements in terms of ATP cost based on the

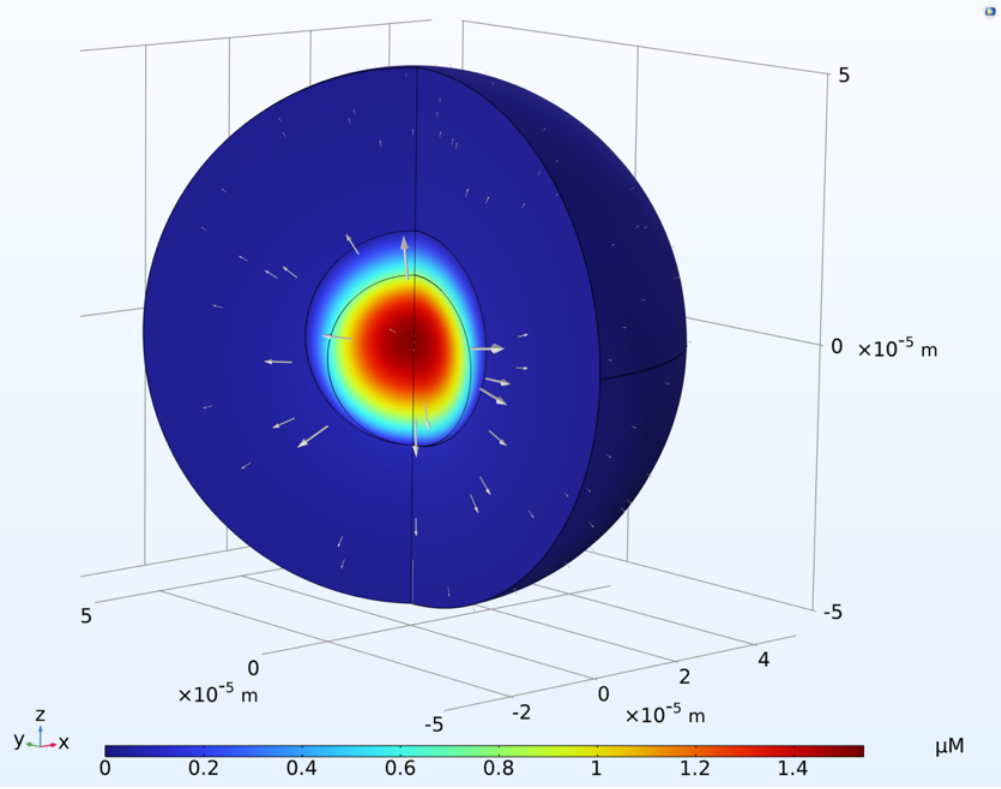

Fig 3.5. Arg production in ANME-SRB aggregate. Interior domain represents the ANME that produce the AA. Arrows indicate velocity vectors for the AA.

work conducted by Kempes et al. (2017), for cell volumes on the order of a typical ANME cell (~10^{-19} m^3). Introducing maintenance energy has a large effect on the growth efficiency, causing a reduction from 52.3% down to approximately 1% at a maintenance power requirement approaching 10^{-16} W $cell^{-1}$ (Figure 3.6).

Protein leakage. Microbial mats containing ANME-SRB consortia have been analyzed for EPS composition in biofilms, showing that a significant portion of the organic carbon leaked is contained in proteins (Krüger et al. 2008). Protein leakage was thus considered as a mechanism to influence model growth efficiencies. Growth efficiencies as a function of protein leakages relative to protein use in growth shows a pronounced decrease with increasing leakage (Figure 3.7).

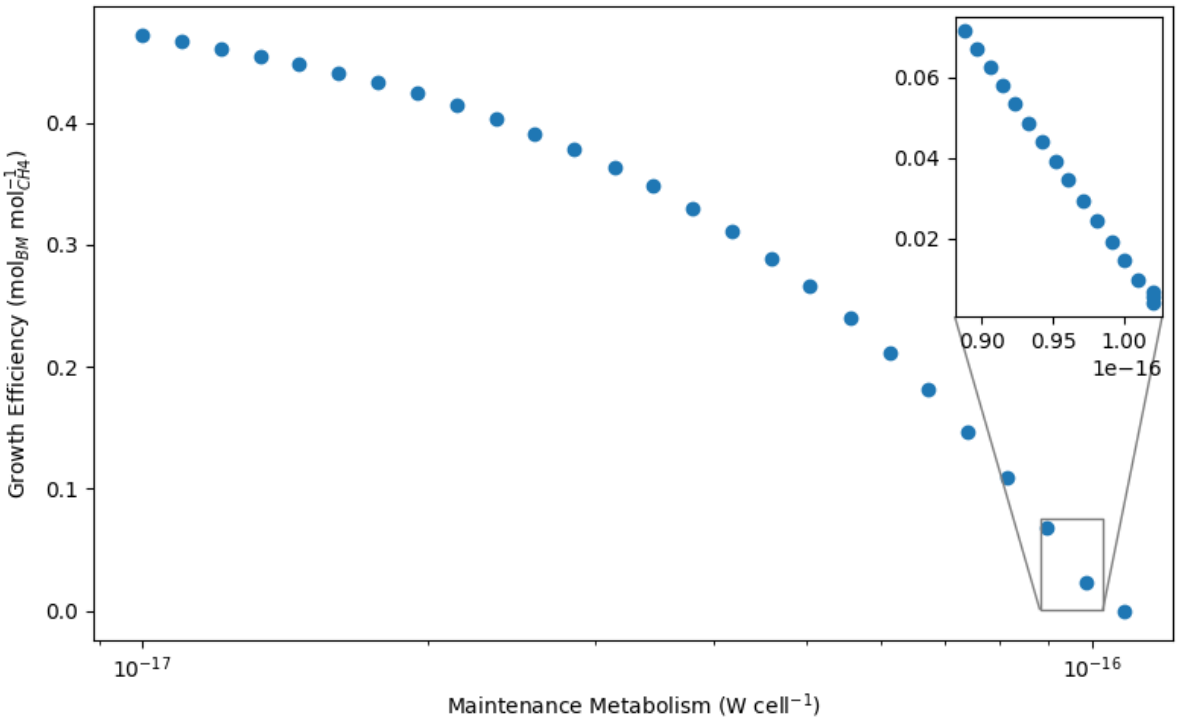

Fig 3.6. Growth efficiency as a function of maintenance metabolism power requirements. The energy threshold comes from Kempes et al. (2017) corresponding to cells with similar volume to an ANME cell. Inset plot shows a finer resolution of energy variance where the ANME growth efficiency approaches observed data.

This decrease in growth efficiency is similar to the relationship seen in Figure 3.6. However, protein leakage did not have as strong of an effect as maintenance energy did; the growth efficiency was 0.1 $mol_{C\text{-}BM}$ mol_{CH4}^{-1} at 51% increased protein demand, which is still 10x larger than the observed ANME growth efficiencies. The reduction in growth efficiency seen here is likely in large part due to the additional ATP requirement involved with protein synthesis and amino acid production in the model, but such reduction is only

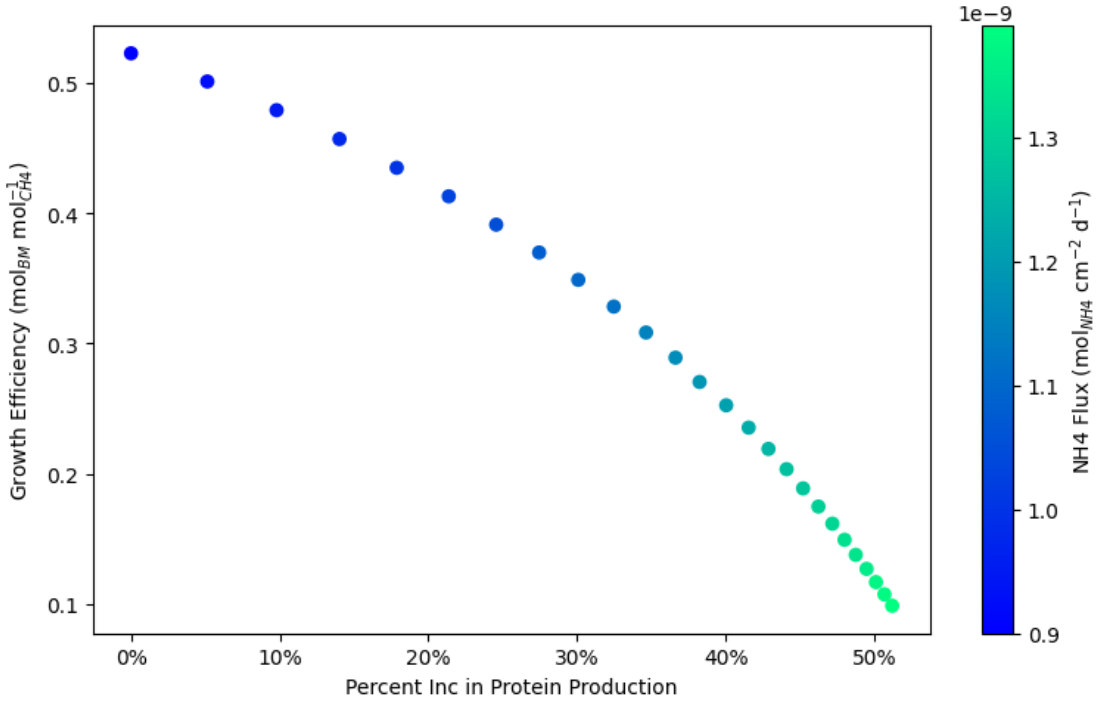

Fig 3.7. Growth efficiency as a function of protein leakage. The percent increase is the amount of protein produced beyond what is required for growth.

achieved when the ANME cell is producing 7.68-times more organic C in protein than is incorporated into biomass.

NH_4^+ limitation and N_2 fixation. Using the maintenance energy constraint that returns the accurate growth efficiencies for NH_4^+ as the nitrogen source, we also investigated triggers for N_2 fixation in ANME. To represent conditions in which

ammonium availability becomes limiting, we decreased the maximum uptake flux. Doing so resulted in net uptake of N_2 and stimulation of the nitrogenase pathway (Figure 3.8).

In the complete absence of NH_4^+, all the N requirements for growth are being fulfilled by N_2 uptake (Figure 3.8B). The growth efficiency drops by 57% with this switch in N sources (Figure 3.8A). Low ammonium availability is shown to stimulate N_2 fixation at uptake fluxes < 9.41 x 10^{-12} mol_{NH4} cm^{-2} d^{-1}.

