From metagenomes to metabolism

Systematically assessing the metabolic flux feasibilities for “Candidatus Accumulibacter” species during anaerobic substrate uptake

Páez-Watson, Timothy; van Loosdrecht, Mark C.M.; Wahl, S. Aljoscha

DOI
10.1016/j.watres.2023.121028

Publication date
2024

Document Version
Final published version

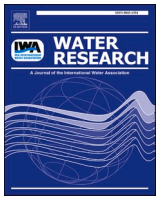
Published in
Water Research

Water Research 250 (2024) 121028

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



From metagenomes to metabolism: Systematically assessing the metabolic flux feasibilities for "*Candidatus* Accumulibacter" species during anaerobic substrate uptake

Timothy Páez-Watson [*], Mark C.M. van Loosdrecht, S.Aljoscha Wahl [1]

Department of Biotechnology, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Metagenome-assembled genomes
Metabolic predictions
Elementary flux modes
Pathway thermodynamics

ABSTRACT

With the rapid growing availability of metagenome assembled genomes (MAGs) and associated metabolic models, the identification of metabolic potential in individual community members has become possible. However, the field still lacks an unbiassed systematic evaluation of the generated metagenomic information to uncover not only metabolic potential, but also feasibilities of these models under specific environmental conditions. In this study, we present a systematic analysis of the metabolic potential in species of "*Candidatus* Accumulibacter", a group of polyphosphate-accumulating organisms (PAOs). We constructed a metabolic model of the central carbon metabolism and compared the metabolic potential among available MAGs for "*Ca.* Accumulibacter" species. By combining Elementary Flux Modes Analysis (EFMA) with max-min driving force (MDF) optimization, we obtained all possible flux distributions of the metabolic network and calculated their individual thermodynamic feasibility. Our findings reveal significant variations in the metabolic potential among "*Ca.* Accumulibacter" MAGs, particularly in the presence of anaplerotic reactions. EFMA revealed 700 unique flux distributions in the complete metabolic model that enable the anaerobic uptake of acetate and its conversion into polyhydroxyalkanoates (PHAs), a well-known phenotype of "*Ca.* Accumulibacter". However, thermodynamic constraints narrowed down this solution space to 146 models that were stoichiometrically and thermodynamically feasible (MDF $>$ 0 kJ/mol), of which only 8 were strongly feasible (MDF $>$ 7 kJ/mol). Notably, several novel flux distributions for the metabolic model were identified, suggesting putative, yet unreported, functions within the PAO communities. Overall, this work provides valuable insights into the metabolic variability among "*Ca.* Accumulibacter" species and redefines the anaerobic metabolic potential in the context of phosphate removal. More generally, the integrated workflow presented in this paper can be applied to any metabolic model obtained from a MAG generated from microbial communities to objectively narrow the expected phenotypes from community members.

1. Introduction

Microbial ecology research strongly relies on cultivation independent approaches since most bacterial species are, to date, unculturable (Steen et al., 2019). Instead, data is generated from analysing microbial communities directly in their natural environments, often through metagenome analysis. The rapid development of high throughput sequencing technologies has resulted in a growing number of metagenome-assembled genomes (MAGs) representing members from various microbial communities (for example Singleton et al. (2021)). Over the years, MAGs generated from ecological samples have been linked to potential functional guilds in microbial communities based on the presence of specific genes (Wrighton et al., 2012; Anantharaman et al., 2016). Particularly in the context of water engineering, MAG's metabolic potentials encompass functions like exopolysaccharide synthesis (Dueholm et al., 2023), nitrogen, phosphorus and iron removal (Kristensen et al., 2021) and even mutualistic interactions amongst species (Fujii et al., 2022).

The metabolic potential derived from MAGs is the initial step towards attaining a mechanistic understanding of the physiology of community members– *i.e.*, what they actually do within the community. Constraint based methods like flux balance analysis (FBA), provide tools

* Corresponding author.
E-mail address: T.W.PaezWatson@tudelft.nl (T. Páez-Watson).
[1] Lehrstuhl für Bioverfahrenstechnik, Friedrich-Alexander University Erlangen-Nürnberg, Paul-Gordan-Straße 3, Erlangen 91,052, Germany.

https://doi.org/10.1016/j.watres.2023.121028
Received 1 September 2023; Received in revised form 6 December 2023; Accepted 16 December 2023
Available online 17 December 2023

to predict metabolic functions and have been successfully applied to study monocultures (Famili et al., 2003; O'Brien et al., 2015). Efforts have been made to extend their applications towards understanding metabolic interactions (Du et al., 2022; Kuppa et al., 2023), resource allocation (Sharma and Steuer 2019), microbial biosynthesis (Páez-Watson et al., 2023), or even inferring functional guilds (Muller et al., 2018) from microbial communities in ecology (Dillard et al. (2021) explain and discuss the available methods more in depth).

Transitioning from a metabolic network to metabolic flux predictions requires addressing key assumptions regarding (i) objective functions of the cells, (ii) considered constraints (or limits) on intracellular reactions and (iii) the chosen environmental conditions for the simulations (Kim et al., 2022). As the number of high-quality MAGs continues to grow, so does the number of metabolic models aiming at predicting community functions. However, critically evaluating the feasibility of metabolic pathways in the specific environmental context is crucial for accurately defining constraints on intracellular reactions, thereby addressing assumptions ii and iii. In this work we propose an integrated workflow that integrates existing methodologies in constraint-based modelling and pathway thermodynamics to address these assumptions. As a case study, we apply this workflow to critically assess the anaerobic metabolic capabilities of the well-studied community of Phosphate Accumulating Organisms (PAOs).

PAOs are considered the main microbial group contributing to the enhanced biological phosphorus removal (EBPR) process and have been extensively studied for many decades (Mino et al., 1987; Van Loosdrecht et al. 1997; Oehmen et al., 2007; Nielsen et al., 2019). Among the PAOs, "*Ca.* Accumulibacter" has emerged as a highly studied genus due to its complex metabolism and role in the EBPR cycle. "*Ca.* Accumulibacter" thrives by utilizing a dynamic interplay of storage polymers, including polyphosphate, glycogen, and polyhydroxyalkanoates (PHAs), in the alternating anaerobic/aerobic cycles of the EBPR ecosystem (Mino et al., 1998). Despite numerous attempts, pure cultures of this organism are still lacking, emphasizing the critical importance of studying their MAGs for gaining insights into their contribution to the EBPR process.

Several groups have proposed different biochemical models that could explain the internal metabolism of "*Ca.* Accumulibacter". Each of these models suggests uniquely different ways as to how "*Ca.* Accumulibacter" obtains the reducing equivalents (NADH) required for storage of fatty acids as PHAs during anaerobic periods. Comeau et al. (1986) proposed that the complete tricarboxylic acid (TCA) cycle is active anaerobically as a source of NADH for the accumulation of polyhydroxybutyrate (PHB) (Fig. 1A). Mino et al. (1987) demonstrated anaerobic depletion of glycogen and suggested it as the source of NADH by means of glycolysis (Fig. 1B). Pereira et al. (1996) measured polyhydroxyvalerate (PHV) as well as PHB in EBPR sludge and proposed a mechanism for its accumulation using both glycolysis and the TCA cycle, albeit with a secondary back-flux through the 'left' branch of the TCA cycle to balance NADH production (Fig. 1C). Conversely, Hesselmann et al. (2000) suggested a model including glycolysis and a split TCA cycle (Fig. 1D), introducing the use of one of the anaplerotic reactions (catalysing the conversion from pyruvate to oxaloacetate) while Yagci et al. (2003) introduced the concept of the glyoxylate shunt to bypass certain reactions within the TCA cycle (Fig. 1E). Burow et al. (2008) provided experimental evidence of the functioning of the glyoxylate shunt during the anaerobic phase of EBPR by using enzymatic inhibitors supporting the model from Yagci et al. (2003). Conversely, Zhou et al. (2009) provide evidence for acetate uptake without glycogen use supporting a full anaerobic TCA cycle. Modelling approaches such as flux balance analysis (FBA) on PAOs have proposed so-called varying stoichiometry (or metabolic flexibility) which could explain the lack of consensus over all these proposed stoichiometries (da Silva et al. 2020).

This apparent complexity (and hence lack of consensus) for understanding the transformations in central carbon metabolism during the anaerobic uptake of acetate showcases the need for systematic re-evaluation of potential stoichiometries and reaction limitations within a given environmental context. Furthermore, the newly available MAGs associated to individual species of "*Ca.* Accumulibacter" (Petriglieri et al., 2022) represent a great opportunity to study potential differences amongst species from this complex genus. In their research, numerous species of the genus "*Ca.* Accumulibacter" were assigned, and a general comparison was made on the difference in potential metabolic guilds harvested in these newly available MAGs.

