The Pulse of Metabolism

Analysing "Candidatus Accumulibacter" dynamic flows

Páez Watson, T.W.

DOI
10.4233/uuid:38767bfb-23a1-4cd2-8f59-71929cea307b

Publication date
2025

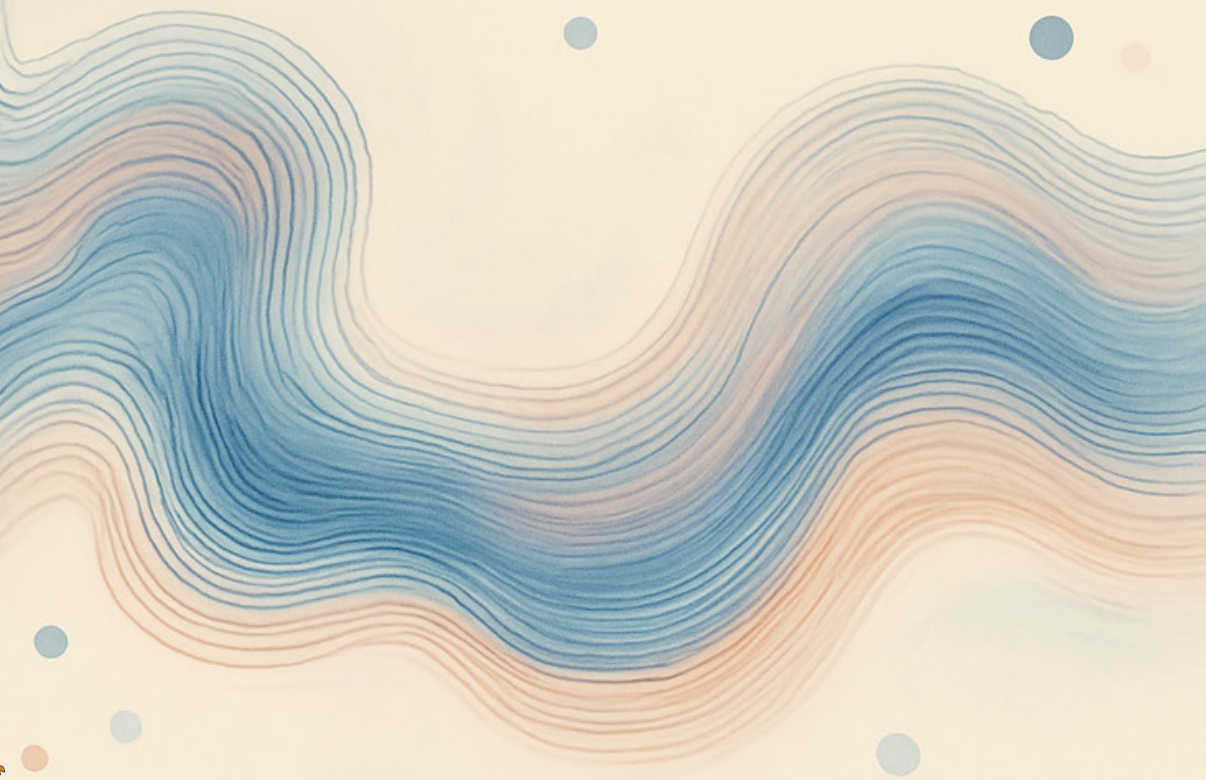
Document Version
Final published version

Citation (APA)
Páez Watson, T. W. (2025). *The Pulse of Metabolism: Analysing "Candidatus Accumulibacter" dynamic flows*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:38767bfb-23a1-4cd2-8f59-71929cea307b

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

The Pulse of Metabolism

Analysing "Candidatus Accumulibacter" dynamic flows



Timothy Páez Watson

The Pulse of Metabolism

Analysing "*Candidatus* Accumulibacter" dynamic flows

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates

to be defended publicly on
Friday 02 May 2025 at 10:00 o'clock

by

Timothy William PÁEZ WATSON

Master of Science in Life Science and Technology,
Delft University of Technology, the Netherlands
born in Quito, Ecuador

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. Mark C. M. van Loosdrecht	Delft University of Technology, promotor
Prof. dr. Aljoscha Wahl	Delft University of Technology, copromotor
Independent members:	
Prof.dr. D. Brdjanovic	Delft University of Technology
Prof.dr. F.J. Bruggeman	Vrije Universiteit Amsterdam
Prof. P.H. Nielsen	Aalborg University, Denmark
Prof.dr. R. Steuer	Leipzig University, Germany
Dr. K. Olavarria Gamez	Wageningen University & Research
Prof.dr. U. Hanefeld	Delft University of Technology, reserve

The research presented in this thesis was performed in the Environmental Biotechnology section of the Department of Biotechnology, Faculty of Applied Sciences of Delft University of Technology, The Netherlands. This PhD research was supported by the SIAM Gravitation Grant 024.002.00, the Netherlands Organization for Scientific Research (NWO).