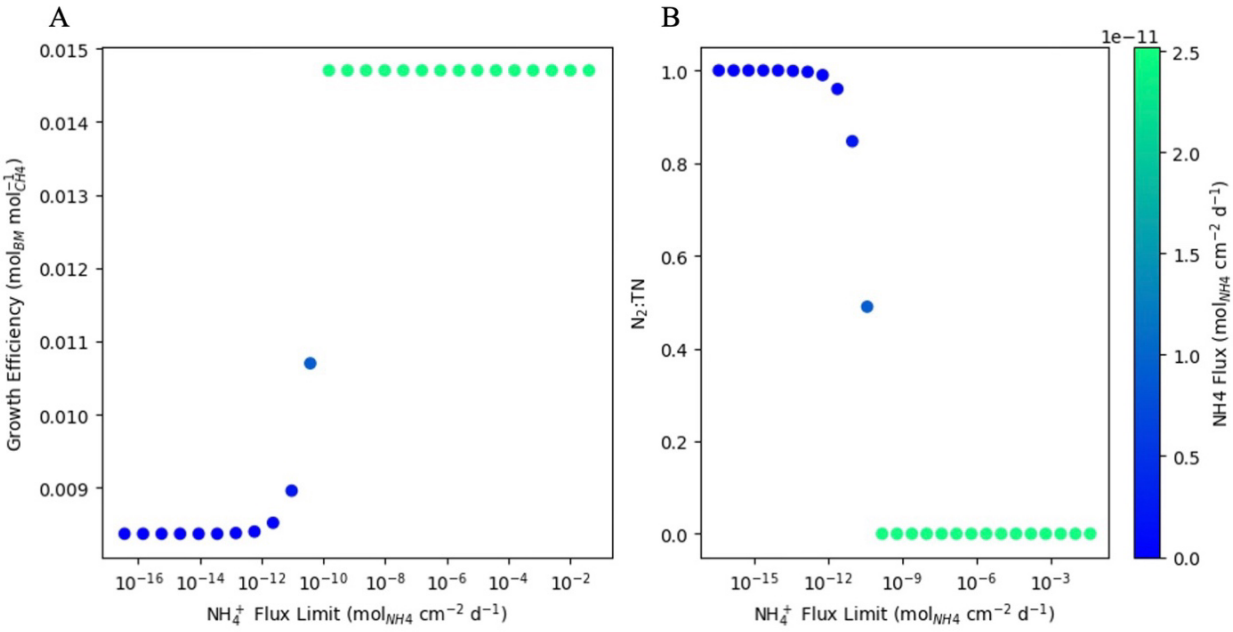

Fig 3.8. NH_4^+ limitation simulations where the flux limit imposed in the medium was gradually decreased; colors correspond to the actual uptake flux of NH_4^+ into the cell. A) growth efficiency as a function of decreasing ammonium uptake flux limits. B) N_2:TN is the proportion of the total N taken up to that of N_2.

Redox homeostasis. Increased production of reduced ferredoxin was imposed by placing greater flux constraints than the baseline model on the reaction pathway described in *Rxn 3.8*. While the goal was for the model to use the excess reducing equivalents to drive N_2 fixation, we did not see any evidence that this was the case in our ANME model here. Instead, the flux balance model used the reduced ferredoxin to generate increasing fluxes of H_2 through a hydrogenase reaction (Figure 3.9). These

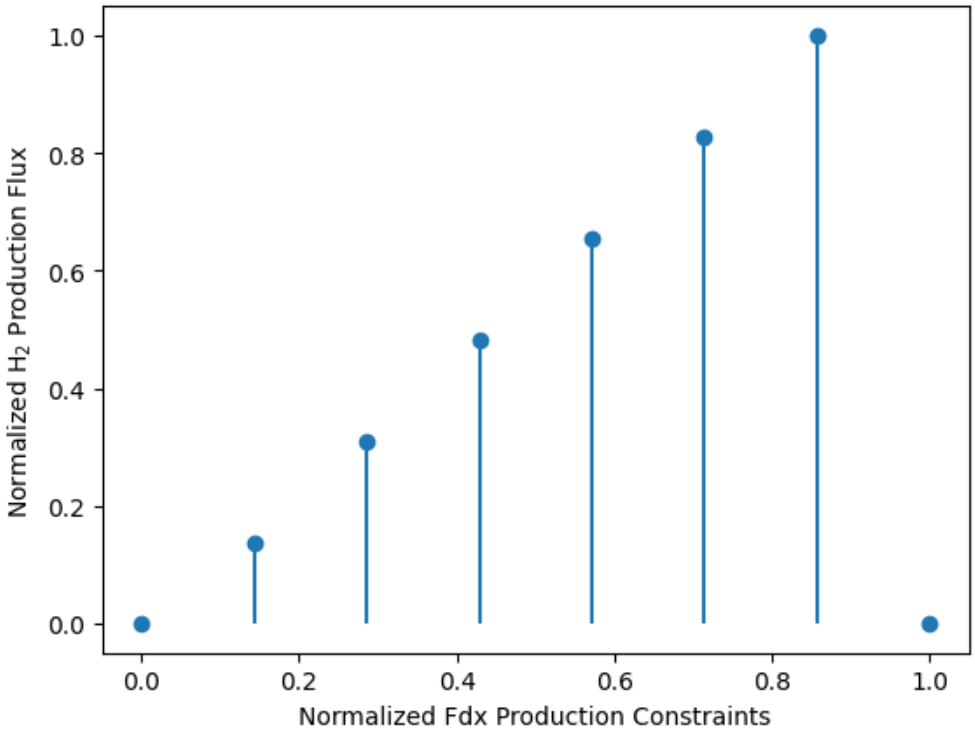

Fig 3.9. H_2 production as a function of increasing reduced ferredoxin production. Fluxes normalized using the expression $\frac{(x-x_{min})}{(x_{max}-x_{min})}$, where x is the flux of H_2 or Rxn. 8. Fdx: ferredoxin

higher H_2 fluxes occurred until the constraint imposed on *Rxn 3.8* became too large and the model could not find a solution. The additional CO_2 production, proportional to the generation of H_2, associated with this reaction further stimulated the carbon monoxide dehydrogenase pathway causing CO generation (Figure 3.10B), which has been reported to have toxic effects on ANME in higher concentrations (Meulepas et al. 2010). Simulations were also run using a knock-out on the H_2-ferredoxin pathway and increasing protein production to see if blocking the hydrogenase pathway while increasing N demands may cause the cell to pivot towards N_2 fixation (Figure 3.10B). Rather than stimulating nitrogenase, however, the cell only pushed more CO_2 and reduced ferredoxin into CO production.

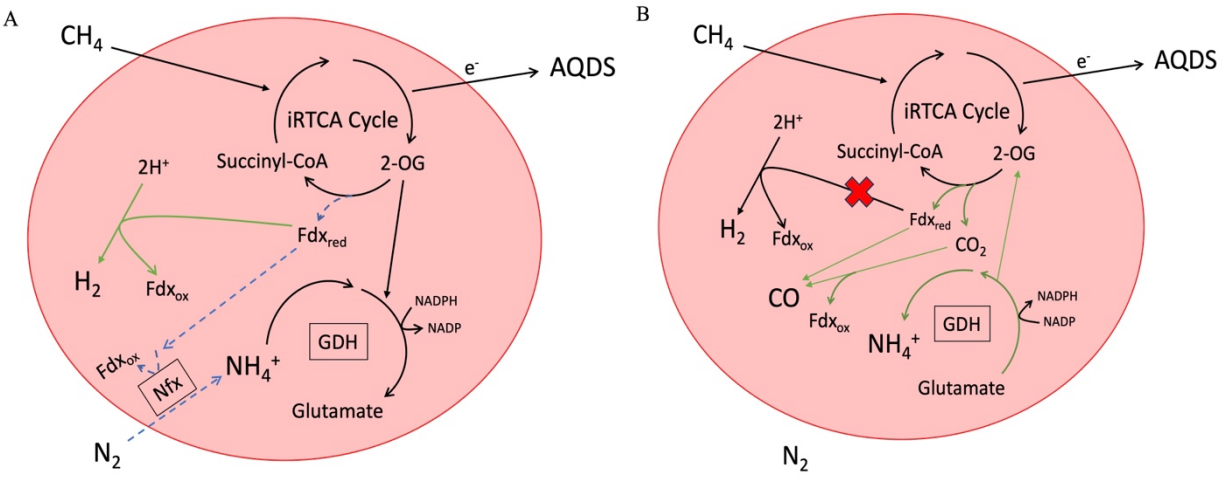

Fig 3.10. A) Diagram showing hypothesized N_2 fixation redox balance mechanism (blue dotted arrows) and the realized hydrogenase pathway used by the model (green solid arrows); B) diagram of hydrogenase knock-out simulation showing the dominant carbon monoxide dehydrogenase pathway used instead (green solid arrows); Nfx: nitrogen fixation; Fdx: ferredoxin; iRTCA: incomplete reductive tricarboxylic acid cycle; GDH: glutamate dehydrogenase; 2-OG: 2-oxoglutarate

Discussion

Growth efficiency in ANME

Through our 3D reactive transport model simulations presented here, we establish explicit descriptions of growth efficiencies in ANME-SRB aggregates that fall into the range of observed values based on experimental sediment incubations (Nauhaus et al. 2007; Orphan et al. 2009; Girguis et al. 2005). By calculating accurate reaction energetics based on activities of spatially resolved substrates, we have shown that using the relationships described by *Eqs. 3.9 & 3.10* (Heijnen et al. 1992; Heijnen & Kleerebezem 2010), growth efficiencies for slow growing microbial species can be estimated. This is notable given that this approach is established from observations of faster growing microorganisms with higher catabolic yields than those generated in AOM-associated consortia. The addition of a maintenance energy term in growth efficiency expression (*Eq. 3.9*) may provide a more physiologically accurate description for methane oxidizing

archaea, since many marine microorganisms will optimize use of organic C for maintaining cell functions over direct assimilation for biosynthesis in energy limited settings (del Giorgio et al. 1998). General descriptions of maintenance energy requirements based on catabolic reaction rates and thermodynamic yields have been derived (Tijhuis et al. 1993; Harder 1997); however, with syntrophic ANME archaea having yet to be grown in pure culture, it remains difficult to constrain how their cell-specific energetic and C yields.

There is also the possibility that the low growth efficiencies seen in the literature are in part caused by including spatial heterogeneity in activity patterns across the aggregate, where some cells may be in a dormant state that is not uncommon for microorganisms surviving in the deep biosphere (Jørgensen & Marshall 2016). However, N incorporation in ANME-SRB aggregates is nearly spatially homogenous in both experimental (McGlynn et al. 2015; Wegener et al. 2015) and modeling studies (He et al. 2019; 2021). Observations such as those lend support for there being physiological controls on energy allocation that have yet to be fully explained in microbial assemblages catalyzing the anaerobic oxidation of methane.

The need to establish a minimum energy requirement for the model ANME cell became all the more apparent when the focus shifted to investigating growth efficiency using the FBA model. Running the model without imposing constraints on their energy use showed growth efficiencies 0.5-0.6 mol_{BM} mol_{CH4}^{-1}, which is on the order of bacterial growth efficiencies calculated for photosynthetic plankton (Carlson et al. 2007; del Girogio et al. 1998). However, assimilation of over half of the consumed methane does not align with what has been shown in previous experimental studies that report growth

efficiencies of 1-2% (~0.01-0.02 $mol_{C\text{-}BM}$ mol_{CH4}^{-1}) (Nauhaus et al. 2007; Orphan et al. 2009). Given their extremely low catabolic yields estimated from our 3D reactive-transport model (e.g., -12 to -28 kJ mol_{CH4}^{-1}) and the high energetic demand associated with autotrophic growth, ANME cells are not highly optimized for growth, as has further been validated in ^{13}C assimilation studies (Wegener et al. 2008; Nauhaus et al. 2007). Therefore, maintaining their cellular machinery might be the strategy they divert the majority of their energy production to.

When a minimum maintenance power requirement on the order of 10^{-16} W $cell^{-1}$ (Kempes et al. 2017) was imposed, we did indeed see growth efficiencies more in line with those observed in bulk sediment incubations. Although, the range of maintenance metabolism requirements discussed in Kempes et al. (2017) varies by over an order of magnitude, thus a more detailed analysis is required to constrain the maintenance costs for ANME specifically. A critical next step in understanding the life history of AOM-associated consortia is establishing how power utilization is controlled within cell metabolisms, a factor that has been difficult to study in natural marine sedimentary systems (Bradley et al. 2018).