Stoichiometric modelling simply calculates the stoichiometries of a network while balancing the production and consumption of metabolites – *i.e.,* maintaining steady state. While FBA requires an optimization objective to identify stoichiometries (Orth et al., 2010), elementary flux mode analysis (EFMA) is a powerful tool that offers a systematic approach to assess all minimal combinations possible in a given metabolic network (Terzer and Stelling 2008; Maarleveld et al., 2013). Flux modes can be especially strong when dealing with highly interwoven metabolic networks (such as central carbon metabolism) and has been famously applied to understand the TCA cycle as much more than just energy generation (Sweetlove et al., 2010). The high number of flux

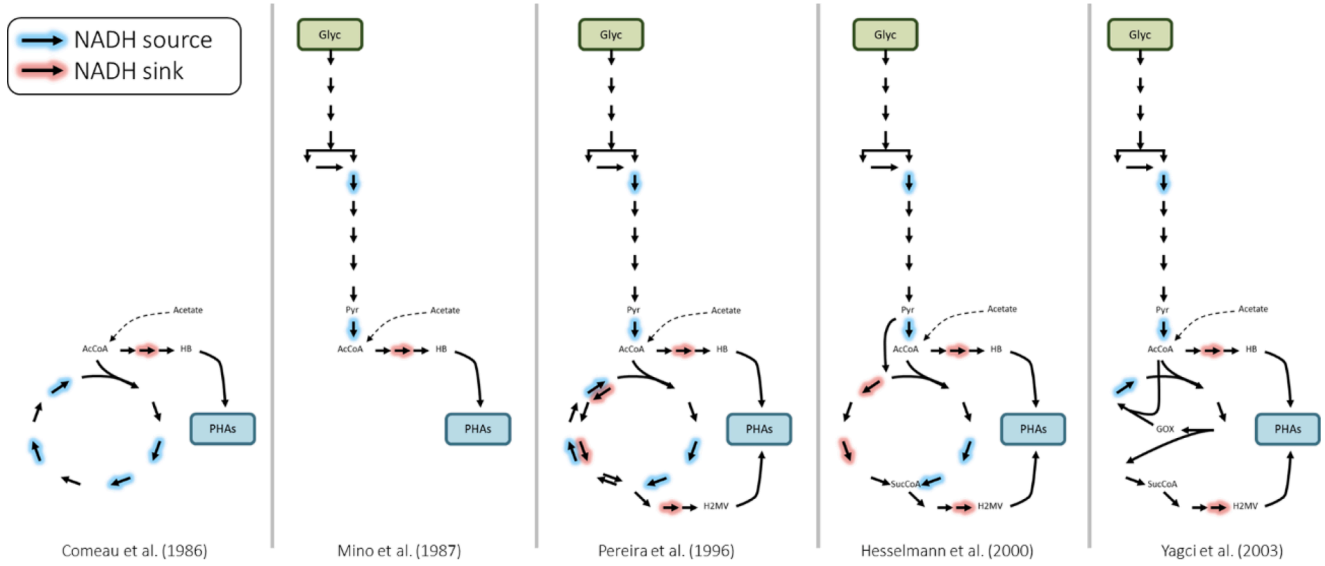


Fig. 1. Schematic representation of the proposed metabolic transformations in PAOs during the anaerobic uptake of acetate over time (Comeau et al., 1986; Mino et al., 1987; Pereira et al., 1996; Hesselmann et al., 2000; Yagci et al., 2003).

modes possible from a metabolic network (due to its combinatorial effect) severely limits its application.

Approaches to filter (or minimize) flux modes from a network are based on supposed irreversibilities from individual reactions (Gerstl et al., 2016; Peres et al., 2017). Nevertheless, the *ad hoc* definition of reaction irreversibilities neglects the context of the given reaction – *i.e.,* the metabolic conditions in which it is happening and its relative contribution to the overall pathway stoichiometry. In contrast, methods such as the max-min driving force (MDF) optimize the overall thermodynamics driving force of a pathway (Noor et al., 2014) without presumptions on the reaction irreversibilities. Thus, its combination with EFMA could hold the key for the identification of the most probably stoichiometries from a pathway given a specific environmental context.

In this work, we analysed the anaerobic metabolism of PAOs to illustrate a systematic methodology for deriving consistent metabolic insights from MAGs. We assess the metabolic potential in central carbon metabolism of nineteen high quality MAGs of "*Ca.* Accumulibacter" species by employing EFMA and highlight differences amongst these species. Further, we determine which potential metabolic flux models contribute to the most feasible model solutions during the anaerobic uptake and storage of acetate into PHAs.

2. Materials and methods

2.1. Model construction

A metabolic model was built to represents reactions from glycolysis, TCA cycle, anaplerotic reactions, PHA synthesis reactions, and glycogen degradation pathways. The stoichiometry of each reaction was obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, using the available MAGs for "*Ca.* Accumulibacter" as the reference. Specific network reactions and stoichiometry is detailed in *Supplementary materials (Table S1)*. The chosen pathways to be incorporated in this metabolic model were focused on substrate uptake and the model was utilized to test current hypothetical *operations* of metabolism to uptake acetate into PHAs (as summarized in Fig. 1). For building more complex metabolic models, we recommend the user to apply available methods for model construction such as gapseq (Zimmermann et al., 2021), CarveMe (Machado et al., 2018), amongst others.

The model, consisting of 45 metabolites connected by 43 reactions includes exchange reactions for metabolites such as PHA, glycogen, polyphosphate, CO_2, and H_2O. Model reactions were formulated and implemented in Excel in the form of a stoichiometric matrix **S**, in which the rows and columns signify the metabolites and the reactions respectively. All metabolites in the **S** matrix are considered to be in steady state by the relation:

$$\boldsymbol{S}.\boldsymbol{v} = 0$$

where v represents a vector containing the fluxes of each reaction (column in **S**) (Orth et al., 2010). We did not explicitly model $FADH_2$, all electron transporters were modelled as NADH. Similarly, all reactions were considered to be reversible at this early stage of the model. PHAs are modelled as poly-hydroxybutyrate (PHB) and hydroxy 2-methyl-valerate (PH2MV) as a proxy to distinguish PHAs resulting from either acetyl CoA or propionyl CoA as done previously (Páez-Watson et al., 2023).

2.2. Enzyme and gene annotation

Genome analysis was conducted to investigate the presence of genes in species of "*Ca.* Accumulibacter" related to the reactions from the metabolic model. The MAG sequences of 19 species of "*Ca.* Accumulibacter" were obtained from the European Nucleotide Archive and downloaded when available in the format WGS Set EMBL. Standard genome annotations from the database was used to verify the presence of the genes related to our metabolic network. In case these genes were not annotated in all genomes, a set of proteins from Escherichia coli K12 for catalysing the reactions in the network was used as a reference (see *Supplementary materials, Table S3*). BLAST analysis was then performed (min_identity 30 %, evalue e-12) to compare these protein sequences with the reference set, and sequence alignment was used to assess conservation and identify potential orthologs or homologs.

2.3. Elementary flux mode analysis

The metabolic network was subjected to elementary flux mode (EFM) analysis using the *efm* MATLAB tool developed by Terzer and Stelling (2008). For the analysis, reactions for glycogen and polyphosphate degradation, PHA synthesis and acetate consumption were set as irreversible since this is the observed phenotype from the anaerobic phase of EBPR. This approach aimed to systematically evaluate the model's capabilities so no further reactions were set as irreversible. The resulting elementary flux modes (EFM) were all normalized to 1 mol of acetate uptake. Variables of interest, including glycogen degradation, PHB and PH2MV synthesis, polyphosphate degradation, and CO_2 export, were examined for their distribution among the EFMs using Python and visualized with violin plots. Relevant EFMs matching the stoichiometry proposed in the literature (specifically the proposed models from Comeau et al. (1986), Mino et al. (1987), Pereira et al. (1996), Hesselmann et al. (2000) and Yagci et al. (2003)) were manually filtered based on the presence and direction of their active reactions within the EFMs.

2.4. MDF calculations and normalization

We evaluated the thermodynamic feasibility of each solution to the metabolic model (*i.e.* an individual EFM) using the concept of Minimum Driving Force (MDF) (Noor et al., 2014). For a reaction to be considered feasible, a negative value for the Gibbs free energy (Δ_rG') is required. The Gibbs free energy of each reaction is determined by the (optimized) concentration of the metabolites in a given range and thus depends on the context of that reaction within the model solution (*i.e.* which and how many reactions are active in the current EFM). For each EFM, the resulting MDF is obtained from the reaction with the lowest thermodynamic driving force (*i.e.* the lowest – Δ_rG'). If this MDF (– Δ_rG') is > 0, it indicates that all the reactions can operate and the EFM is considered feasible. Hence, maximization of the driving force of an EFM can be achieved by the following linear optimization problem:

$$\begin{aligned} &Maximize\ \boldsymbol{B} \\ &\boldsymbol{x},\ \boldsymbol{B} \\ &Subject\ to - \left(\Delta_r G'^{o} + RT \cdot \boldsymbol{S}^T \boldsymbol{x}\right) \geq \boldsymbol{B} \\ &ln(\boldsymbol{C_{min}}) \leq \boldsymbol{x} \leq ln(\boldsymbol{C_{max}}) \end{aligned}$$

where **B** represents the minimum driving force of all reactions in the EFM, $\boldsymbol{x}$ a vector containing the molar concentrations of the metabolites in **S** within a range of concentrations (C_{min} and C_{max}). Important to note is that the MDF is influenced by the presence/absence of reactions and their direction, but not affected by the flux that each reaction could carry.

For the MDF analysis of each EFM, a custom stoichiometric matrix (**S_adjusted**) was generated by modifying the **S** matrix as follows. Reactions involving metabolites for which the estimation of $\Delta_fG^{o'}$ is highly uncertain were removed (reactions of glycogen and polyphosphate degradation, PHB and PH2MV polymerization) as well as all external reactions from the model. Furthermore, the reaction definitions were adjusted to fit the directionalities of each EFM (either positive or negative fluxes). Reactions carrying no flux (zero values) were removed. The resulting matrix was converted into a tab-separated value (tsv) format for further calculations. MDF calculations were performed using the Equilibrator pathway tool (version 0.4.7) developed by Noor et al. (2014) in the Equilibrator package. As settings for these calculations, we

used default conditions such as a pMg (potential of magnesium) of 3, an ionic strength of 250 mM, a cytosolic pH (potential of hydrogen) of 7.5, and a degree of confidence of 0.95. Simulations varying this degree of confidence were also performed, but the outcome of the model did not affect the results (data not shown). Metabolites in the optimizer were allowed to vary within the default physiologically expected ranges (0.001 to 10 mM), except for phosphate (10 mM), and CO_2 (0.01 mM).