Printed by: Proefschriftspecialist
Cover design by: Giovanni Páez Daqui
ISBN: 978-94-6384-765-0

An electronic version of this dissertation is available at: http://repository.tudelft.nl/

Printing of this thesis was financially supported by the Netherlands Society of Medical Microbiology (NVMM) and the Royal Netherlands Society for Microbiology (KNVM).

Una sonrisa no cuesta nada,

pero significa un mundo para quien lo necesita.

Table of Contents

Summary

Biology is full of complexities, and the more we learn, the more we realize how much remains unknown. A major debate in microbiology is whether DNA alone dictates an organism's function or if metabolism and energy flows play an equally fundamental role. This question is particularly relevant for microbes in dynamic environments, where survival depends on metabolic adaptability.

This thesis focuses on "*Candidatus* Accumulibacter", a key microorganism in wastewater treatment that removes excess phosphorus from water. These bacteria endure feast-famine cycles by storing and utilizing energy reserves as conditions change. While extensively studied, much remains unknown about their metabolic strategies and how environmental factors shape their function. This research combines computational models, laboratory cultivation, and multi-omics analysis to explore how "*Ca*. Accumulibacter" optimizes its metabolism.

Chapter 1 introduces the central debate: Is DNA the sole blueprint for microbial function, or do metabolism and energy constraints shape microbial behavior? It traces the shift from biochemical models to genome-centric approaches and highlights the potential of a metabolism-first perspective. It also contextualizes "*Ca*. Accumulibacter" within existing research, outlining its role in biological phosphorus removal and summarizing past findings.

Chapter 2 investigates extracellular polymeric substances (EPS) produced by "*Ca*. Accumulibacter", revealing novel glycans and glycoproteins that challenge genome-based predictions. These biomolecules are crucial for biofilm formation and microbial interactions, emphasizing the need for direct biochemical analysis alongside genetic data.

Chapter 3 uses elementary flux mode analysis (EFMA) to map the metabolic potential of "*Ca*. Accumulibacter". While genome annotations suggest flexibility, thermodynamic constraints limit feasible metabolic strategies, highlighting the role of energy availability in shaping microbial function.

Chapter 4 introduces the development of the Conditional Flux Balance Analysis (cFBA) Toolbox, an open-source Python framework for modeling metabolism in fluctuating environments. Unlike conventional models that assume steady-state conditions, cFBA enables dynamic predictions of resource allocation over time.

Chapter 5 explores the impact of temperature on "*Ca*. Accumulibacter" metabolism using cFBA. The findings confirm that biomass synthesis is mainly

aerobic but also uncover metabolic shifts at lower temperatures that influence phosphorus removal efficiency and microbial competition.

Chapter 6 examines how "*Ca*. Accumulibacter" metabolizes multiple substrates simultaneously, revealing unexpected synergies that enhance survival in microbial communities. Combining experimental enrichment cultures with cFBA, this study identifies key metabolic trade-offs and resource optimization strategies.

Finally, Chapter 7 synthesizes the thesis findings, advocating for a shift beyond genome-based interpretations toward a metabolism-centric understanding of microbial function. It discusses broader implications for microbial ecology, wastewater engineering, and metabolic modeling, emphasizing the need for multi-omics approaches and potential applications in synthetic biology.

By integrating experimental and computational approaches, this research deepens our understanding of how "*Ca*. Accumulibacter" thrives in fluctuating environments. More broadly, it highlights the importance of metabolism and energy flows in shaping microbial function, offering insights that extend beyond wastewater treatment to microbial ecology and engineered bioprocesses.

Samenvatting

Biologie is vol complexiteit, en hoe meer we leren, hoe meer we beseffen hoeveel er nog onbekend is. Een belangrijke discussie in de microbiologie is of DNA alleen de functie van een organisme bepaalt, of dat metabolisme en energiestromen een even fundamentele rol spelen. Deze vraag is vooral relevant voor microben in dynamische omgevingen, waar overleving afhankelijk is van metabole aanpassingsvermogen.