N_2 fixation impacting growth

Marine microbial groups that use diazotrophy are typically observed to be slow growing due to the high energy investment required with reducing N_2 to NH_3 (LaRoche et al. 2005; Falcón et al. 2005). While phototrophic cyanobacteria like *Trichodesmium* still have cell-specific growth rates several orders of magnitude larger than have been estimated in ANME-2/*Desulfosarcina* (DSS) consortia (e.g., *Trichodesmium:* 0.1-0.24 d^{-1} (Breitbarth et al. 2007); ANME-2/DSS: 0.003-0.008 d^{-1} (Nauhaus et al. 2007; Orphan et

al. 2009)), there is a significant reduction in growth compared to when assimilating reduced nitrogen sources directly. This was also described here in our 3D reactive-transport model, where the growth efficiencies when using NH_4^+ were reduced by 13-16% when using N_2 fixation to meet N demands, depending on which energy conservation strategy was used (Figure 3.4). The impact of N source on ANME growth efficiency was widened in the FBA model, showing a 57% reduction when N_2 fixation was at maximum (Figure 3.8A). These results demonstrate a substantial change to the potential growth of ANME-SRB aggregates and one that is qualitatively expected due to the large ATP investment required for using nitrogenase; however, nanoSIMS imagery of $^{15}N_2$ incorporation into aggregates has been reported to cause a 20-times reduction in growth than when assimilating NH_4^+ for biosynthesis (Dekas et al. 2009). This large of an impact on growth was not observed in either of the modeling studies here, leading to questions about the purpose of N_2 fixation in ANME-SRB aggregates.

Perhaps the answer lies in the sharing of reduced N sources between syntrophic partners, as proposed in the Dekas et al. (2009) study, causing reduced growth throughout the aggregate. We implemented this idea in our 3D reactive-transport model, where leakage of 0.1 mol N per mol CH_4 oxidized was required to achieve growth efficiencies on the order of their observations. This is a very large amount of N being fixed and leaked, considering their small ammonium demands required for growth calculated here. The FBA simulations show reduced growth efficiencies at over 50% increased protein production, but this still exceeds empirical values for ANME-SRB consortia growth efficiencies by 10x. These results suggest that protein and amino acid leakage is only partially responsible for driving down growth efficiencies.

Previous growth studies on C and N incorporation into ANME-SRB aggregates have indicated that much of the *de novo* protein production is not strongly linked to growth (Orphan et al. 2009) which contrasts with our implementation. There is also evidence that a significant portion of the EPS content in microbial mats containing ANME-SRB aggregates is protein (Krüger et al. 2008). This may be an indication that N_2 fixation serves as a method to supplement the surrounding syntrophic community with higher-quality bioavailable C and N sources for growth. This would likely alleviate some of the impact of the 'net oxidation effect' that ANME encounter due to CH_4 being highly reduced, requiring them to also assimilate CO_2 for biosynthesis (Nauhaus et al. 2007; Wegener et al. 2008). A possible situation could be that a subpopulation of N_2 fixers actually alleviates the energetic and C cost burden for autotrophic growth in the aggregate by producing AAs or proteins, similar to aggregations of small diazotrophic cyanobacteria (Masuda et al. 2020). NanoSIMS imagery of ^{15}N incorporation into biomass following sediment incubations with labeled N_2 often shows a slight enrichment gradient from the aggregate interior to its edge, suggesting that diazotrophs at the center of the consortia may in fact release a portion of reduced ^{15}N sources for use by neighboring cells (Dekas et al. 2009; Metcalfe et al. 2021).

NH_4^+ limitation stimulating N_2 fixation

Previous studies of diazotrophy in ANME-SRB aggregates have suggested the role of N_2 fixation is due to the impact of NH_4^+ limitation within the surrounding porewater (Dekas et al. 2018; Metcalfe et al. 2021). Using the flux at which N_2 fixation was stimulated in our FBA ammonium limitation simulations (9.41×10^{-12} mol_{NH4} cm^{-2} d^{-1}) and Fick's law, assuming a diffusive boundary layer thickness of 0.1 mm, this flux can

be translated into a concentration of approximately 0.11 nM. This is significantly larger than the 1.6 pM ammonium uptake for growth calculated in our 3D reactive-transport model; however, encountering ammonium diffusive limitation at this level seems highly unlikely in sediments where diazotrophic ANME have been found, given that the ammonium porewater concentrations are on the order of tens to hundreds of μM (Dekas et al. 2018).

In general, porewater in marine sediments is often replete with ammonium up into the mM range depending on the rates of loading and remineralization of organic matter (Kristensen et al. 1999; Brodecka-Goluch et al. 2019). Previous estimates of ammonium inhibition thresholds in benthic diazotrophs began at ≥ 100 μM NH_4^+ (Kessler et al. 2001; Knapp 2012), however, experiments by Dekas et al. (2018) on nitrogenase sensitivity showed activation of N_2 fixation at < 25 μM NH_4^+. While it has been posited that intra-aggregate heterogeneity in ammonium availability may be possible (Metcalfe et al. 2021), the short diffusive length scales coupled with the ANME's slow growth argues against NH_4^+ limitation being the main driver of N_2 fixation, especially given the demands for N we have observed in our models here. Furthermore, evidence that ANME-associated SRB are also able to fix nitrogen (Dekas et al. 2016; Metcalfe et al. 2021) begs the question as to why both consortia members would require the expression of nitrogenase despite the often plentiful local N conditions found in natural sediments.

Redox homeostasis

Because of the unlikely substrate limitation given the low NH_4^+ demands for growth calculated here, we speculate the role of diazotrophic activity in ANME-SRB consortia may be to maintain an intracellular redox balance. A similar mechanism has

been suggested by recent evidence in *Geobacter sulfurreducens*, where high-energy electrons generated in C fixation are used to reduce N_2 and protons to NH_4^+ and H_2 respectively under low-energy conditions (Ortiz-Medina et al. 2023). Redox balancing during C fixation using nitrogenase regulation is also described in *Rhodospirillum rubrum* (Wang et al. 2010) and in *Azotobacter vinelandii* by altering the distribution of carbon between different pools and affecting the internal C/N ratio (Inomura et al. 2018). These previous two studies also provide evidence that simultaneous assimilation of reduced N sources while fixing N_2 may be a viable strategy to alleviate some of the energetic burdens, lending support to the idea that ANME-SRB aggregates may still fix nitrogen under ammonium-replete conditions. AOM-associated consortia often maintain their metabolic activity across the entire aggregate (Orcutt & Meile 2008; McGlynn et al. 2015; He et al. 2019), possibly contributing to a situation where excess reducing power in the form of electrons generated may be a concern for the cell, requiring a method for rebalancing.

Our FBA simulations presented here lend support that internal redox balancing in ANME leads to increased H_2 production in the cell (Figure 3.9). However, this is due to the activation of a hydrogenase pathway that is often seen in methanogens which may not be conserved in ANME given recent genomic evidence (Chadwick et al. 2022). Knock-out simulations for the hydrogenase pathway also did not stimulate N_2 fixation within the cell, leading instead to further CO production. Despite increasing protein production as well to induce the need for more N, this likely only served to drive down ATP availability causing a disadvantageous situation for the cell to use N_2 fixation in tandem. In either case, generating excess CO as a method for offloading electrons will lead to

adverse effects for the cell, given its recorded toxic effects on ANME and SRB (Meulepas et al. 2010; Van Houten et al. 1995).

We cannot be certain of the intracellular conditions here, specifically the pH, as FBA models do not account for concentrations; however, it is not uncommon for there to be low-pH settings in sediments where ANME-SRB aggregates are found (Yanagawa et al. 2013). Acidic conditions have been seen to promote higher redox states and subsequent increases in H_2 production in anaerobic bacteria cultures (Nakashimada et al. 2002); thus, if a buildup of protons occurs in the periplasmic space of ANME cells, this may require additional H_2 production to maintain homeostasis through the use of nitrogenase given the lack of membrane-bound hydrogenases (Chadwick et al. 2022). The exact mechanism for proton pumping in ANME is also still not fully understood (McGlynn et al. 2015; 2017), leaving room to speculate if perhaps N_2 fixation aids in regulating intracellular pH and redox state through H_2 production. This idea may be further supported by recent transcriptomic evidence that indicates the upregulation of nitrogenase expression in some ANME when using AQDS for electron capture (Yu et al. 2022).

Conclusions

The data from the two models presented here shows agreement with observed growth efficiencies in ANME-SRB aggregates catalyzing anaerobic methane oxidation (Girguis et al. 2005; Nauhaus et al. 2007; Orphan et al. 2009). Additional work will need to be conducted to elucidate the energetic drivers causing this extreme reduction in growth efficiency and C assimilation relative to other faster-growing microorganisms,

whether that is through intracellular maintenance requirement costs or additional biosynthesis. Understanding the mechanisms behind the growth strategies of AOM-associated consortia will provide vital knowledge for how life at the energetic edge thrives and further inform their impact on the biogeochemical turnover of methane in marine sediments.

Investigations of N_2 fixation activity in our models here show that while we do see an impact on growth associated with the energetic impact derived from using nitrogenase, these results do not align with previous estimates of ANME-SRB aggregate growth on N_2 (Dekas et al. 2009). Ammonium limitation does stimulate N_2 activity, as is commonly observed in diazotrophs (Knapp 2012), but this is likely not the sole driver for the expression of nitrogenase enzymes in ANME-SRB given the often high availability of reduced N sources found around methane cold seeps (Joye et al. 2004; Dekas et al. 2018). Future studies should focus on the intracellular conditions leading to nitrogenase activity in ANME-SRB aggregates, such as further testing the sensitivity of the enzyme to ammonium as has been previously studied in sediment incubations (Dekas et al. 2018) and its potential role in maintaining redox conditions. In light of the absence of ANME pure cultures, conducting studies on their internal mechanisms may prove difficult; however, uncovering the role that N_2 fixation serves in ANME-SRB aggregates will help to bolster our understanding of benthic marine diazotrophy and its purpose beyond the production of reduced N sources.

References

Alberty, R. A. (1994). Thermodynamics of the nitrogenase reactions. *Journal of Biological Chemistry*, *269*(10), 7099-7102.

Boetius, A., Ravenschlag, K., Schubert, C. J., Rickert, D., Widdel, F., Gieseke, A., ... & Pfannkuche, O. (2000). A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature*, *407*(6804), 623-626.

Boudreau, B. P. (1997). *Diagenetic models and their implementation* (Vol. 505). Berlin: Springer.

Bowles, M. W., Samarkin, V. A., Hunter, K. S., Finke, N., Teske, A. P., Girguis, P. R., & Joye, S. B. (2019). Remarkable capacity for anaerobic oxidation of methane at high methane concentration. *Geophysical Research Letters*, *46*(21), 12192-12201.

Bradley, J. A., Amend, J. P., & LaRowe, D. E. (2018). Bioenergetic controls on microbial ecophysiology in marine sediments. *Frontiers in Microbiology*, *9*, 180.

Breitbarth, E., Oschlies, A., & LaRoche, J. (2007). Physiological constraints on the global distribution of Trichodesmium–effect of temperature on diazotrophy. *Biogeosciences*, *4*(1), 53-61.

Brodecka-Goluch, A., Siudek, P., & Bolałek, J. (2019). Impact of sampling techniques on the concentration of ammonia and sulfide in pore water of marine sediments. *Oceanological and Hydrobiological Studies*, *48*(2), 184-195.

Carlson, C. A., Del Giorgio, P. A., & Herndl, G. J. (2007). Microbes and the dissipation of energy and respiration: from cells to ecosystems. *Oceanography*, *20*(2), 89-100.