The MDF and optimized Δ_rG' values for each reaction in the models were determined and stored. The MDF distribution of all solutions was visualized using a swarm plot to identify subsets of models that were

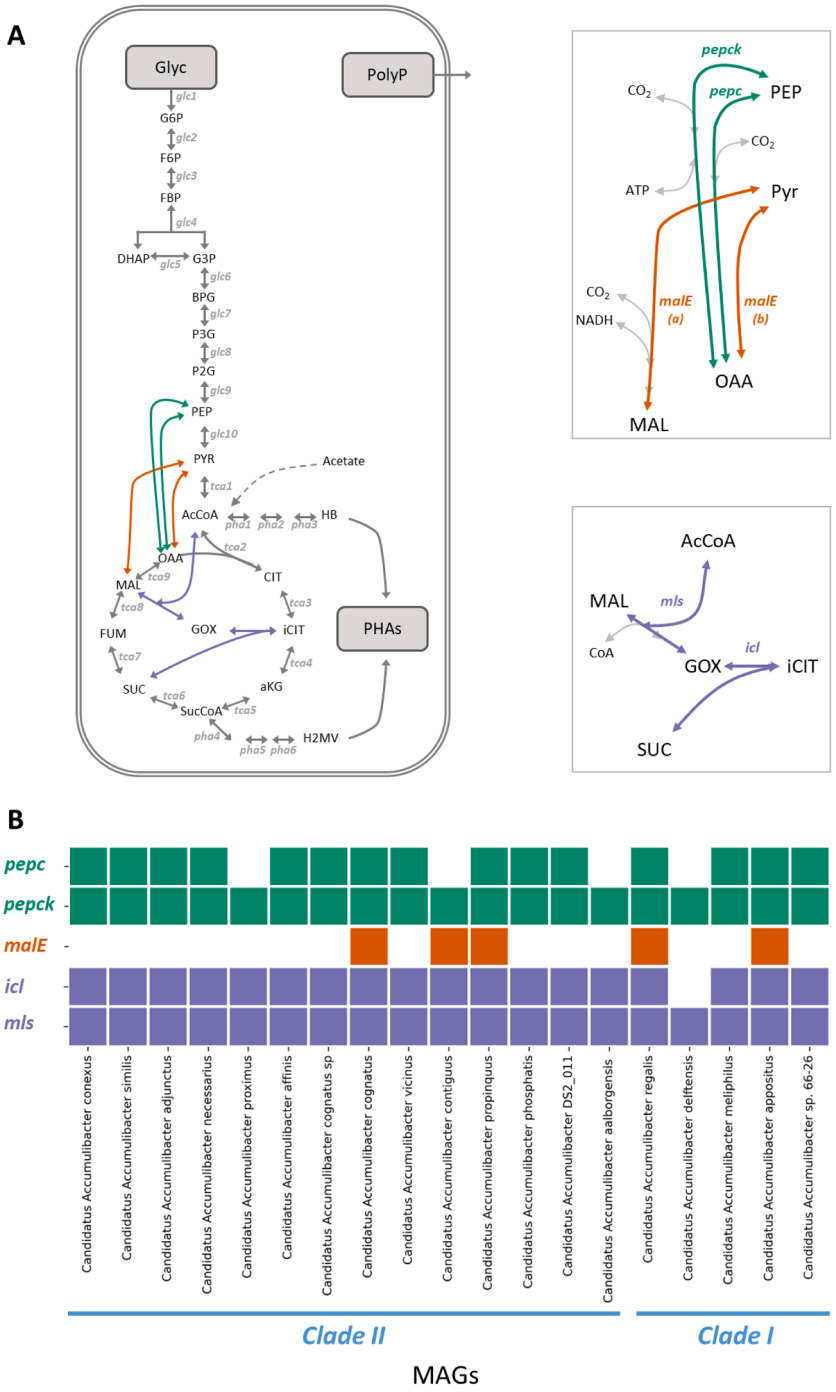


Fig. 2. (A) Schematic representation of the PAOs metabolic model used in this research. (Left) Model schematic including all reactions as reversible except for reactions degrading glycogen (Glyc), polyphosphate (PolyP), acetate uptake and polyhydroxyalkanoate (PHAs) accumulation. Reaction names were given to each reaction and is indicated in italics next to each arrow. (Right) Closer inspection into the anaplerotic routes connecting phosphoenolpyruvate or pyruvate with malate or Oxaloacetate. These reactions are catalysed by the indicated enzymes: PEP-carboxykinase (pepck), PEP-carboxylase (pepc) and malic enzyme (malE). **(B)** Functional potential present in the MAGs of "Ca. Accumulibacter" species related to the presence (filled) or absence (empty) of the genes involved in anaplerotic reactions and/or glyoxylate shunt reactions of central carbon metabolism. MAGs are separated based on their corresponding types (also referred to as clades). Abbreviations: glucose 6 phosphate (G6P), fructose 6 phosphate (F6P), fructose 1, 6 biphosphate (FBP), dihydroxyacetone phosphate (DHAP), glyceraldehyde 3 phosphate (G3P), biphosphoglycerate (BPG), phosphate 3 glycerate (P3G), phosphate 2 glycerate (P2G), phosphoenolpyruvate (PEP), pyruvate (Pyr), acetyl-CoA (AcCoA), citrate (CIT), isocitrate (iCIT), α-keto gluterate (aKG), succinyl-CoA (SucCoA), succinate (SUC), fumarate (FUM), malate (MAL), oxaloacetate (OAA), hydroxyl-butyrate (HB), hydroxyl-e-methyl-valerate (H2MV).

thermodynamically feasible. Feasible reactions were manually selected, and a stoichiometry scheme derived. The achieved optimized Δ_rG' values for each reaction in all the solutions were used to create a bar plot in Python, representing the distribution of Δ_rG'.

To simulate anaerobic conditions, the additional constraint to limit the range of concentrations of NADH and of NAD^+ (0.1 – 10 mM and 0.001 – 0.01 mM respectively) was added to reflect known increase of $NADH/NAD^+$ up to 10 fold from aerobic to anaerobic conditions (Bekers et al., 2015). MDF optimization was then performed on all EFMs with these conditions and resulting MDF and Δ_rG' where analysed in a similar fashion using Python.

2.5. Code availability

The code utilized in this study together with all the data required to adapt or reproduce the simulations is available at GitLab Project ID: 48,899,967 (https://gitlab.com/delft_paos/from-metagenomes-to-metabolism-paos).

3. Results

3.1. The core metabolic model of "Ca. Accumulibacter" is not conserved amongst species

We built a metabolic model based on all previous reports on the metabolism of PAOs. This model included all the reactions involved in polyphosphate and glycogen degradation, PHA synthesis, glycolysis, TCA cycle, glyoxylate shunt and anaplerotic reactions (Fig. 2A). PHAs are modelled as poly-hydroxybutyrate (PHB) and hydroxy 2-methyl-valerate (PH2MV) as a proxy to distinguish PHAs resulting from either acetyl CoA or propionyl CoA.

We confirmed the presence of the genes related to the model in the available MAGs for "*Ca.* Accumulibacter" species (species considered in Table S2). The analysis revealed the absence of genes associated with the glyoxylate shunt and anaplerotic reactions in some species (Fig. 2B.). Specifically, the majority of MAGs analysed lacked the gene coding for malic enzyme (*malE*). Additionally, while all analysed MAGs possessed the gene for phosphoenolpyruvate carboxykinase (*pepck*), some MAGs did not contain the gene for phosphoenolpyruvate carboxylase (*pepc*). In other words, some species of "*Ca.* Accumulibacter" lack the pair *pepc/pepck*, and the majority lack the enzyme *malE*.

3.2. Elementary flux modes analysis (EFMA) reveals more than 700 possible metabolic model solutions to explain the anaerobic metabolism of "Ca. Accumulibacter"

We conducted EFMA on the metabolic model of PAOs. This resulted in individual, elementary flux modes (EFM), each of which contains the fluxes for every reaction such that metabolites are balanced. Thus, each EFM is a unique metabolic model solution that represents a specific, balanced phenotype. Here, we narrowed the EFMs to only solutions that encompassed both acetate uptake and PHA accumulation since this represents the observed metabolism during the anaerobic phase of EBPR. This selection resulted in 700 unique metabolic model solutions. With these 700 solutions, 700 sets of yield coefficients are obtained (Fig. 3).

Among these solutions, we identified flux distributions that have been previously proposed in the literature, including studies by Comeau et al. (1986), Mino et al. (1998), Hesselmann et al. (2000) and Yagci et al. (2003) (highlighted in Fig. 3). The solution corresponding to the model proposed by Pereira et al. (1996) could not be identified since the double net flux of reactions in the TCA cycle is mathematically not possible. Notably, the range of solutions obtained with EFMA is much wider than that of the proposed models alone, suggesting a potentially greater flexibility in terms of polyphosphate/glycogen utilization, PHA accumulation, and even CO_2 incorporation.