Deze thesis richt zich op "*Candidatus* Accumulibacter", een cruciale micro-organisme in afvalwaterzuivering dat overtollig fosfor uit water verwijdert. Deze bacteriën doorstaan cycli van overvloed en schaarste door energie op te slaan en te benutten wanneer de omstandigheden veranderen. Ondanks uitgebreide studie blijft veel onbekend over hun metabole strategieën en hoe omgevingsfactoren hun functioneren beïnvloeden. Dit onderzoek combineert computationele modellen, laboratoriumcultivatie en multi-omische analyses om te onderzoeken hoe "*Ca*. Accumulibacter" zijn metabolisme optimaliseert.

Hoofdstuk 1 introduceert het centrale debat: is DNA het enige bouwplan voor microbiële functies, of bepalen metabolisme en energiebeperkingen het gedrag van micro-organismen? Het hoofdstuk schetst de verschuiving van biochemische modellen naar genoomgerichte benaderingen en belicht het potentieel van een metabolismegerichte visie. Het ook plaatst "*Ca*. Accumulibacter" in de bestaande wetenschappelijke context, met een overzicht van zijn rol in biologische fosforverwijdering en een samenvatting van eerdere bevindingen.

Hoofdstuk 2 onderzoekt de extracellulaire polymere stoffen (EPS) die "*Ca*. Accumulibacter" produceert en onthult nieuwe glycaan- en glycoproteïnestructuren die genoomgebaseerde voorspellingen tegenspreken. Deze biomoleculen zijn essentieel voor biofilmvorming en microbiële interacties, wat het belang onderstreept van directe biochemische analyse naast genetische gegevens.

Hoofdstuk 3 maakt gebruik van elementary flux mode analysis (EFMA) om het metabole potentieel van "*Ca*. Accumulibacter" in kaart te brengen. Hoewel genoomannotaties metabole flexibiliteit suggereren, tonen thermodynamische beperkingen aan dat slechts een beperkt aantal strategieën haalbaar is. Dit benadrukt de invloed van energiebeperkingen op microbiële functies.

Hoofdstuk 4 introduceert de ontwikkeling van de Conditional Flux Balance Analysis (cFBA) Toolbox, een open-source Python-framework voor het modelleren van metabolisme in fluctuerende omgevingen. In tegenstelling tot

conventionele modellen die uitgaan van een steady-state, stelt cFBA onderzoekers in staat om dynamische voorspellingen te maken over de verdeling van hulpbronnen in de tijd.

Hoofdstuk 5 onderzoekt de invloed van temperatuur op het metabolisme van "*Ca*. Accumulibacter" met behulp van cFBA. De resultaten bevestigen dat biomassa-synthese voornamelijk in aerobe omstandigheden plaatsvindt, maar onthullen ook metabole verschuivingen bij lagere temperaturen. Deze veranderingen beïnvloeden de efficiëntie van fosforverwijdering en microbiële competitie.

Hoofdstuk 6 bestudeert hoe "*Ca*. Accumulibacter" meerdere substraten tegelijkertijd metaboliseert en onthult onverwachte synergiën die de overleving binnen microbiële gemeenschappen bevorderen. Door experimentele verrijkingsculturen te combineren met cFBA, identificeert dit onderzoek belangrijke metabole afwegingen en strategieën voor efficiënter gebruik van hulpbronnen.

Hoofdstuk 7 vat de bevindingen van deze thesis samen en pleit voor een paradigmaverschuiving van een genoomgerichte naar een metabolismegeoriënteerde benadering van microbiële functies. Het hoofdstuk bespreekt bredere implicaties voor microbiële ecologie, afvalwatertechnologie en metabole modellering en benadrukt de noodzaak van multi-omische methoden en mogelijke toepassingen binnen de synthetische biologie.

Door experimentele en computationele methoden te integreren, biedt dit onderzoek diepgaand inzicht in hoe "*Ca*. Accumulibacter"floreert in wisselende omgevingen. In bredere zin benadrukt het de centrale rol van metabolisme en energiestromen bij microbiële functies en biedt het inzichten die verder reiken dan afvalwaterzuivering, met implicaties voor microbiële ecologie en biotechnologische processen.

1

GENERAL INTRODUCTION

All the images in this chapter were illustrated by Julie Watson to make the content more accessible and easier to understand.