Chadwick, G. L., Skennerton, C. T., Laso-Pérez, R., Leu, A. O., Speth, D. R., Yu, H., ... & Orphan, V. J. (2022). Comparative genomics reveals electron transfer and syntrophic mechanisms differentiating methanotrophic and methanogenic archaea. *PLoS Biology*, *20*(1), e3001508.

Del Giorgio, P. A., & Cole, J. J. (1998). Bacterial growth efficiency in natural aquatic systems. *Annual Review of Ecology and Systematics*, *29*(1), 503-541.

Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-sea archaea fix and share nitrogen in methane consuming microbial consortia. *Science*, *326*(5951), 422-426.

Dekas, A. E., Chadwick, G. L., Bowles, M. W., Joye, S. B., & Orphan, V. J. (2014). Spatial distribution of nitrogen fixation in methane seep sediment and the role of the ANME archaea. *Environmental Microbiology*, *16*(10), 3012-3029.

Dekas, A. E., Connon, S. A., Chadwick, G. L., Trembath-Reichert, E., & Orphan, V. J. (2016). Activity and interactions of methane seep microorganisms assessed by parallel transcription and FISH NanoSIMS analyses. *The ISME Journal*, *10*(3), 678-692.

Dekas, A. E., Fike, D. A., Chadwick, G. L., Green-Saxena, A., Fortney, J., Connon, S. A., ... & Orphan, V. J. (2018). Widespread nitrogen fixation in sediments from diverse deep-sea sites of elevated carbon loading. *Environmental Microbiology*, *20*(12), 4281-4296.

Deppenmeier, U. (2002). Redox-driven proton translocation in methanogenic Archaea. *Cellular and Molecular Life Sciences CMLS*, *59*, 1513-1533.

Dickson, A. G., & Goyet, C. (1994). *Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water. Version 2* (No. ORNL/CDIAC-74). Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States).

Egger, M., Riedinger, N., Mogollón, J. M., & Jørgensen, B. B. (2018). Global diffusive fluxes of methane in marine sediments. *Nature Geoscience*, *11*(6), 421-425.

Falcón, L. I., Pluvinage, S., & Carpenter, E. J. (2005). Growth kinetics of marine unicellular N2-fixing cyanobacterial isolates in continuous culture in relation to phosphorus and temperature. *Marine Ecology Progress Series*, *285*, 3-9.

Faria, J. P., et al. (2023). "ModelSEED v2: High-throughput genome-scale metabolic model reconstruction with enhanced energy biosynthesis pathway prediction." bioRxiv: 2023.2010.2004.556561.

Girguis, P. R., Cozen, A. E., & DeLong, E. F. (2005). Growth and population dynamics of anaerobic methane-oxidizing archaea and sulfate-reducing bacteria in a continuous-flow bioreactor. *Applied and Environmental Microbiology*, *71*(7), 3725-3733.

González-Cabaleiro, R., Ofiţeru, I. D., Lema, J. M., & Rodríguez, J. (2015). Microbial catabolic activities are naturally selected by metabolic energy harvest rate. *The ISME Journal*, *9*(12), 2630-2641.

Harder, J. (1997). Species-independent maintenance energy and natural population sizes. *FEMS Microbiology Ecology*, *23*(1), 39-44.

Heijnen, J. J., & Kleerebezem, R. (2010). Bioenergetics of microbial growth. Encyclopedia of Industrial Biotechnology: Bioprocess. *Bioseparation and Cell Technology2010*, 1-24.
Heijnen, J. J., & Van Dijken, J. P. (1992). In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnology and Bioengineering*, *39*(8), 833-858.
He, X., Chadwick, G., Kempes, C., Shi, Y., McGlynn, S., Orphan, V., & Meile, C. (2019). Microbial interactions in the anaerobic oxidation of methane: model simulations constrained by process rates and activity patterns. *Environmental Microbiology*, *21*(2), 631-647.
He, X., Chadwick, G. L., Kempes, C. P., Orphan, V. J., & Meile, C. (2021). Controls on interspecies electron transport and size limitation of anaerobically methane-oxidizing microbial consortia. *Mbio*, *12*(3), 10-1128.
Hubbell, D. H., & Kidder, G. (2009). Univ Fla IFAS Ext Publ SL16. *Biological Nitrogen Fixation*.
Inomura, K., Bragg, J., Riemann, L., & Follows, M. J. (2018). A quantitative model of nitrogen fixation in the presence of ammonium. *PLoS One*, *13*(11), e0208282.
Jin, Q., & Bethke, C. M. (2003). A new rate law describing microbial respiration. *Applied and Environmental Microbiology*, *69*(4), 2340-2348.
Jin, Q., & Bethke, C. M. (2005). Predicting the rate of microbial respiration in geochemical environments. *Geochimica et Cosmochimica Acta*, *69*(5), 1133-1143.
Jin, Q., & Bethke, C. M. (2007). The thermodynamics and kinetics of microbial metabolism. *American Journal of Science*, *307*(4), 643-677.
Jørgensen, B. B., & Des Marais, D. J. (1990). The diffusive boundary layer of sediments: oxygen microgradients over a microbial mat. *Limnology and Oceanography*, *35*(6), 1343-1355.
Jørgensen, B. B., & Marshall, I. P. (2016). Slow microbial life in the seabed. *Annual Review of Marine Science*, *8*, 311-332.
Joye, S. B., Boetius, A., Orcutt, B. N., Montoya, J. P., Schulz, H. N., Erickson, M. J., & Lugo, S. K. (2004). The anaerobic oxidation of methane and sulfate reduction in sediments from Gulf of Mexico cold seeps. *Chemical Geology*, *205*(3-4), 219-238.
Kempes, C. P., van Bodegom, P. M., Wolpert, D., Libby, E., Amend, J., & Hoehler, T. (2017). Drivers of bacterial maintenance and minimal energy requirements. *Frontiers in Microbiology*, *8*, 31.
Kessler, P. S., Daniel, C., & Leigh, J. A. (2001). Ammonia switch-off of nitrogen fixation in the methanogenic archaeon Methanococcus maripaludis: mechanistic features and requirement for the novel GlnB homologues, NifI1 and NifI2. *Journal of Bacteriology*, *183*(3), 882-889.
Knapp, A. N. (2012). The sensitivity of marine N2 fixation to dissolved inorganic nitrogen. *Frontiers in Microbiology*, *3*, 374.
Knittel, K., Lösekann, T., Boetius, A., Kort, R., & Amann, R. (2005). Diversity and distribution of methanotrophic archaea at cold seeps. *Applied and Environmental Microbiology*, *71*(1), 467-479.
Korth, B., Rosa, L. F., Harnisch, F., & Picioreanu, C. (2015). A framework for modeling electroactive microbial biofilms performing direct electron transfer. *Bioelectrochemistry*, *106*, 194-206.
Kristensen, E., & Hansen, K. (1999). Transport of carbon dioxide and ammonium in bioturbated (Nereis diversicolor) coastal, marine sediments. *Biogeochemistry*, *45*, 147-168.
Krüger, M., Blumenberg, M., Kasten, S., Wieland, A., Känel, L., Klock, J. H., ... & Seifert, R. (2008). A novel, multi-layered methanotrophic microbial mat system growing on the sediment of the Black Sea. *Environmental Microbiology*, *10*(8), 1934-1947.
Le Gal, K., Schmidt, E. E., & Sayin, V. I. (2021). Cellular redox homeostasis. *Antioxidants*, *10*(9), 1377.
Luxem, K. E., Kraepiel, A. M., Zhang, L., Waldbauer, J. R., & Zhang, X. (2020). Carbon substrate re-orders relative growth of a bacterium using Mo-, V-, or Fe-nitrogenase for nitrogen fixation. *Environmental Microbiology*, *22*(4), 1397-1408.
Ma, Y., Zhu, C., Ma, P., & Yu, K. T. (2005). Studies on the diffusion coefficients of amino acids in aqueous solutions. *Journal of Chemical & Engineering Data*, *50*(4), 1192-1196.
Malvankar, N. S., Vargas, M., Nevin, K. P., Franks, A. E., Leang, C., Kim, B. C., ... & Lovley, D. R. (2011). Tunable metallic-like conductivity in microbial nanowire networks. *Nature Nanotechnology*, *6*(9), 573-579.
Masuda, T., Inomura, K., Takahata, N., Shiozaki, T., Sano, Y., Deutsch, C., ... & Furuya, K. (2020). Heterogeneous nitrogen fixation rates confer energetic advantage and expanded ecological niche of unicellular diazotroph populations. *Communications Biology*, *3*(1), 172.
McGlynn, S. E., Chadwick, G. L., Kempes, C. P., & Orphan, V. J. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature*, *526*(7574), 531-535.

McGlynn, S. E. (2017). Energy metabolism during anaerobic methane oxidation in ANME archaea. *Microbes and Environments*, *32*(1), 5-13.
Metcalfe, K. S., Murali, R., Mullin, S. W., Connon, S. A., & Orphan, V. J. (2021). Experimentally-validated correlation analysis reveals new anaerobic methane oxidation partnerships with consortium-level heterogeneity in diazotrophy. *The ISME Journal*, *15*(2), 377-396.
Meulepas, R. J., Jagersma, C. G., Khadem, A. F., Stams, A. J., & Lens, P. N. (2010). Effect of methanogenic substrates on anaerobic oxidation of methane and sulfate reduction by an anaerobic methanotrophic enrichment. *Applied Microbiology and Biotechnology*, *87*, 1499-1506.
Murali, R., Yu, H., Speth, D. R., Wu, F., Metcalfe, K. S., Crémière, A., ... & Orphan, V. J. (2023). Physiological potential and evolutionary trajectories of syntrophic sulfate-reducing bacterial partners of anaerobic methanotrophic archaea. *PLoS Biology*, *21*(9), e3002292.
Nakashimada, Y., Rachman, M. A., Kakizono, T., & Nishio, N. (2002). Hydrogen production of Enterobacter aerogenes altered by extracellular and intracellular redox states. *International Journal of Hydrogen Energy*, *27*(11-12), 1399-1405.
Nauhaus, K., Treude, T., Boetius, A., & Krüger, M. (2005). Environmental regulation of the anaerobic oxidation of methane: a comparison of ANME-I and ANME-II communities. *Environmental Microbiology*, *7*(1), 98-106.
Nauhaus, K., Albrecht, M., Elvert, M., Boetius, A., & Widdel, F. (2007). In vitro cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate. *Environmental Microbiology*, *9*(1), 187-196.
Orcutt, B., & Meile, C. (2008). Constraints on mechanisms and rates of anaerobic oxidation of methane by microbial consortia: process-based modeling of ANME-2 archaea and sulfate reducing bacteria interactions. *Biogeosciences*, *5*(6), 1587-1599.
Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D., & DeLong, E. F. (2002). Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proceedings of the National Academy of Sciences*, *99*(11), 7663-7668.
Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis?. *Nature Biotechnology*, *28*(3), 245-248.
Ortiz-Medina, J. F., Poole, M. R., Grunden, A. M., & Call, D. F. (2023). Nitrogen Fixation and Ammonium Assimilation Pathway Expression of Geobacter sulfurreducens Changes in Response to the Anode Potential in Microbial Electrochemical Cells. *Applied and Environmental Microbiology*, *89*(4), e02073-22.
Popovic, M. (2019). Thermodynamic properties of microorganisms: determination and analysis of enthalpy, entropy, and Gibbs free energy of biomass, cells and colonies of 32 microorganism species. *Heliyon*, *5*(6).
Roychoudhury, A. N. (2004). Sulfate respiration in extreme environments: a kinetic study. *Geomicrobiology Journal*, *21*(1), 33-43.
Scheller, S., Yu, H., Chadwick, G. L., McGlynn, S. E., & Orphan, V. J. (2016). Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science*, *351*(6274), 703-707.
Schink, B. (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews*, *61*(2), 262-280.
Schulz, H. D. (2000). Quantification of early diagenesis: dissolved constituents in marine pore water. In *Marine geochemistry* (pp. 85-128). Berlin, Heidelberg: Springer Berlin Heidelberg.
Storck, T., Virdis, B., & Batstone, D. J. (2016). Modelling extracellular limitations for mediated versus direct interspecies electron transfer. *The ISME Journal*, *10*(3), 621-631.
Stewart, P. S. (2003). Diffusion in biofilms. *Journal of Bacteriology*, *185*(5), 1485-1491.
Strycharz-Glaven, S. M., Snider, R. M., Guiseppi-Elie, A., & Tender, L. M. (2011). On the electrical conductivity of microbial nanowires and biofilms. *Energy & Environmental Science*, *4*(11), 4366-4379.
Thauer, R. K., Jungermann, K., & Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews*, *41*(1), 100-180.
Tijhuis, L., Van Loosdrecht, M. C., & Heijnen, J. (1993). A thermodynamically based correlation for maintenance Gibbs energy requirements in aerobic and anaerobic chemotrophic growth. *Biotechnology and Bioengineering*, *42*(4), 509-519.