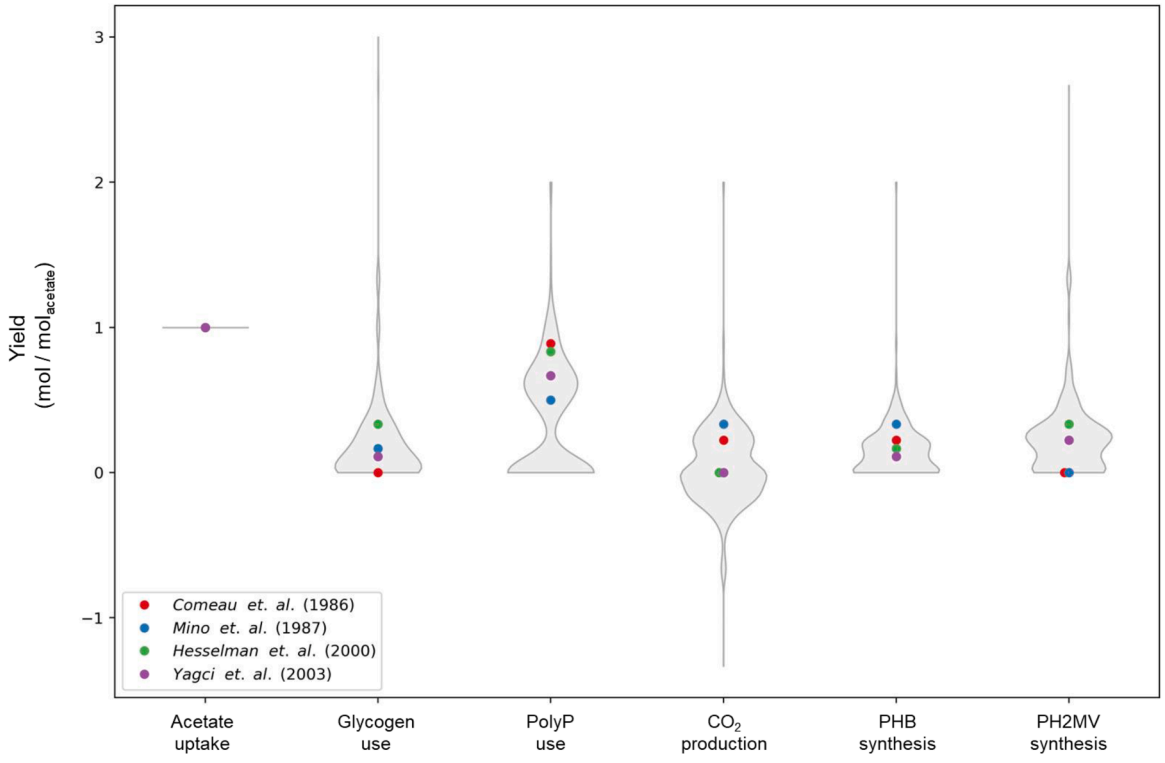


Fig. 3. Model yields of glycogen and polyphosphate use, CO_2 production and PHB and PH2MV synthesis (mole) per mole of acetate consumed in the metabolic model of PAOs.The violin plots illustrate the distribution of stoichiometric values based on 700 unique solutions. Published solutions from previous studies by Comeau et al., (1986); Mino et al., (1998); Hesselmann et al., (2000) and Yagci et al., (2003) are indicated by differently coloured circles and lines. Negative stoichiometries indicate the reverse direction of the proposed reactions, such as the production or consumption of CO_2 within the different solutions.

3.3. A small subset of the metabolic model solutions are thermodynamically feasible

To further reduce the 700 metabolic model solutions, the thermodynamic feasibility of each solution was evaluated by means of calculating the MDF. The MDF algorithm maximizes the driving force (defined as the maximum - $\Delta_r G'$ for each reaction) of a given metabolic model solution (Noor et al., 2014) (see Materials and Methods). MDF suggests that only 20.9 % of the 700 metabolic model solutions result in a thermodynamically feasible pathway, *i.e.* all individual metabolic reactions proceed in the required direction with a negative $\Delta_r G'$ (Fig. 4A). The models proposed by Mino et al. (1998), Hesselmann et al. (2000) and Yagci et al. (2003) are within the feasible solutions. Nevertheless, these proposed models show an MDF value close to equilibrium (MDF of 1.18, 1.8 and 0.14 kJ/$mol_{acetate}$ respectively). In contrast, for the model proposed by Comeau et al. (1986) no thermodynamically feasible solution (MDF = -4.06 kJ/$mol_{acetate}$) was found.

The obtained 146 thermodynamically feasible metabolic model solutions can be further reduced considering that a significant dissipation of energy is required to obtain a significant reaction rate. From the current solutions, 6 metabolic models indicated the highest energy dissipation with MDF values above 7 kJ/$mol_{acetate}$ (Fig. 4A and B for model operations with an MDF of 7.46 kJ/$mol_{acetate}$). All these solutions include either the combination of *pepc/pepck* or the combination of *pepc* with *malE*. Notably, all these metabolic model solutions result in the same yield coefficients for glycogen degradation matching that of PH2MV accumulation and no PHB accumulation.

3.4. Dependency of reaction feasibility on the NADH/NAD^+ ratio

The thermodynamic driving force of each metabolic model solution was the result of a combination of the individual reaction's $\Delta_r G'$. Depending on the set of active reactions in each EFM, a specific reaction can either be thermodynamically feasible or infeasible. The reason for this observation is coupling of reactions via shared metabolites, which make the system more or less constrained depending on the EFM. Specific reactions consistently exhibited high infeasibilities ($\Delta_r G' > 0$) across multiple EFMs. Fig. 5 represents an overview of the achieved $\Delta_r G'$ values for each reaction on both the forward and, when applicable, reverse directions in all 700 EFMs (see also *Table S1* for detailed information on stoichiometry of each reaction and the forward and reverse operation).

Reactions that consistently displayed thermodynamic infeasibilities include *glc5, glc6, tca7, tca9, pepc, pha1 and pha4* in the forward direction (Fig. 5A) and *glc10, tca1, tca2, tca3, tca4, tca5, pepck, icl, mls* and *malE* in the reverse direction (Fig. 5B). Thus, these reactions (in the indicated direction) contribute to the infeasibility of the solutions where these reactions are included. Conversely, reactions that consistently exhibited negative $\Delta_r G'$ values include *tca1, tca2, pepck, mls, MalE, pha3, pha5* and *pha6* in the forward direction and *tca7, tca9* and *pepc* in the reverse direction; indicating their strong contribution to feasible networks.

The MDF optimization was further constrained to reflect expected changes in the ratio of NADH/NAD^+ when considering anaerobic *vs* aerobic conditions (Bekers et al., 2015) to showcase the limitations posed under anaerobic conditions. For this analysis, we fixed the NADH/NAD^+ ratio to be 0.1 and 10 which reflects the change in this ratio between aerobic and anaerobic conditions. With these fixed values, we performed MDF and represented the distribution of $\Delta_r G'$ under each condition (Fig. 5). This modification led to changed $\Delta_r G'$ distributions for reactions involving oxidation or reduction of NAD(H). Specifically, reactions generating NADH resulted in higher $\Delta_r G'$ values leading to potential infeasibilities. Notably, reactions *glc6, tca7* and *tca9* running in forward direction (Fig. 5A) were the strongest affected by the change in the NADH/NAD^+ since their $\Delta_r G'$ became more positive at a higher ratio potentially indicating their (even stronger) infeasibility under anaerobic conditions.

4. Discussion

4.1. Bridging metagenomics with metabolic function predictions

This research paper presents an integrated workflow designed to extract feasible metabolic predictions within a specific environmental context from a MAG's metabolic model. This systematic analysis serves

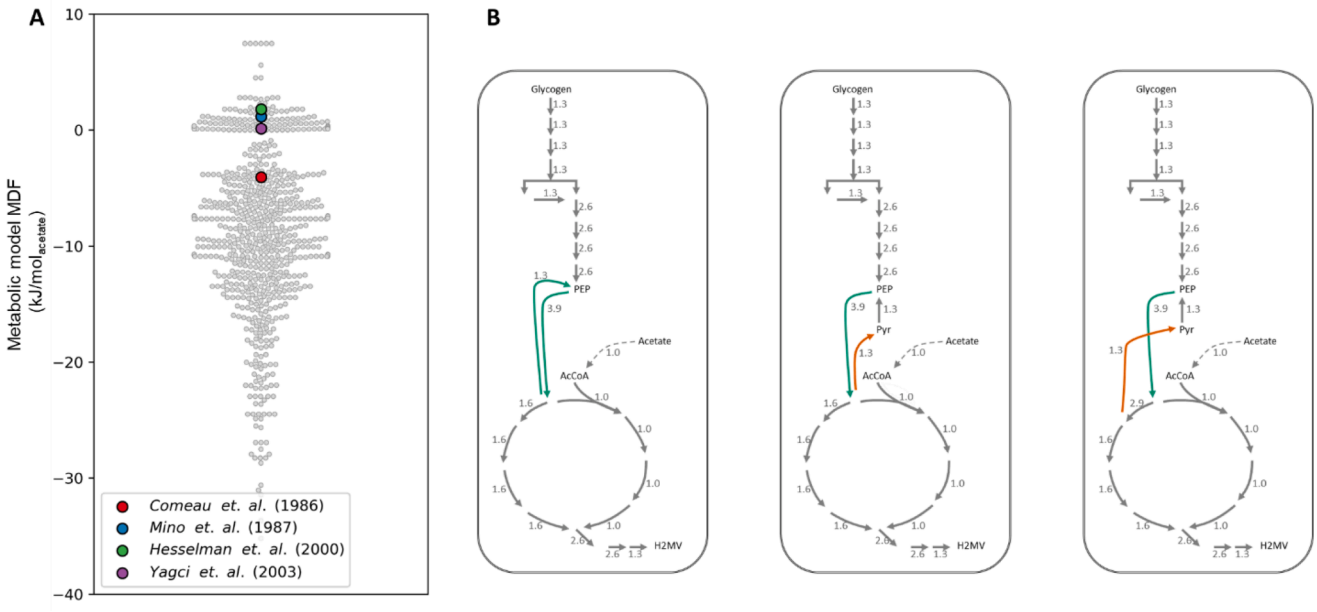


Fig. 4. (A) A swarm plot displaying the distribution of the Minimal Driving Force (MDF) achieved by all 700 metabolic model solutions. The MDF results based on the models proposed by Comeau et al., (1986), Mino et al., (1998), Hesselmann et al., (2000) and Yagci et al., (2003) are highlighted in distinct colours and sizes. **(B)** Flux maps with highest MDF of 7.46 kJ/ $mol_{acetate}$. Among the top metabolic model solutions, we showcase three flux maps out of a total of six. The six top metabolic models pertain to three solution pairs, each with and without polyphosphate degradation (but solutions with the same combination of reactions). Note that polyphosphate degradation was excluded from the thermodynamic analysis due to the high uncertainty of the $\Delta_f G^o$ of this polymer.