Van Houten, R. T., Van der Spoel, H., van Aelst, A. C., Hulshoff Pol, L. W., & Lettinga, G. (1996). Biological sulfate reduction using synthesis gas as energy and carbon source. *Biotechnology and Bioengineering*, *50*(2), 136-144.

Wang, D., Zhang, Y., Welch, E., Li, J., & Roberts, G. P. (2010). Elimination of Rubisco alters the regulation of nitrogenase activity and increases hydrogen production in Rhodospirillum rubrum. *International Journal of Hydrogen Energy*, *35*(14), 7377-7385.

Wegener, G., Niemann, H., Elvert, M., Hinrichs, K. U., & Boetius, A. (2008). Assimilation of methane and inorganic carbon by microbial communities mediating the anaerobic oxidation of methane. *Environmental Microbiology*, *10*(9), 2287-2298.

Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E., & Boetius, A. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature*, *526*(7574), 587-590.

Yanagawa, K., Morono, Y., De Beer, D., Haeckel, M., Sunamura, M., Futagami, T., ... & Inagaki, F. (2013). Metabolically active microbial communities in marine sediment under high-CO2 and low-pH extremes. *The ISME Journal*, *7*(3), 555-567.

Yongquan, Z. H. O. U., Chunhui, F. A. N. G., Yan, F. A. N. G., & Fayan, Z. H. U. (2013). Volumetric and transport properties of aqueous NaB (OH) 4 solutions. *Chinese Journal of Chemical Engineering*, *21*(9), 1048-1056.

Yu, H., Skennerton, C. T., Chadwick, G. L., Leu, A. O., Aoki, M., Tyson, G. W., & Orphan, V. J. (2022). Sulfate differentially stimulates but is not respired by diverse anaerobic methanotrophic archaea. *The ISME Journal*, *16*(1), 168-177.

Zeebe, R. E., & Wolf-Gladrow, D. (2001). *CO2 in seawater: equilibrium, kinetics, isotopes* (No. 65). Gulf Professional Publishing.

Zheng, Y., Wang, H., Liu, Y., Zhu, B., Li, J., Yang, Y., ... & Zhao, F. (2020). Methane-dependent mineral reduction by aerobic methanotrophs under hypoxia. *Environmental Science & Technology Letters*, *7*(8), 606-612.

CHAPTER 4

SUMMARY

Through the work presented here, we have shown that symbolic regression can successfully reconstruct simple concentration-dependent rate expressions from co-registered observations of rates and chemical concentrations (C → R; Figure 2.1). In addition to the structure of the rate laws, it also provides the associated rate constants. For complex rate expressions, our exploratory work indicates that the integration of expert knowledge that prompts the use of complex operators may be necessary. This may include, for example, representations of thermodynamic constraints that indirectly depend on the local concentrations through the reaction energetics. Our results also indicate that for microbially-mediated reactions, insight into the abundance of the active microbial population catalyzing a reaction is important (B → R; Figure 2.1). Without that, the symbolic regression approach may still provide a good fit to the data but loses its interpretability and does not provide insight into the true drivers of biogeochemical process rates. Future work may aim at identifying the underlying connections between reaction energetics and the microbial population catalyzing reactions (E → B; Figure 2.1) by exploring energetic controls on microbial growth (Heijnen et al. 1992; Heijnen & Kleerebezem 1999). Additionally, performing targeted experiments in concert with ecological theory on the organization of ecological networks (Vallino 2010; Wu et al. 2022) to uncover connections between the environmental conditions and microbial dynamics (C → B; Figure 2.1).

Our modeling efforts on anaerobic methane-oxidizing consortia have further shown that we can accurately simulate ammonium-based growth efficiencies that match observed values, in both our 3D reactive-transport model and flux balance analysis. These results also indicated that ammonium limitation in ANME-SRB aggregates is an unlikely driver of N_2 fixation activity given the small N demands required for growth. While we did see a reduction in model growth efficiencies computed from the energetic impacts due to diazotrophy, experimental studies have described observed growth efficiencies substantially less than those calculated here (Dekas et al. 2018). To explore whether N_2 fixation served an alternative role for ANME-SRB consortia beyond N supplementation for growth, we simulated amino acid and protein leakage in our two models. These simulations showed no stimulation of N_2 fixation despite the increased demand for N. Our results indicate that in an ammonium-limited setting, using N_2 fixation to support protein leakage would require several orders of magnitude larger protein production to match observed growth efficiencies. We further investigated whether N_2 fixation may be used as a mechanism for intracellular redox state balancing. Our results did not reveal a stimulation of diazotrophic activity when higher fluxes of reducing equivalents were generated. However, given that nitrogenase genes are still shown to be transcribed in ammonium-replete experimental settings (Yu et al. 2022), further investigation into the use of diazotrophy for redox homeostasis is required. Future works should focus on the intracellular mechanisms for post-translational control of nitrogenase enzymes in ANME as well as patterns of gene and protein expression to better inform models like those used in this study.

References

Dekas, A. E., Fike, D. A., Chadwick, G. L., Green-Saxena, A., Fortney, J., Connon, S. A., ... & Orphan, V. J. (2018). Widespread nitrogen fixation in sediments from diverse deep-sea sites of elevated carbon loading. *Environmental Microbiology*, *20*(12), 4281-4296.

Heijnen, J. J., & Kleerebezem, R. (1999). Bioenergetics of microbial growth. *Encyclopedia of bioprocess technology: Fermentation, biocatalysis, and bioseparation*, *1*, 267-291.

Heijnen, J. J., & Van Dijken, J. P. (1992). In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnology and Bioengineering*, *39*(8), 833-858.

Vallino, J. J. (2010). Ecosystem biogeochemistry considered as a distributed metabolic network ordered by maximum entropy production. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1545), 1417-1427.

Wu, Q., Guthrie, M. J., & Jin, Q. (2022). Physiological Acclimation Extrapolates the Kinetics and Thermodynamics of Methanogenesis From Laboratory Experiments to Natural Environments. *Frontiers in Ecology and Evolution*, *10*, 838487.

Yu, H., Skennerton, C. T., Chadwick, G. L., Leu, A. O., Aoki, M., Tyson, G. W., & Orphan, V. J. (2022). Sulfate differentially stimulates but is not respired by diverse anaerobic methanotrophic archaea. *The ISME Journal*, *16*(1), 168-177.

APPENDIX A

SUPPLEMENTARY MATERIAL

CHAPTER 2: IDENTIFYING MACROSCOPIC RATE EXPRESSIONS USING SYMBOLIC REGRESSION

Model files and codes can be found at: https://bitbucket.org/MeileLab/kyle/src/master/

A.1. STEADYSED description

The model data presented here was generated using the multicomponent reaction-transport model STEADYSED described in Wang & Van Cappellen (1996). Physical and chemical parameters for this implementation are shown in Table S2.1. These correspond to Site S4 described by the source paper. The reaction network used couples transformations between C, O, N, S, Fe, and Mn compounds including organic matter remineralization and multiple secondary redox reactions.

Table S2.1. STEADYSED model parameters

Parameters		Value
Physical Parameters	Temperature (ºC)	5
	Salinity (‰)	35
	Sedimentation Rate (cm yr^{-1})	0.20
	Bulk Dry Sediment Density (g cm^{-3})	2.40
	Porosity (%)	80
Bottom Water Chemistry	O_2 (µmol L^{-1})	231
	NO_3^- (µmol L^{-1})	1.5
	Mn^{2+} (µmol L^{-1})	2.0
	Fe^{2+} (µmol L^{-1})	0
	SO_4^{2-} (mmol L^{-1})	28
	CH_4 (µmol L^{-1})	0
	NH_4^+ (µmol L^{-1})	22
	TS (mmol L^{-1})	0
	pH	7.9
	Alkalinity (meq L^{-1})	2.5

A.2 Chemostat Model

The data for our biomass-dependent reaction rates was generated using a simple chemostat system with a single population of denitrifying bacteria. Model parameters and initial conditions are listed in Table S2.2. Subsections A.2.2-A.2.6 include the scripts for the model implementation used for the results shown in the text.

A.2.1 Parameters

Table S2.2. Dentrifier Chemostat Model Parameters

Parameter	Value	Description
Temperature (ºC)	8	
Volume (m^3)	0.01	
pH	8.2	
Time (d)	1000	Time duration of simulations
Q (m^3 s^{-1})	10^{-8}	Flow rate
sAc	-1.525	Acetate used per reaction
sNO3	-1.6	Nitrate used per reaction
sB	-1	Biomass used per reaction
Acetate (mol m^{-3})	0	Initial concentration of Ac
NO_3^- (mol m^{-3})	0	Initial concentration of nitrate
B (cell m^{-3})	10^6	Initial cell concentration
k (mol $cell^{-1}$ d^{-1} or m^6 mol^{-1} $cell^{-1}$ s^{-1})	10^{-5}	Rate constant
K_m^{Ac} (mol m^{-3})	0.01	Affinity constant for Acetate
K_m^{NO3} (mol m^{-3})	0.05	Affinity constant for nitrate
Ac_{in} (mol m^{-3})	10	Inflow concentration of Ac
$NO_3^-{}_{in}$ (mol m^{-3})	20	Inflow concentration of NO_3^-
B_{in} (cell m^{-3})	0	Inflow of cells
[HCO_3^-] (mol m^{-3})	2.14	Concentration used for thermodynamic calculations (speciation calculation done at pH = 8.2)

$[CO_2]$ (mol m^{-3})	0.0136	Concentration used for thermodynamic calculations (speciation calculation done at pH = 8.2)
$[N_2]$ (mol m^{-3})	10	Concentration used for thermodynamic calculations
$[NH_4^+]$ (mol m^{-3})	1	Concentration used for thermodynamic calculations