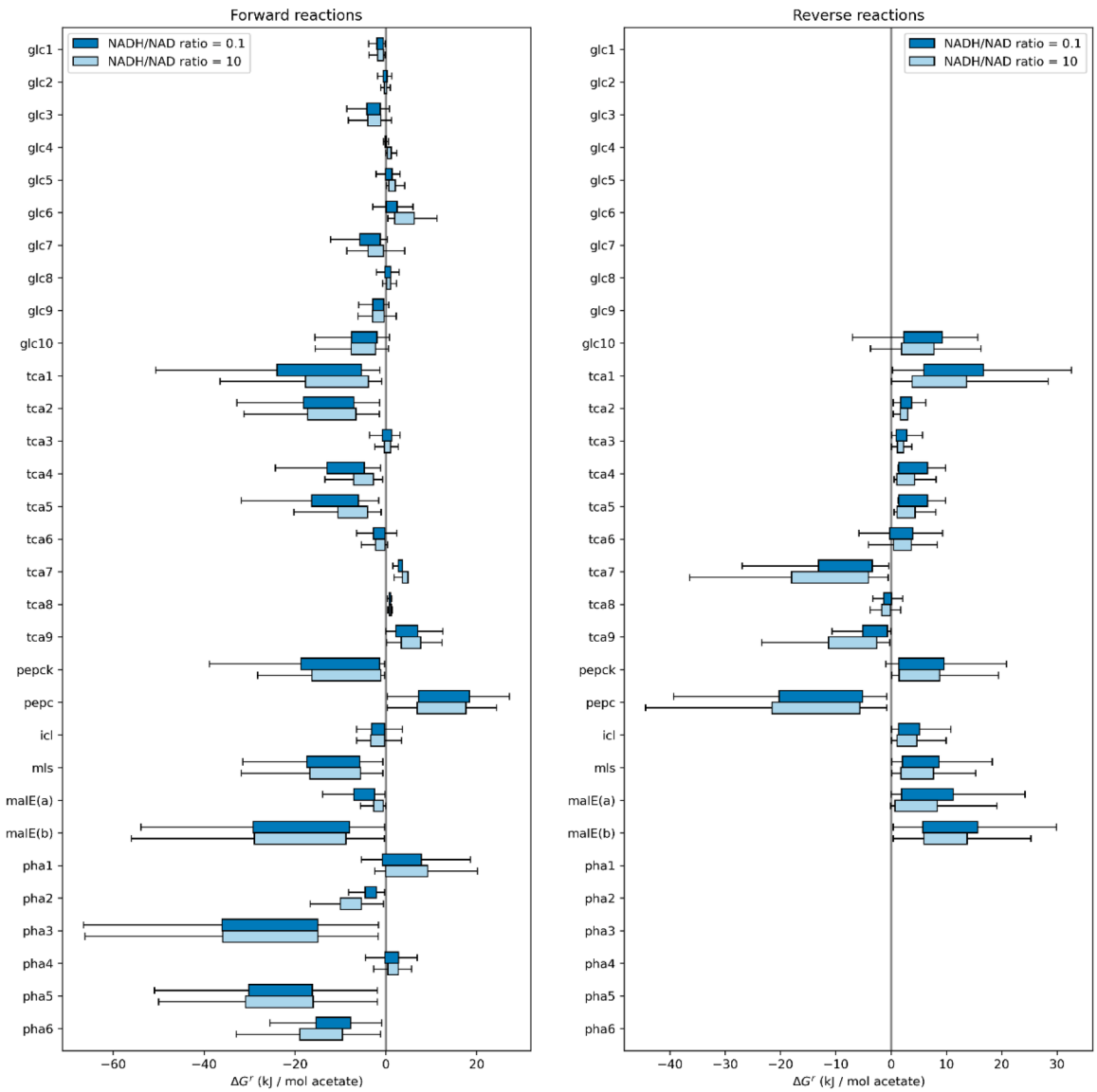


Fig. 5. Boxplot distribution of the calculated $\Delta_r G'$ for each reaction based on the 700 Elementary Flux Modes (EMF) solutions and respective Minimal Driving Force (MDF) optimization. Dark and light blue indicate different NADH/NAD^+ ratios that were fixed for the MDF analysis simulating changes from aerobic to anaerobic conditions respectively. Distribution of the reactions in **(A)** forward or **(B)** reverse operation. Absent boxplot indicates that the reaction in its given operation does not contribute to the solution space of the model.

as a crucial step towards transitioning from descriptive studies in microbiome research towards more mechanistic and predictive studies, also called eco-systems frameworks (Lawson et al., 2019; McDaniel et al., 2021). With the increasing availability of high quality MAGs, research has rapidly focused on assigning metabolic functions to given species based on gene presence/absence (Singleton et al., 2021; Woodcroft et al., 2018; Aalismail et al., 2019; Kristensen et al., 2021; Qiu et al., 2019; Singleton et al., 2022) and, in limited cases, integrating multi-omic data (McDaniel et al., 2021; Ishii et al., 2013; Camejo et al., 2019; Roume et al., 2015; Scarborough et al., 2020). We emphasize the importance of subjecting metabolic networks to critical examination within the specific environmental context of the community, aided by modelling tools like those presented in this work, to evaluate all potential functionalities of community members in a consistent manner. This essential step precedes the integration of multi-species metabolic models aimed at predicting (or even engineering) microbial communities.

Generating metabolic models from MAGs is becoming easier; more advanced, automated tools are developed and support model construction from metagenomes (Bekiaris and Klamt 2020; Zimmermann et al., 2021; Machado et al., 2018). However, caution must be taken when performing simulations on metabolism. We have showcased just how diverse the possible solutions to a simple metabolic model can be (by employing EFMA) and more importantly how many of these potential solutions represent thermodynamic infeasibilities (by using MDF analysis). Our approach adds a general/universal test on the stoichiometric solutions – universal as thermodynamics are calculated for all possible network solutions (reaction combinations) and do not rely on a specific assumptions on one reaction or only on energy generation cycles (Fritzemeier et al., 2017) as done with tools such as *gap-seq* or *CarveMe*.

The analysis presented in this paper faces two limitations related to input data. Firstly, meta-omic data from environmental samples (due to its inherent complexity) poses challenges on sequence completeness. Even assuming complete MAGs, we are limited to studying metabolic reactions with documented descriptions, leaving out many potential genes/enzymes not catalogued in databases. For instance, it is estimated that in the genome of *E. coli* the function of 35 % of coding sequences is

unknown (Ghatak et al., 2019)). The former is a technical limitation that can be overcome with better quality data. Thus we recommend the application of this workflow with confidence on MAGs generated combining both short and long read sequencing data (as demonstrated in Singleton et al. (2021)). Conversely, the latter limitation persists beyond the advancements of current omic-techniques. Metagenomics, metatranscriptomics and metaproteomics are insufficient to address functionality of coding sequences. To uncover novel functions within organisms, a different approach must be taken. The combination of metabolic modelling predictions with experimental validation could be powerful in indicating how *incomplete* our state-of-art knowledge on metabolism is.

4.2. Physiological differences amongst "Ca. Accumulibacter" species

We applied our workflow to "*Ca.* Accumulibacter" species since these organisms have been vastly studied and recently an exponential growth in the available genomes has become available (Petriglieri et al., 2022). In contrast to previous studies, which focused on finding '*the correct*' stoichiometry for anaerobic acetate uptake in PAOs, our analysis provides a holistic analysis of underlying metabolic potential within a basic model of central carbon metabolism. Building upon the concept of 'metabolic flexibility' proposed by da Silva et al. (2020) for PAOs, we emphasize the idea of understanding metabolism in a non-linear manner, where multiple metabolic branches (here seen as nodes) can coexist. However, the absence of essential genes in a network can limit this 'flexibility' and may indicate differences in the ecological niches of individual "*Ca.* Accumulibacter" species.

One notable observation when comparing the MAGs of "*Ca* Accumulibacter" species was the absence of the anaplerotic route enzyme *malE* in almost all studied species, and the *pepc/pepck* pair in certain species (Fig. 2B). We found that the presence of these reactions (especially their combination) contributed to the highest driving forces of the metabolic models (Fig. 4B). This suggests that the species lacking *malE* and/or the pair *pepck/pepc* may be more thermodynamically limited and would exhibit a considerably lower flexibility during anaerobic substrate uptake. On the other hand, the species "*Ca.* Accumulibacter cognatus", "*Ca.* Accumulibacter propinquus", "*Ca.* Accumulibacter regalis" and "*Ca.* Accumulibacter appositus" exhibited the complete genetic potential for the model studied, indicating both stoichiometric and thermodynamic flexibility during anaerobic substrate uptake.