A.2.2 main.mlx

Main script for the chemostat model implemented in MATLAB. "savevar" data frames were exported to be used as data to fit the SR models.

```
% chemostat with denitrification
global sAc sNO3 sB Q V k KmAc KmNO3 in Ft Gr0dnf_ana Gr0dnf_cat Rgas
T dGdis_Ac gamma Gdnf_cat


% start and end time
tstart   =  0; % start time in d
tend     =  1000; % end time in d


% write balanced DNF equation
% cat: CH3COO- + 1.6 NO3- + 0.6 H+ --> 0.8 N2 + 2 HCO3- + 0.8 H2O
% ana: 0.725 CH3COO- + 0.2 NO3- + 0.6 H+ --> B + 0.45 CO2 + 0.65 H2O
% ana: 0.525 CH3COO- + 0.2 NH4+ + 0.325 H+ --> B + 0.05 CO2 + 0.45
H2O


%sAc = -1.725; % acetate used per R (NO3)
sAc = -1.525; % acetate used per R (NH4)
sNO3 = -1.6; % nitrate used per R
sB = 1; % biomass per R
% assumes both cat and ana occur simultaneously


% calculating dGrxn0
Gf0Ac = -369.3; % kJ/mol
Gf0NO3 = -111.3; % kJ/mol
Gf0N2 = 18.72; % kJ/mol
Gf0H = 0; % kJ/mol
Gf0HCO3 = -586.9; % kJ/mol
Gf0B = -67; % kJ/mol
Gf0H2O = -237.18; % kJ/mol
Gf0CO2 = -386.02; % kJ/mol
Gf0NH4 = -79.37; % kJ/mol
```

```
Gr0dnf_cat = (0.8*Gf0N2 + 2*Gf0HCO3 + 0.8*Gf0H2O - Gf0Ac -
1.6*Gf0NO3 - 0.6*Gf0H)*1000;
Gr0dnf_ana = (Gf0B + 0.05*Gf0CO2 + 0.45*Gf0H2O - 0.525*Gf0Ac -
0.2*Gf0NH4 - 0.325*Gf0H)*1000;


% calculating dGrxn
Rgas = 8.31446; % J mol^-1 K^-1
T = 273+8; % K


% calculating dGdis
% dGdis = 200+18·[(6 - NoC)]^1.8 + exp[{(-0.2 - ox)^2}^0.16
*(3.6+0.4·NoC)]
% from Gonzalez-Cabaleiro et al. 2015 supp.
NoC_Ac = 2; % # of C in e- donor
degRed_Ac = 4; % oxidation state of C source


dGdis_Ac = 200+18*((6 - NoC_Ac))^(1.8) + exp(((3.8 -
degRed_Ac)^2)^(0.16) *(3.6+ 0.4*NoC_Ac));




% state variables
% 1: acetate
% 2: nitrate
% 3: dnf_B




% concentration in inflow
in_Ac = 10;  % 10 mol/m3
in_NO3 = 20; % 20 mol/m3
in_B = 0; % cell/m3
in =[in_Ac, in_NO3, in_B];


% initial conditions
%yini=[0,0,1]; % use for R w/o B
yini=[0,0,1e6]; % use for R w/ B


% other paramters
Q_in = 1; % m3/s
V = 0.01; % m3
k = 1e-5; % mol/cell/d


% consider estimating Ks based on transport (i.e., Gonzalez-
Cabaleiro)
```

```
KmAc = 10e-3; % mol/m3
KmNO3 = 50e-3; % mol/m3


options = []
options =

     []
options = odeset(options,'NonNegative',[1 2 3])
```

regular forward simulation

```
% using R1
p=0;
Q = 1e-8; % m3/s
savevar_R1 = zeros([100 9]);
for i=1:100
    Q = Q*1.1;

    [t,y] = ode15s(@rhs_R1,[tstart tend],yini,options);


    % reconstruct the rate
    kprime = k.*y(:,3)*Ft;
    R1 = k.*y(:,1).*y(:,2).*y(:,3);


    p=p+1;
    savevar_R1(p,:) = [y(end,:), kprime(end), Q, R1(end), gamma, Ft,
Gdnf_cat];
end


% using R2
p=0;
Q = 1e-8; % m3/s
savevar_R2 = zeros([100 11]);
for i=1:100
    Q = Q*1.1;


    [t,y] = ode15s(@rhs_R2,[tstart tend],yini,options);


    % reconstruct the rate
    kprime = k.*y(:,3)*Ft;
    R2 = k .* y(:,2)./(KmNO3+y(:,2)) .* y(:,1)./(KmAc+y(:,1)) .*
y(:,3);


    % monod expression
    monod_R2_NO3 = y(:,2)./(KmNO3+y(:,2));
```

```
    monod_R2_Ac = y(:,1)./(KmAc+y(:,1));


    p=p+1;
    savevar_R2(p,:) = [y(end,:), kprime(end), Q, R2(end), gamma, Ft,
Gdnf_cat, monod_R2_Ac(end), monod_R2_NO3(end)];
end


% using R3
p=0;
Q = 1e-8; % m3/s
savevar_R3 = zeros([100 9]);
for i=1:100
    Q = Q*1.1;


    [t,y] = ode15s(@rhs_R3,[tstart tend],yini,options);


    % reconstruct the rate
    kprime = k.*y(:,3)*Ft;
    R3 = k.*y(:,1).*y(:,2).*y(:,3)*Ft;

    p=p+1;
    savevar_R3(p,:) = [y(end,:), kprime(end), Q, R3(end), gamma, Ft,
Gdnf_cat];
end


% using R4
p=0;
Q = 1e-8; % m3/s
savevar_R4 = zeros([100 11]);
for i=1:100
    Q = Q*1.1;

    [t,y] = ode15s(@rhs_R4,[tstart tend],yini,options);


    % reconstruct the rate
    kprime = k.*y(:,3)*Ft;
    R4 = k.* y(:,2)./(KmNO3+y(:,2)) .* y(:,1)./(KmAc+y(:,1)) .*
y(:,3) *Ft;

    % monod expression
    monod_R4_NO3 = y(:,2)./(KmNO3+y(:,2));
    monod_R4_Ac = y(:,1)./(KmAc+y(:,1));


    p=p+1;
    savevar_R4(p,:) = [y(end,:), kprime(end), Q, R4(end), gamma, Ft,
Gdnf_cat, monod_R4_Ac(end), monod_R2_NO3(end)];
```

```
end


% using R w/ B
p=0;
Q = 1e-8; % m3/s
savevar_Rd = zeros([100 9]);
for i=1:100
     Q = Q*1.1;


    [t,y] = ode15s(@rhs_dG,[tstart tend],yini,options);


    % reconstruct the rate
    Rd = k.*y(:,3) .* y(:,2)./(KmNO3+y(:,2)) .*
y(:,1)./(KmAc+y(:,1)) *Ft;
    kprime = k.*y(:,3)*Ft;

    p=p+1;
    savevar_Rd(p,:) = [y(end,:), kprime(end), Q, Rd(end), gamma, Ft,
Gdnf_cat];
end


figure
subplot(3,2,1), plot(savevar_R1(:,5),savevar_R1(:,6),'o'), title('R1
vs. Q'), xlabel('Q (m^3/s)'), ylabel('R1 (mol/m^3/s)')
subplot(3,2,2), plot(savevar_R2(:,5),savevar_R2(:,6),'o'), title('R2
vs. Q'), xlabel('Q (m^3/s)'), ylabel('R2 (mol/m^3/s)')
subplot(3,2,3), plot(savevar_R3(:,5),savevar_R3(:,6),'o'), title('R3
vs. Q'), xlabel('Q (m^3/s)'), ylabel('R3 (mol/m^3/s)')
subplot(3,2,4), plot(savevar_R4(:,5),savevar_R4(:,6),'o'), title('R4
vs. Q'), xlabel('Q (m^3/s)'), ylabel('R4 (mol/m^3/s)')
subplot(3,2,5), plot(savevar_Rd(:,5),savevar_Rd(:,6),'o'), title('Rd
vs. Q'), xlabel('Q (m^3/s)'), ylabel('Rd (mol/m^3/s)')
```

A.2.3 rhs_R1.m

Reaction rate R_1 model script.

```
function dy=rhs_R1(t,y)
global sAc sNO3 Q V k in Ft Gr0dnf_ana Gr0dnf_cat Rgas T dGdis_Ac gamma
Gdnf_cat Gdnf_ana

ac = max(0,y(1));
no3 = max(0,y(2));
dnf_B = max(0,y(3));

R1 = k*ac*no3*dnf_B;
```

```
% calculating Grxn for gamma
hco3 = 2.139938155; % mol/m3 (calculated in excel at pH=8.2)
co2 = 0.013594676; % mol/m3
n2 = 10; % mol/m3
h = 6.30957E-06; % mol/m3
nh4 = 1; % mol/m3

valmin = 1e-20;
Qdnf_cat =
(n2^(0.2)*hco3^(2))/(max(valmin,ac)*max(valmin,no3^(1.6))*h^(0.6));
Qdnf_ana = co2^(0.05)/(max(valmin,ac^(0.525))*nh4^(0.2)*h^(0.325));

Gdnf_cat = (Gr0dnf_cat + Rgas*T*log(Qdnf_cat))/1000;
Gdnf_ana = (Gr0dnf_ana + Rgas*T*log(Qdnf_ana))/1000;

gamma = Gdnf_cat/-(Gdnf_ana+dGdis_Ac); % mol_B/mol_eD

% calculating Ft
dG_atp = 50; % kJ/mol
chi_Ft = 8; % 2 e- transferred per reaction; J&B 2005
m_Ft = 2/3; % 2 H+ translocated per pair of electrons and 3 req to make
one ATP

fx = Gdnf_cat + m_Ft*dG_atp; % -dG_dnit - dG_atp
Ft=max(0,1-exp(fx/(chi_Ft*(Rgas/1000)*T)));


dy(1) = sAc*R1 +(Q/V)*in(1) -(Q/V)*ac;
dy(2) = sNO3*R1 +(Q/V)*in(2) -(Q/V)*no3;
dy(3) = gamma*R1 +(Q/V)*in(3) -(Q/V)*dnf_B;

dy=dy';
end
```

A.2.4 rhs_R2.m

Reaction rate R_2 model script.

```
function dy=rhs_R2(t,y)
global sAc sNO3 Q V k KmAc KmNO3 in Ft Gr0dnf_ana Gr0dnf_cat Rgas T
dGdis_Ac gamma Gdnf_cat

ac = max(0,y(1));
no3 = max(0,y(2));
dnf_B = max(0,y(3));

R2 = k * (no3/(KmNO3+no3)) * (ac/(KmAc+ac))*dnf_B;

% calculating Grxn for gamma
hco3 = 2.139938155; % mol/m3 (calculated in excel at pH=8.2)
co2 = 0.013594676; % mol/m3
```

```
n2 = 10; % mol/m3
h = 6.30957E-06; % mol/m3
nh4 = 1; % mol/m3

valmin = 1e-20;
Qdnf_cat =
(n2^(0.2)*hco3^(2))/(max(valmin,ac)*max(valmin,no3^(1.6))*h^(0.6));
Qdnf_ana = co2^(0.05)/(max(valmin,ac^(0.525))*nh4^(0.2)*h^(0.325));

Gdnf_cat = (Gr0dnf_cat + Rgas*T*log(Qdnf_cat))/1000;
Gdnf_ana = (Gr0dnf_ana + Rgas*T*log(Qdnf_ana))/1000;

gamma = Gdnf_cat/-(Gdnf_ana+dGdis_Ac); % mol_B/mol_eD

% calculating Ft
dG_atp = 50; % kJ/mol
chi_Ft = 8; % 2 e- transferred per reaction; J&B 2005
m_Ft = 2/3; % 2 H+ translocated per pair of electrons and 3 req to make
one ATP

fx = Gdnf_cat + m_Ft*dG_atp; % -dG_dnit - dG_atp
Ft=max(0,1-exp(fx/(chi_Ft*(Rgas/1000)*T)));


dy(1) = sAc*R2 +(Q/V)*in(1) -(Q/V)*ac;
dy(2) = sNO3*R2 +(Q/V)*in(2) -(Q/V)*no3;
dy(3) = gamma*R2 +(Q/V)*in(3) -(Q/V)*dnf_B;

dy=dy';
end
```

A.2.5 rhs_R3.m

Reaction rate R_3 model script.

```
function dy=rhs_R3(t,y)
global sAc sNO3 Q V k in Ft Gr0dnf_ana Gr0dnf_cat Rgas T dGdis_Ac gamma
Gdnf_cat

ac = max(0,y(1));
no3 = max(0,y(2));
dnf_B = max(0,y(3));


% calculating Grxn for gamma
hco3 = 2.139938155; % mol/m3 (calculated in excel at pH=8.2)
co2 = 0.013594676; % mol/m3
n2 = 10; % mol/m3
h = 6.30957E-06; % mol/m3
nh4 = 1; % mol/m3

valmin = 1e-20;
```

```
Qdnf_cat =
(n2^(0.2)*hco3^(2))/(max(valmin,ac)*max(valmin,no3^(1.6))*h^(0.6));
Qdnf_ana = co2^(0.05)/(max(valmin,ac^(0.525))*nh4^(0.2)*h^(0.325));

Gdnf_cat = (Gr0dnf_cat + Rgas*T*log(Qdnf_cat))/1000;
Gdnf_ana = (Gr0dnf_ana + Rgas*T*log(Qdnf_ana))/1000;

gamma = Gdnf_cat/-(Gdnf_ana+dGdis_Ac); % mol_B/mol_eD

% calculating Ft
dG_atp = 50; % kJ/mol
chi_Ft = 8; % 2 e- transferred per reaction; J&B 2005
m_Ft = 2/3; % 2 H+ translocated per pair of electrons and 3 req to make
one ATP

fx = Gdnf_cat + m_Ft*dG_atp; % -dG_dnit - dG_atp
Ft=max(0,1-exp(fx/(chi_Ft*(Rgas/1000)*T)));


R3 = k*ac*no3*Ft*dnf_B;


dy(1) = sAc*R3 +(Q/V)*in(1) -(Q/V)*ac;
dy(2) = sNO3*R3 +(Q/V)*in(2) -(Q/V)*no3;
dy(3) = gamma*R3 +(Q/V)*in(3) -(Q/V)*dnf_B;

dy=dy';
end
```

A.2.6 rhs_R4.m

Reaction rate R_4 model script.

```
function dy=rhs_R4(t,y)
global sAc sNO3 Q V k KmAc KmNO3 in Ft Gr0dnf_ana Gr0dnf_cat Rgas T
dGdis_Ac gamma Gdnf_cat

ac = max(0,y(1));
no3 = max(0,y(2));
dnf_B = max(0,y(3));


% calculating Grxn for gamma
hco3 = 2.139938155; % mol/m3 (calculated in excel at pH=8.2)
co2 = 0.013594676; % mol/m3
n2 = 10; % mol/m3
h = 6.30957E-06; % mol/m3
nh4 = 1; % mol/m3

valmin = 1e-20;
Qdnf_cat =
(n2^(0.2)*hco3^(2))/(max(valmin,ac)*max(valmin,no3^(1.6))*h^(0.6));
```

```
Qdnf_ana = co2^(0.05)/(max(1e-100,ac^(0.525))*nh4^(0.2)*h^(0.325));

Gdnf_cat = (Gr0dnf_cat + Rgas*T*log(Qdnf_cat))/1000;
Gdnf_ana = (Gr0dnf_ana + Rgas*T*log(Qdnf_ana))/1000;

gamma = Gdnf_cat/-(Gdnf_ana+dGdis_Ac); % mol_B/mol_eD

% calculating Ft
dG_atp = 50; % kJ/mol
chi_Ft = 8; % 2 e- transferred per reaction; J&B 2005
m_Ft = 2/3; % 2 H+ translocated per pair of electrons and 3 req to make
one ATP

fx = Gdnf_cat + m_Ft*dG_atp; % -dG_dnit - dG_atp
Ft=max(0,1-exp(fx/(chi_Ft*(Rgas/1000)*T)));

R4 = k * (no3/(KmNO3+no3)) * (ac/(KmAc+ac)) *Ft*dnf_B;


dy(1) = sAc*R4 +(Q/V)*in(1) -(Q/V)*ac;
dy(2) = sNO3*R4 +(Q/V)*in(2) -(Q/V)*no3;
dy(3) = gamma*R4 +(Q/V)*in(3) -(Q/V)*dnf_B;

dy=dy';
end
```

A.3 SR model parameterization

A.3.1 STEADYSED

```
default_pysr_params = dict(
    populations=30,
    procs='auto',
    model_selection='best'
)

model = pysr.PySRRegressor(niterations=100,
                           population_size=50,
                           binary_operators=['*'],
                           **default_pysr_params
)
```

Fig S2.1. Parametrization for the symbolic regressor used in the model runs involving STEADYSED data. Model parameter complexity was kept to a minimum to see how well a simple setup performed at constructing the expressions.

A.3.2 Chemostat

```
default_pysr_params = dict(
    populations=24, # more pops increase the diversity of equations found; (Rule of Thumb =
3*num_cores)
    procs=8, # cores dedicated to Julia operations
```

```
    model_selection='best', # uses a combination of 'accuracy' and 'complexity' to select best eq
    ncyclesperiteration = 1000, # number of mutations to run; per 10 samples of pop, per iteration
    nested_constraints = {'exp': {'exp': 0}, 'log': {'log':0}},
    # nested_constraints = {'monod_Ac': {'monod_Ac': 0, 'monod_NO3': 0},
    #                       'monod_NO3': {'monod_Ac': 0, 'monod_NO3': 0}}, # ensures algo won't use
nested Monod functions
    maxsize = 15, # length of equation including operators; goal here is a length of 9 max (e.g.,
k*B*monod_Ac*monod_NO3*Ft)
    warm_start=False, # only True if you want the previous result to be where the regressor begins
    adaptive_parsimony_scaling=100.0, # increasing helps control the weights from only optimizing
complex expressions
    turbo=True, # can speed up search evaluation
    weight_optimize=0.001, # optimizes weights more frequently (helps at high
ncyclesperiteration)
    verbosity = 0 # set to 1 for detailed model output statements
)

model = pysr.PySRRegressor(niterations=1500, # total number of generations
                           population_size=80,
                           binary_operators=binary_operators,
                        #    unary_operators = ['monod_Ac(x) = x / (x + 10f-3)',
                        #                       'monod_NO3(x) = x / (x + 50f-3)','exp'],
                           unary_operators=['exp','log'],
                        #    extra_sympy_mappings={"monod_Ac": lambda x: x / x + 1e-2,
                        #                          "monod_NO3": lambda x: x / x + 5e-2},
                           **default_pysr_params
)
```

Fig S2.2. Parameterization for the symbolic regressor used in the model runs involving data from our chemostat simulations. Due to the expressions in these microbial rate expressions being more complicated than the rate laws used in STEADYSED, more tuning of the regressor was required. Comments beside each parameter describe what it does and why it is used here.

A.4 Rate constant unit conversion

The discrepancy in the numerical value of the rate constant determined for the oxidation of Fe^{2+} using O_2 with PySR vs. k_{FeOx} from STEADYSED ($0.14*10^9$ M^{-1} yr^{-1}) is due to a difference in units. The input to PySR was the rate in µmol cm_{total}^{-3} yr^{-1} and concentrations in µM. For a rate expression $R = k^{PySR}[Fe^{2+}][O_2]$, $k^{PySR} = 0.112$ µmol cm_{total}^{-3} yr^{-1} μM^{-2} (the units of $R/[Fe^{2+}]/[O_2]$). Division by porosity (0.8 $cm_{fluid}^3/cm_{total}^3$) and converting some units gives $0.112 *10^{-6}$ mol cm_{total}^{-3} yr^{-1} * $10^3 cm_{fluid}^3/10^6$ mol * $L/10^{-6}$ mol * cm_{total}^3 /0.8 $cm_{fluid}^3 = 0.14*10^9$ M^{-1} yr^{-1}.

A.5 Monod warm-start resulting in correct R_4 expression

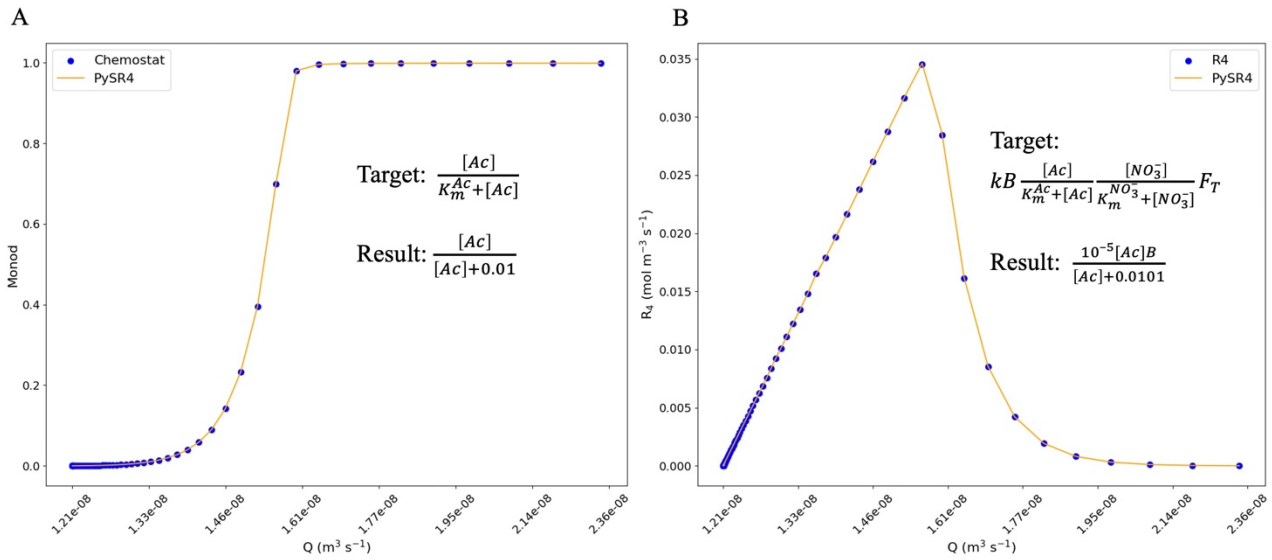

Fig S2.3. PySR simulation results showing A) Monod expression of acetate for rate law R_4 and B) using a warm-start from the Monod result to target the R_4 expression. This shows PySR can return a rate law consisting of Monod kinetic expressions using only basic operators and co-dependent concentration fields. The Monod expression for NO_3^- and F_T are both 1 which is why they were not used in the resulting expression.

A.6 GPLearn vs. PySR

Our original explorations for using symbolic regression to describe reaction rate expressions were conducted using the Python package gplearn. Here is a brief description of the parameterization for that particular model and how its functionality compares to that of PySR for the results discussed above.

Parametrization of the symbolic regressor in gplearn begins with setting the population size and tournament size for each generation. This defines the number of trees that are evaluated for their fitness and compete against each other to find that which has the best fit. The winning tree will then be used to fill the next generation's population.

The regressor will iterate over the set number of generations or until it reaches the stopping criteria imposed, which is defined as the mean absolute error by default. The function set to be used for regression must also be provided here as a list of the operators or custom functions. To control the size of the programs generated by the model, the parsimony coefficient must be tuned. Larger values will penalize the model and create smaller trees, which can help with bloat and keeping the resulting program interpretable. In addition, the genetic mutation parameters will also need to be defined to introduce complexity into the programs. The probability of the mutations occurring is controlled by floating point values, with a maximum sum of < 1.0. For the application presented here, the combination used is p_subtree_mutation = 0.2 and p_crossover = 0.8. Subtree mutation takes the tournament winner for a generation and selects a random subtree to be replaced by a randomly generated replacement. This new mutant program is used as the parent for the next generation. The crossover mutation also selects a subtree from the winner of a tournament and runs a new tournament to generate a donor that is used to swap subtrees with. Using these mutations introduces complexity and new possible programs into the generations, much like our biological mechanisms of genetic inheritance. As is evident, parameterization requires some experimentation to find the best approach depending on the problem.