Furthermore, it is important to note that *malE* and *pepc/pepck* play crucial roles in controlling flux from the TCA cycle towards gluconeogenesis, replenishing bacterial glycogen reserves (Sauer and Eikmanns 2005). Therefore, while our research focused on the anaerobic uptake of acetate, the absence of these enzymes could have broader consequences during aerobic conversions of "*Ca.* Accumulibacter" that have not been assessed. Previous studies have shown that the lack of *malE* leads to imbalanced metabolic rates, altered storage reserves (Zhang et al., 2016) and substantial reduction of growth yields (Netzer et al., 2004). Similarly, the *pepc/pepck* pair has been shown to participate in futile cycles in *E. coli* (Hua et al., 2003). Overall, a better understanding of the absence of these genes through the complete EBPR anaerobic/aerobic cycle is needed.

4.3. Redefining the anaerobic metabolic potential of "Ca. Accumulibacter"

Decades of research into PAO biochemistry have led to the progressive proposition of individual models to explain the anaerobic uptake of acetate. Here we examine individual model operations proposed over the years in the context of our results and propose an updated metabolic model for the anaerobic uptake of acetate in "*Ca.* Accumulibacter".

Firstly, the stoichiometric analysis aided by EFMA resulted in 700 possible unique metabolic model solutions to the model (Fig. 3). This number of possible stoichiometries explains the lack of consensus on the proposed model operations over the years. We found that the suggested operations by Comeau et al. (1986), Mino et al. (1998), Hesselmann et al. (2000) and Yagci et al. (2003) are indeed stoichiometrically possible. However, the model proposed by (Pereira et al., 1996) did not appear in these solutions; since this model suggests the simultaneous operation of the forward and reverse direction of a portion of the TCA cycle (see Fig. 1), which as a net flux is impossible. EFMA can therefore aid to identify first-hand the possible operations of a metabolic model, without depending on experimental observations.

EFMA generated 700 metabolic model solutions, of which 21 % were also thermodynamically feasible. Many of these solutions, however, operate close to thermodynamic equilibrium. The model solutions with the highest thermodynamic feasibility made use of a split TCA cycle together with an additional cycle using either *pepc/pepck* or *pepc/malE* (Fig. 4B). These solutions result in an MDF of more than 7 $kJ/mol_{acetate}$, by means of converting all the carbon from acetate to CO_2, leading to the high energy dissipation of the system. By integrating the distribution of each individual reaction's Δ_rG' values under anaerobic conditions (Fig. 5), we outline the metabolic model for "*Ca.* Accumulibacter" in Fig. 6, highlighting reactions that represent strong thermodynamic infeasibilities. Below we discuss on the implications of specific reactions on the model for "*Ca.* Accumulibacter".

The notion that the full anaerobic TCA cycle can occur to generate enough NADH for the accumulation of PHAs is highly debated in literature. At the centre of this debate lies succinate dehydrogenase (reaction *tca5* in our network), a membrane bound enzyme that transfers the electrons from succinate to FADH. In the absence of a strong electron acceptor (like oxygen), it's thermodynamically impossible to transfer these electrons to NADH. We modelled this reaction with NADH rather than FADH, lumping the overall transfer of electrons from succinate to FADH, to the quinone pool and finally to NADH, and confirm the thermodynamic impossibility of this reaction (Fig. 5). Some theories suggest succinate-to-fumarate conversion could occur through a novel cytochrome b/b6 fusion protein (Martín et al., 2006) or electron bifurcation mechanisms (Oyserman et al., 2016). However, attempts at assessing this activity have failed (Gabriel Guedes da Silva, 2021). Future work could be done using this model framework to study these alternative routes and its effects on changing the thermodynamics and yields of the anaroibc stoichiometry. We conclude that, given the current evidence, the conventional succinate dehydrogenase reaction cannot occur under anaerobic conditions.

We further add to the notion that the full TCA cycle cannot occur anaerobically since malate dehydrogenase (*tca9* in our network) represents a thermodynamically infeasible reaction in anaerobic conditions. This reaction has been known to have very high energy requirements (Miller and Smith-Magowan 1990) thus several hypotheses exist to explain the function of the full TCA cycle. One argument is that the concentration of oxaloacetate is extremely low in cells (below the used lower bound concentration in the MDF analysis = 1 μM), although this could result in other reactions that utilize oxaloacetate to become infeasible (Noor et al., 2014). Assuming this is the case, the redox state of the cell (*i.e.* $NADH/NAD^+$) would be a major determinant of the feasibility of this reaction. Although not measured in "*Ca.* Accumulibacter" (nor in any microbial community of this kind), the $NADH/NAD^+$ ratios of bacteria and yeast when changing from aerobic to anaerobic conditions increases as much as 10 fold (Wimpenny and Firth 1972; Bekers et al., 2015; Liu et al., 2019), which would pull this reaction towards the infeasibility region. Further, van der Rest, Frank, and Molenaar (2000) described that this reaction could also be catalysed by a membrane bound enzyme that reduces a quinone pool, thus making the reaction more thermodynamically favourable. However, da Silva et al. (2020) found a high preference of malate dehydrogenase for NADH. Altogether the evidence shows that malate dehydrogenase cannot operate towards generating oxaloacetate and NADH under anaerobic conditions and instead is more likely to operate under

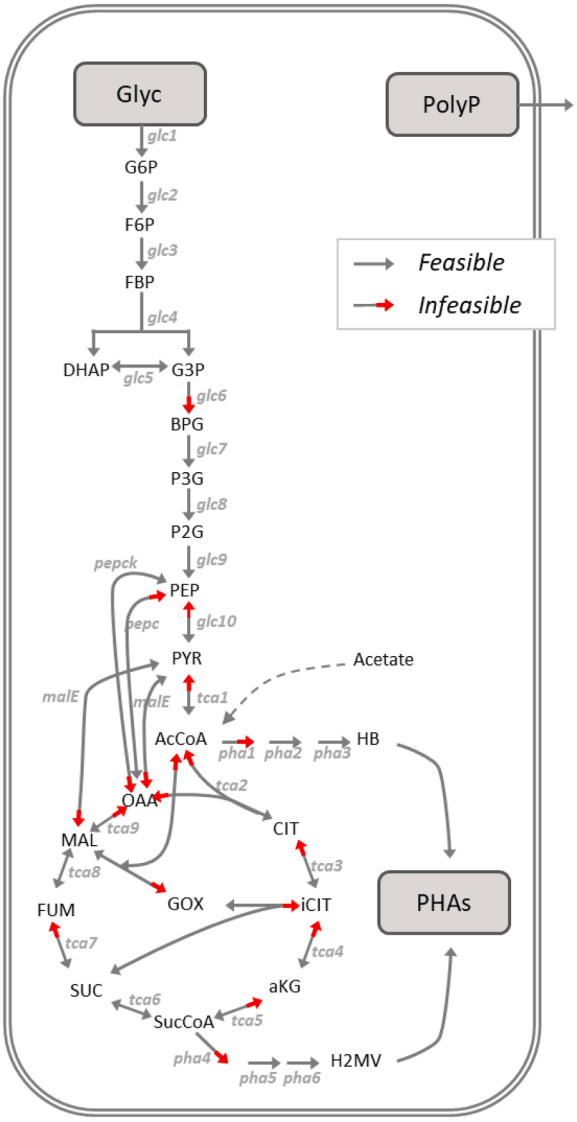


Fig. 6. Redefinition of the PAOs metabolic model for anaerobic uptake of acetate and accumulation of PHAs. Scheme represents the reactions (grey arrows) connecting metabolites participating in central carbon metabolism. Based on thermodynamic analysis, each reaction of the pathway has been defined as feasible (when $\Delta_rG' < 0$ kJ/mol$_{acetate}$), irreversible (when the reverse reaction could contribute to the stoichiometry, but it's $\Delta_rG' > 0$ kJ/molacetate) or bottleneck (when $\Delta_rG' > 0$ kJ/mol$_{acetate}$ but no other alternatives are possible).

anaerobic conditions in the opposite reaction, a phenomenon previously observed in *E. coli* cells (Chen et al., 2011).

Finally, we represent three reactions of the metabolic model to be thermodynamically infeasible albeit necessary to obtain the observed phenotype of "*Ca.* Accumulibacter". The reactions are glyceraldehyde 3-phosphate dehydrogenase (*glc6* in our network), acetoacetyl CoA synthase (*pha1* in our network) and methylmalonyl CoA mutase (*pha4* in our network). Firstly, *glc6* is a well-known thermodynamic bottleneck reaction (Noor et al., 2014). To overcome this bottleneck, alternative glycolytic pathways exist such as the Entner-Doudoroff glycolysis (ED) or the pentose phosphate pathway (PPP) that bypass this reaction. However, genomic (Martín et al., 2006), proteomic (Wilmes et al., 2008) and enzymatic (Guedes da Silva et al. 2019) evidence support the notion that both ED and the complete PPP do not operate in "*Ca.* Accumulibacter". Hence, the reaction *glc6* is the only alternative in the current metabolic model to carry flux from glucose towards lower glycolysis. Further research must be done to explore weather thermodynamic 'tricks' exist to enable this reaction to occur or even if different, yet unknown, glycolytic routes are active in PAOs anaerobically. On the other hand, the reactions *pha1* and *pha4* involved in the synthesis of PHAs are immediately followed by two reactions (*pha2* and *pha3* for PHB and *pha5* and *pha6* for PH2MV) with very high thermodynamic driving force (Fig. 5). Although this process has not been studied for PHA synthesis, a very similar set of reactions to *pha1* in archaea (acetoacetyl-CoA thiolase) catalyze a similarly thermodynamically 'uphill' reaction connected to 'downhill' reactions. Vögeli et al. (2018) found evidence that substrate channeling via enzyme complexes exists and could explain the pathway to occur despite this bottleneck. A similar mechanism could be at play in "*Ca.* Accumulibacter" since PHA synthesis is observed anaerobically.

It is important to note that our thermodynamic analysis only considers intracellular metabolic reactions and excludes exchange reactions and reactions involved in polymeric synthesis or degradation. The thermodynamics of exchange reactions (for example, acetate uptake, phosphate export, etc.) not dependent on metabolic network control. On the other hand, polymers are substances that are not completely in aqueous solution as a typical metabolite is (hence they can be *accumulated* to extremely high amounts). Thus, the incorporation of polymers and their *concentration's* contribution to a reaction's thermodynamic driving force requires a deeper physical understanding of polymers as a whole. Further research should be done to understand the complexity that different kinds of polymers contribute to metabolic reactions.

4.4. Trade-offs between stoichiometric yield and thermodynamic feasibility

The metabolic model solutions that resulted in the highest MDF (Fig. 4.B) released the carbon from acetate as CO_2 and formed the more reduced form of PHA (*i.e.* PH2MV). The PH2MV accumulated in these solutions came directly from the amount of glycogen degraded. Although these solutions are thermodynamically the most feasible, they will result in very poor growth yields since none of the acetate accumulates as PHAs. This leads to an interesting relationship between thermodynamic efficiency (proportional to the rate) *vs* the stoichiometric efficiency (*i.e.* yield). This apparent relationship stems from the trade-off between either converting acetate into CO_2 to gain driving force (and thus uptake rates) against conserving the acetate into PHB that later can be used to grow more efficiently (improve biomass yields). Such trade-off between rate and yield has been described previously for growth rate (Pfeiffer et al., 2001; Lipson 2015), but since PAOs grow in a cyclic environment, such relationship has not been previously described and represent an interesting research potential.

In this context, long term enrichments under fast *vs* slow feeding rates could offer insights into microbial communities prioritizing uptake rates or efficiencies. Similarly, applying a substrate mix that alleviate kinetic bottlenecks in substrate uptake rates (as demonstrated in (Qiu et al., 2019)) may enhance uptake rates, revealing previously unexplored metabolic strategies. The integration of such experimental findings with metabolic modelling can illuminate key aspects for metabolic control, fostering hypothesis generation to improve or select microbial communities functions.

5. Conclusions

- Metabolic potential inferred from MAGs-only is misleading since many potential reactions from a metabolic model can be thermodynamically infeasible under specific environmental conditions.
- In the current post-genomic era, we should not predefine reaction directions based on presence of marker genes, but rather systematically evaluate their potential given metabolic and environmental context.

- The combination of EFMA with MDF calculations provides a suitable framework for the assessment of metabolic pathway feasibility given a MAG-derived metabolic model.
- Species of "*Ca.* Accumulibacter" possess different potential in their central carbon metabolism specifically related to anaplerotic reactions.
- The full operation of a classical TCA cycle (oxidating direction) is, according to current knowledge, not possible under anaerobic acetate uptake due to thermodynamic infeasibilities in succinate dehydrogenase and malate dehydrogenase.
- Anaerobic PHA accumulation in "*Ca.* Accumulibacter" results from a trade-off between thermodynamic feasibility and stoichiometric yield.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to improve the clarity and readability of the Python scripts and fractions of this text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Timothy Páez-Watson: Conceptualization, Investigation, Formal analysis, Writing – original draft. **Mark C.M. van Loosdrecht:** Conceptualization, Investigation, Writing – original draft. **S.Aljoscha Wahl:** Conceptualization, Investigation, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A link to the data and code has been shared in the manuscript.

Acknowledgments

This work was supported by SIAM Gravitation Grant 024.002.00, the Netherlands Organization for Scientific Research (NWO).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2023.121028.

References

Aalismail, N.A., Ngugi, D.K., Díaz-Rúa, R., Alam, I., Cusack, M., Duarte, C.M., 2019. Functional metagenomic analysis of dust-associated microbiomes above the red Sea. Sci. Rep. 9, 13741.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 13219.

Bekers, K.M., Heijnen, J.J., Van Gulik, W.M., 2015. Determination of the *in vivo* NAD: NADH ratio in Saccharomyces cerevisiae under anaerobic conditions, using alcohol dehydrogenase as sensor reaction. Yeast 32, 541–557.

Bekiaris, P.S., Klamt, S., 2020. Automatic construction of metabolic models with enzyme constraints. BMC Bioinform. 21, 1–13.

Burow, L.C., Mabbett, A.N., Blackall, L.L., 2008. Anaerobic glyoxylate cycle activity during simultaneous utilization of glycogen and acetate in uncultured Accumulibacter enriched in enhanced biological phosphorus removal communities. ISME J. 2, 1040–1051.

Camejo, P.Y., Oyserman, B.O., McMahon, K.D., Noguera, D.R., 2019. Integrated omic analyses provide evidence that a "*Candidatus* Accumulibacter phosphatis" strain performs denitrification under microaerobic conditions. Msystems 4.

Chen, X., Alonso, A.P., Allen, D.K., Reed, J.L., Shachar-Hill, Y., 2011. Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in E. coli. Metab. Eng. 13, 38–48.

Comeau, Y., Hall, K.J., Hancock, R.E.W., Oldham, W.K., 1986. Biochemical model for enhanced biological phosphorus removal. Water Res. 20, 1511–1521.

da Silva, L.G., Gamez, K.O., Gomes, J.C., Akkermans, K., Welles, L., Abbas, B., van Loosdrecht, M.C.M., Wahl, S.A., 2020. Revealing the metabolic flexibility of "*Candidatus* Accumulibacter phosphatis" through redox cofactor analysis and metabolic network modeling. Appl. Environ. Microbiol. 86.

Dillard, L.R., Payne, D.D., Papin, J.A., 2021. Mechanistic models of microbial community metabolism. Mol. Omics 17, 365–375.

Du, H., Pan, J., Zou, D., Huang, Y., Liu, Y., Li, M., 2022. Microbial active functional modules derived from network analysis and metabolic interactions decipher the complex microbiome assembly in mangrove sediments. Microbiome 10, 1–17.

Dueholm, M.K.D., Besteman, M., Zeuner, E.J., Riisgaard-Jensen, M., Nielsen, M.E., Vestergaard, S.Z., Heidelbach, S., Bekker, N.S., Nielsen, P.H., 2023. Genetic potential for exopolysaccharide synthesis in activated sludge bacteria uncovered by genome-resolved metagenomics. Water Res. 229, 119485.

Famili, I., Förster, J., Nielsen, J., Palsson, B.O., 2003. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc. Natl. Acad. Sci. 100, 13134–13139.

Fritzemeier, C.J., Hartleb, D., Szappanos, B., Papp, B., Lercher, M.J., 2017. Erroneous energy-generating cycles in published genome scale metabolic networks: identification and removal. PLoS Comput. Biol. 13, e1005494.

Fujii, N., Kuroda, K., Narihiro, T., Aoi, Y., Ozaki, N., Ohashi, A., Kindaichi, T., 2022. Metabolic potential of the superphylum Patescibacteria reconstructed from activated sludge samples from a municipal wastewater treatment plant. Microbes Environ. 37, ME22012.

Gerstl, M.P., Jungreuthmayer, C., Müller, S., Zanghellini, J., 2016. Which sets of elementary flux modes form thermodynamically feasible flux distributions? FEBS J. 283, 1782–1794.

Ghatak, S., King, Z.A., Sastry, A., Palsson, B.O., 2019. The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. Nucleic Acids Res. 47, 2446–2454.

Gabriel Guedes da Silva, L. (2021). Life in changing environments: The intriguing cycles of Polyphosphate Accumulating Organisms. [Dissertation (TU Delft), Delft University of Technology]. doi:10.4233/uuid:944a1cc2-9a30-4b84-91b9-5022d689d7f3.

Guedes da Silva L., S. Tomás-Martínez, M.C.M. van Loosdrecht, and S.A. Wahl. 2019. 'The environment selects: modeling energy allocation in microbial communities under dynamic environments', *bioRxiv*: 689174.

Hesselmann, R.P.X., Von Rummell, R., Resnick, S.M., Hany, R., Zehnder, A.J.B., 2000. Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. Water Res. 34, 3487–3494.

Hua, Q., Yang, C., Baba, T., Mori, H., Shimizu, K., 2003. Responses of the central metabolism in Escherichia coli to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. J. Bacteriol. 185, 7053–7067.

Ishii, S., Suzuki, S., Norden-Krichmar, T.M., Tenney, A., Chain, P.S.G., Scholz, M.B., Nealson, K.H., Bretschger, O., 2013. A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. Nat. Commun. 4, 1601.

Kim, M., Sung, J., Chia, N., 2022. Resource-allocation constraint governs structure and function of microbial communities in metabolic modeling. Metab. Eng. 70, 12–22.

Kristensen, J.M., Singleton, C., Clegg, L.A., Petriglieri, F., Nielsen, P.H., 2021. High diversity and functional potential of undescribed "Acidobacteriota" in Danish wastewater treatment plants. Front. Microbiol. 12, 643950.

Kuppa, B., Kumar, D., Umale, S., Zhou, Z., Raman, K., Anantharaman, K., 2023. Metagenome-based metabolic modelling predicts unique microbial interactions in deep-sea hydrothermal plume microbiomes. ISME Commun. 3, 42.