The PySR package has similar functionality to gplearn in so far as being an iterative algorithm that performs genetic programming to optimize a loss function and find a solution. Where its main difference lies, however, is that the actual computation is offloaded into the programming language Julia, resulting in a significantly faster model than gplearn which does its calculations in Python. To build the model, one must set the

populations and number of iterations (e.g., generations) that are to be run. Operators are also selected as parameter inputs, but are split into two categories within the PySR regressor: binary and unary. The model also allows for custom functions to be used, requiring them to be written in Julia syntax. Here, we define operators for the Monod-like terms that dictate substrate dependence on the rate expression. PySR allows for control over the mutation parameters similar to how gplearn does, but this seems to require less tuning to get good results. Some other differences to be aware of in PySR is that constraints can be set on operators to only allow a defined amount of complexity to be used. One application of this functionality is setting constraints on the nesting of Monod operators, as the algorithm has a tendency to apply operators to the same feature multiple times. The PySR package also allows for multiple targets to be selected as inputs for regression, unlike gplearn, but it will simply run them in succession rather than in tandem. The number of features that can be selected is limited to 10. Trying to add in more will slow down computation and may require dimensionality reduction to address which features ultimately matter to describe the function being derived.

APPENDIX B

SUPPLEMENTARY MATERIAL

CHAPTER 3: GROWTH EFFICIENCY AND NITROGEN USE IN ANAEROBIC METHANE OXIDATION: A MICROBIAL CONSORTIUM MODEL AND FLUX BALANCE ANALYSIS

Model files and code can be found at: https://bitbucket.org/MeileLab/kyle/src/master/

B.1 Calculations for NH_4^+ demand in COMSOL model

The concentration of ammonium in the environment required to sustain microbial growth was estimated by first estimating an area of uptake flux from the ANME cell-specific uptake rate of NH_4^+. The minimum ammonium concentration in the environment was then derived using Fick's Law of Diffusion where we assume complete drawdown occurs from the environmental pool to the cell's cytoplasm.

$$F_{NH4} = \frac{R_{NH4,cell}}{A_{cell}} \rightarrow F_{NH4} = -D\frac{C_{NH4} - 0}{d_{BL}} \rightarrow C_{NH4} = \frac{F_{NH4} \cdot d_{BL}}{D}$$
$$\rightarrow C_{NH4} * 10^{-15}\ mol\ m^{-3} = 1.59\ x\ 10^{-9}\ mol\ m^{-3}\ or\ 1.59\ pmol\ L^{-1}$$

where $R_{NH4,cell}$ is the ANME cell-specific uptake rate of NH_4^+ computed in the model with units of *fmol cell*$^{-1}$ *d*$^{-1}$ and A_{cell} is the surface area of the cell in $m^2\ cell^{-1}$. F_{NH4} is the diffusive flux of NH_4^+ in *fmol* m^{-2} d^{-1}, D is the diffusive constant in $m^2\ d^{-1}$, C_{NH4} is the concentration of ammonium required for growth, and d_{BL} is the width of the cell boundary layer.

B.2 Maximum flux constraints on growth medium & solver tolerance in FBA model

Diffusive fluxes were derived from the medium contents used in Nauhaus et al. (2007) as outlined in Widdel & Bak (1992). We use the cell-specific methane uptake rate from Nauhaus et al. (2007) as a methane flux constraint in the FBA simulations by converting it into units of a diffusive flux:

$$F_{CH4} = \frac{R_{Nauhaus} \nu_{C-BM} m_{C-cell}}{A_{cell}} = 6.18x10^{-9}\ mol_{CH4}\ cm^{-2}\ d^{-1}$$

where $R_{Nauhaus}$ is the uptake rate of methane ($5.00x10^{-3}\ mol\ g_{cdw}^{-1}\ d^{-1}$), ν_{C-BM} is the mass ratio of C in biomass to the total molecular weight of the biomass formula used in the FBA model ($0.4212\ g_{C-BM}\ g_{BM}^{-1}$), $m_{c\text{-}cell}$ is the estimated mass of C per cell ($5.9x10^{-14}\ g_C\ cell^{-1}$), and A_{cell} is the surface area of an ANME cell based on a typical cell radius ($2.01x10^{-8}\ cm^2\ cell^{-1}$). The other chemicals in the media were converted to diffusive fluxes as shown in Table S3.1, using Fick's Law and a diffusive boundary layer length d_{BL} of 0.1 mm.

$$F_x = -D_{aq}\frac{C_x - 0}{d_{BL}}$$

Fluxes used for simulations are highlighted in green. AQDS concentration used comes from experiments conducted in Scheller et al. (2016). Given the magnitude of these imposed fluxes, the tolerance of the model solver was set to 10^{-16} to account for the lower flux bounds.

Table S3.1 Diffusive Fluxes used as constraints in FBA model medium

Molecule	C (mol L^{-1})	D_{aq} ($m^2\ s^{-1}$)	Flux (mol $cm^{-2}\ d^{-1}$)
Na^+	0.398556511	1.33E-09	-4.58E-02
Cl^-	0.386193417	2.03E-09	-6.77E-02
Mg^{2+}	0.0147563	7.05E-10	-8.99E-04
Ca^{2+}	0.001020339	7.93E-10	-6.99E-05
SO_4^{2-}	0.028241081	5.12E-10	-1.25E-03
NH_4^+	0.004673683	1.00E-09	-4.04E-04
K^+	0.008176567	1.96E-09	-1.38E-03
PO_4^{2-}	0.001469659	6.90E-10	-8.76E-05
Fe^{2+}	0.000075	7.19E-10	-4.66E-06
Co^{2+}	0.000008	7.32E-10	-5.06E-07
Cu^{2+}	0.0000001	7.33E-10	-6.33E-09
Zn^{2+}	0.000005	7.15E-10	-3.09E-07
MoO_4	0.0000015	1.98E-09	-2.57E-07
Ni^{2+}	0.000001	1.50E-09	-1.30E-07
AQDS	5.00E-05	1.00E-10	-4.58E-02

B.3 Protein Leakage Simulations

Our protein leakage simulations were conducted by increasing the stoichiometric coefficient for the protein required in biomass (shown in Figure S3.1) and is hence tied to the growth rate. Any amount of protein produced above that required amount was then assumed to be leaked from the cell. We calculated the leakage flux in terms of C by using:

$$F_{Leakage} = F_{BM} \nu_{C-Protein} \nu_{leakage}$$

where F_{BM} is the flux of biomass, $\nu_{C-Protein}$ is the C content per protein molecule (37.936), and $\nu_{leakage}$ is the difference between the increased and default protein stoichiometric coefficient.

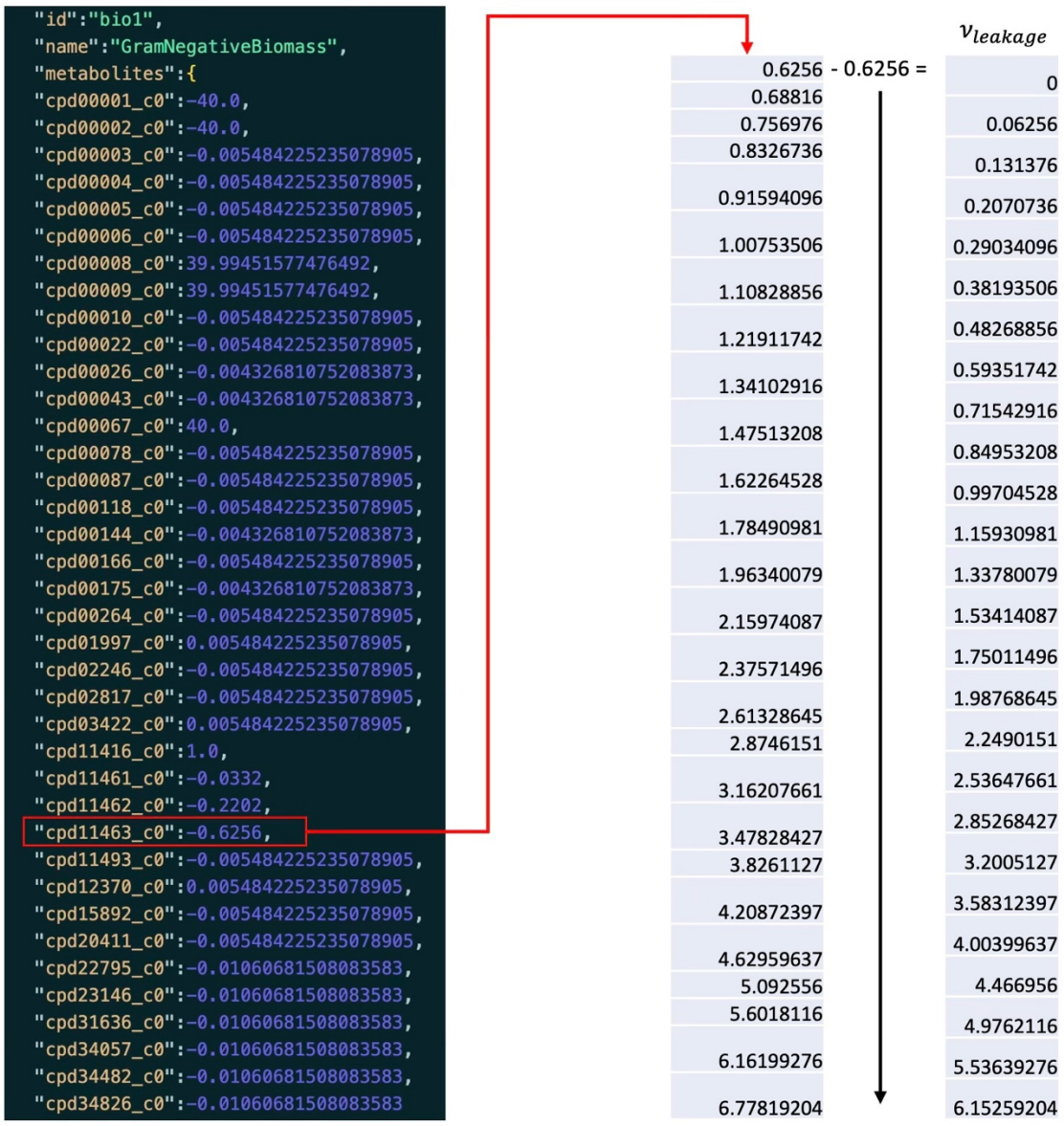

Fig S3.1. Diagram showing how protein leakage was calculated. Code shows the reaction for biomass production (bio1), with the protein (cpd11463_c0) outlined in red. This value was increased incrementally and the difference between that and the default value is the protein that was assumed to be leaked.

References

Nauhaus, K., Albrecht, M., Elvert, M., Boetius, A., & Widdel, F. (2007). In vitro cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate. *Environmental Microbiology*, *9*(1), 187-196.

Scheller, S., Yu, H., Chadwick, G. L., McGlynn, S. E., & Orphan, V. J. (2016). Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science*, *351*(6274), 703-707.

Widdel, F., & Bak, F. (1992). Gram-negative mesophilic sulfate-reducing bacteria. In *The prokaryotes: a handbook on the biology of bacteria: ecophysiology, isolation, identification, applications* (pp. 3352-3378). New York, NY: Springer New York.