Lawson, C.E., Harcombe, W.R., Hatzenpichler, R., Lindemann, S.R., Löffler, F.E., O'Malley, M.A., Martín, H.G., Pfleger, B.F., Raskin, L., Venturelli, O.S., 2019. Common principles and best practices for engineering microbiomes. Nat. Rev. Microbiol. 17, 725–741.

Lipson, D.A., 2015. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. Front. Microbiol. 6, 615.

Liu, Y., Landick, R., Raman, S., 2019. A regulatory NADH/NAD+ redox biosensor for bacteria. ACS Synth. Biol. 8, 264–273.

Maarleveld, T.R., Khandelwal, R.A., Olivier, B.G., Teusink, B., Bruggeman, F.J., 2013. Basic concepts and principles of stoichiometric modeling of metabolic networks. Biotechnol. J. 8, 997–1008.

Machado, D., Andrejev, S., Tramontano, M., Patil, K.R., 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 46, 7542–7553.

Martín, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., Yeates, C., He, S., Salamov, A.A., Szeto, E., 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities'. Nat. Biotechnol. 24, 1263–1269.

McDaniel, E.A., Moya-Flores, F., Beach, N.K., Camejo, P.Y., Oyserman, B.O., Kizaric, M., Khor, E.H., Noguera, D.R., McMahon, K.D., 2021a. Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics. Msystems 6 e00474-21.

McDaniel, E.A., Wahl, S.A., Ishii, S., Pinto, A., Ziels, R., Nielsen, P.H., McMahon, K.D., Williams, R.B.H., 2021b. Prospects for multi-omics in the microbial ecology of water engineering. Water Res. 205, 117608.
Miller, S.L., Smith-Magowan, D., 1990. The thermodynamics of the Krebs cycle and related compounds. J. Phys. Chem. Ref. Data 19, 1049–1073.
Mino, T., Arun, V., Tsuzuki, Y., Matsuo, T., 1987. Effect of phosphorus accumulation on acetate metabolism in the biological phosphorus removal process. Biological Phosphate Removal from Wastewaters. Elsevier.
Mino, T., Van Loosdrecht, M.C.M., Heijnen, J.J., 1998. Microbiology and biochemistry of the enhanced biological phosphate removal process. Water Res. 32, 3193–3207.
Muller, E.E.L., Faust, K., Widder, S., Herold, M., Arbas, S.M., Wilmes, P., 2018. Using metabolic networks to resolve ecological properties of microbiomes. Curr. Opin. Syst. Biol. 8, 73–80.
Netzer, R., Krause, M., Rittmann, D., Peters-Wendisch, P.G., Eggeling, L., Wendisch, V.F., Sahm, H., 2004. Roles of pyruvate kinase and malic enzyme in Corynebacterium glutamicum for growth on carbon sources requiring gluconeogenesis. Arch. Microbiol. 182, 354–363.
Nielsen, P.H., McIlroy, S.J., Albertsen, M., Nierychlo, M., 2019. *Re*-evaluating the microbiology of the enhanced biological phosphorus removal process. Curr. Opin. Biotechnol. 57, 111–118.
Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., Milo, R., 2014. Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput. Biol. 10, e1003483.
O'Brien, E.J., Monk, J.M., Palsson, B.O., 2015. Using genome-scale models to predict biological capabilities. Cell 161, 971–987.
Oehmen, A., Lemos, P.C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L.L., Reis, M.A.M., 2007. Advances in enhanced biological phosphorus removal: from micro to macro scale. Water Res. 41, 2271–2300.
Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.
Oyserman, B.O., Noguera, D.R., Rio, T.G., Tringe, S.G., McMahon, K.D., 2016. Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus* Accumulibacter phosphatis. ISME J. 10, 810–822.
Páez-Watson, T., van Loosdrecht, M.C.M., Wahl, S.A., 2023. Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: application to polyphosphate accumulating organisms. Water Res. 228, 119365.
Pereira, H., Lemos, P.C., Reis, M.A.M., Crespo, J.P.S.G., Carrondo, M.J.T., Santos, H., 1996. Model for carbon metabolism in biological phosphorus removal processes based on *in vivo*13C-NMR labelling experiments. Water Res. 30, 2128–2138.
Peres, S., Jolicœur, M., Moulin, C., Dague, P., Schuster, S., 2017. How important is thermodynamics for identifying elementary flux modes? PLoS One 12, e0171440.
Petriglieri, F., Singleton, C.M., Kondrotaite, Z., Dueholm, M.K.D., McDaniel, E.A., McMahon, K.D., Nielsen, P.H., 2022. Reevaluation of the phylogenetic diversity and global distribution of the genus "*Candidatus* Accumulibacter". Msystems 7, e00016–e00022.
Pfeiffer, T., Schuster, S., Bonhoeffer, S., 2001. Cooperation and competition in the evolution of ATP-producing pathways. Science 292, 504–507.
Qiu, G., Liu, X., Saw, N.M.M.T., Law, Y., Zuniga-Montanez, R., Thi, S.S., Nguyen, T.Q.N., Nielsen, P.H., Williams, R.B.H., Wuertz, S., 2019. Metabolic traits of *Candidatus* Accumulibacter clade IIF strain SCELSE-1 using amino acids as carbon sources for enhanced biological phosphorus removal. Environ. Sci. Technol. 54, 2448–2458.
Roume, H., Heintz-Buschart, A., Muller, E.E.L., May, P., Satagopam, V.P., Laczny, C.C., Narayanasamy, S., Lebrun, L.A., Hoopmann, M.R., Schupp, J.M., 2015. Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. npj Biofilms Microbiomes 1, 1–11.
Sauer, U., Eikmanns, B.J., 2005. The PEP—pyruvate—oxaloacetate node as the switch point for carbon flux distribution in bacteria: we dedicate this paper to Rudolf K. Thauer, director of the max-Planck-institute for terrestrial microbiology in Marburg, Germany, on the occasion of his 65th birthday. FEMS Microbiol. Rev. 29, 765–794.
Scarborough, M.J., Myers, K.S., Donohue, T.J., Noguera, D.R., 2020. Medium-chain fatty acid synthesis by "Candidatus Weimeria bifida" gen. nov., sp. nov., and "Candidatus Pseudoramibacter fermentans" sp. nov. Appl. Environ. Microbiol. 86 e02242-19.
Sharma, S., Steuer, R., 2019. Modelling microbial communities using biochemical resource allocation analysis. J. R. Soc. Interface 16, 20190474.
Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., 2021. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nat. Commun. 12, 1–13.
Singleton, C.M., Petriglieri, F., Wasmund, K., Nierychlo, M., Kondrotaite, Z., Petersen, J. F., Peces, M., Dueholm, M.S., Wagner, M., Nielsen, P.H., 2022. The novel genus,'Candidatus Phosphoribacter', previously identified as Tetrasphaera, is the dominant polyphosphate accumulating lineage in EBPR wastewater treatment plants worldwide. ISME J. 16, 1605–1616.
Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G., Thrash, J.C., 2019. High proportions of bacteria and archaea across most biomes remain uncultured. ISME J. 13, 3126–3130.
Sweetlove, L.J., Beard, K.F.M., Nunes-Nesi, A., Fernie, A.R., George Ratcliffe, R., 2010. Not just a circle: flux modes in the plant TCA cycle. Trends Plant Sci. 15, 462–470.
Terzer, M., Stelling, J., 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24, 2229–2235.
van der Rest, M.E., Frank, C., Molenaar, D., 2000. Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of Escherichia coli. J. Bacteriol. 182, 6892–6899.
Van Loosdrecht, M.C.M., Smolders, G.J., Kuba, T., Heijnen, J.J., 1997. Metabolism of micro-organisms responsible for enhanced biological phosphorus removal from wastewater, Use of dynamic enrichment cultures. Antonie Van Leeuwenhoek 71, 109–116.
Vögeli, B., Engilberge, S., Girard, E., Riobé, F., Maury, O., Erb, T.J., Shima, S., Wagner, T., 2018. Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. Proc. Natl. Acad. Sci. 115, 3380–3385.
Wilmes, P., Andersson, A.F., Lefsrud, M.G., Wexler, M., Shah, M., Zhang, B., Hettich, R. L., Bond, P.L., VerBerkmoes, N.C., Banfield, J.F., 2008. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. ISME J. 2, 853–864.
Wimpenny, J.W.T., Firth, A., 1972. Levels of nicotinamide adenine dinucleotide and reduced nicotinamide adenine dinucleotide in facultative bacteria and the effect of oxygen. J. Bacteriol. 111, 24–32.
Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A.F., Hoelzle, R.D., Lamberton, T.O., McCalley, C.K., Hodgkins, S.B., 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560, 49–54.
Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337, 1661–1665.
Yagci, N., Artan, N., Çokgör, E.U., Randall, C.W., Orhon, D., 2003. Metabolic model for acetate uptake by a mixed culture of phosphate-and glycogen-accumulating organisms under anaerobic conditions. Biotechnol. Bioeng. 84, 359–373.
Zhang, Ye, Smallbone, L.A., DiCenzo, G.C., Morton, R., Finan, T.M., 2016. Loss of malic enzymes leads to metabolic imbalance and altered levels of trehalose and putrescine in the bacterium Sinorhizobium meliloti. BMC Microbiol. 16, 1–13.
Zhou, Y., Pijuan, M., Zeng, R.J., Yuan, Z., 2009. Involvement of the TCA cycle in the anaerobic metabolism of polyphosphate accumulating organisms (PAOs). Water Res. 43, 1330–1340.
Zimmermann, J., Kaleta, C., Waschina, S., 2021. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. Genome Biol. 22, 1–35.