Biology is incredible. Yet it can seem frustrating and complex. Did you ever find yourself lost in memorizing a sea of molecule names, cell parts and mechanisms? You're not alone. Even with the best teachers in the world, understanding biological mechanisms is not easy. There is a reason for this – a reason so humbling that we often prefer to ignore it. Or actively avoid confronting it. It's true! Students don't fully understand biology. Neither do teachers. Not even renowned scientists. In fact, sometimes the more we learn, the more we realize how much we don't know. Why is that?

It's rooted in the way we seek to understand the world around us. We try to break down complex things, such as sight, intelligence, and love, into simpler components. In doing so, we search for small, simple units that, collectively, emerge as complex phenomena [1]. Scientists refer to these basic units as *fundamental* principles or elements. Once we identify these *fundamental principles,* we can understand the root cause for many complex behaviours in the world around us. Simple! Yet not. The problem is that what we call fundamental *today* is not the same as what we thought was fundamental *yesterday*, nor will it be the same *tomorrow* [2]. With each new technological advance, these 'fundamental' units are broken down into even simpler components, revealing newly discovered 'fundamental' ones. Physics and elemental particles is a field where this change is evident (Figure 1).

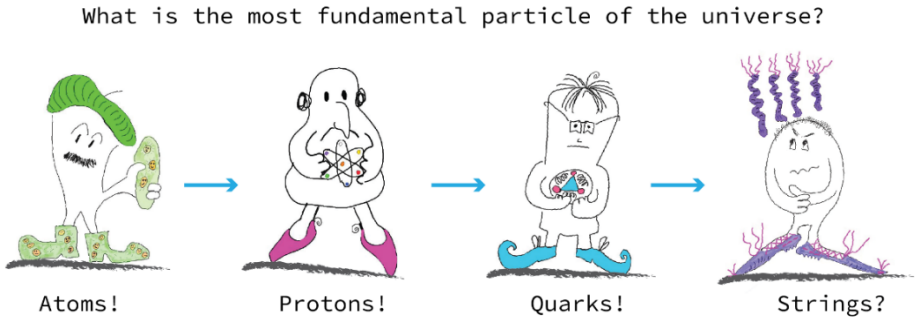


Figure 1*. Evolution over the centuries of our understanding of what constitutes the fundamental particles in the universe.*

The atom was once thought to be the fundamental unit of matter (the word itself means 'indivisible' in Greek). In the early 20th century, we discovered that atoms are actually composed of electrons, neutrons and protons –the fundamental particles at that time[3]. With the advent of more advanced technology, scientists challenged this notion by discovering that protons and neutrons are themselves composed of even smaller constituents called quarks –the current fundamental particles *today*[4]. If our technology becomes advanced enough, the fundamental

constituents might one day be vibrating strings [5]. And since technology advances at a near-exponential rate, the frequency with which new fundamental principles appear is also much faster!

Coming back to my first sentence: biology is incredible. Yet, just as in physics, it is complicated by what we consider to be fundamental. The current mainstream notion is that DNA, a molecule capable of carrying information, is the root cause of life and complex phenomena. This perspective shapes how we approach biology education, how journalists and the public discuss, and how scientists conduct research. In the following pages, I will contrast this view with a different perspective on what is fundamental to biology. I'll aim to show that challenging this notion can help us better discover and understand biological systems. Then I'll apply it to explore a unique microbe that indiscriminately stores fats, sugars and stones. But we'll get there. Eventually.

Fundamental Changes in Understanding Biology

To start, let's go back to the beginning: the origin of life*. I'm writing this in 2024. It should be pretty clear how life emerged by now, right?

Life emerged in the early stages of our planet. Inside a warm, shallow body of water, neutral gas mixtures sparked by lightning led to the formation of organic molecules –amino acids, sugars and even nucleotides [6,7]. Eventually, this *primordial soup* gave rise to more complex molecules (Figure 2: Primordial soup). RNA, among them, with catalytic power, became the first catalyst to enable self-replication. Over time, RNA's function was replaced by the more stable molecules: proteins and DNA, which gave rise to life as we know it. And thus, life emerged.

Or did it?

Different theories challenge this version of the origin of life. One of the most likely scenarios suggests that life did not emerge in a *primordial soup*, but in a completely different environment: deep inside the ocean, specifically at hydrothermal vents [8]. These vents –cracks in the ocean bed from where heat emanates –provide an ideal setting for the chemical formation of organic molecules containing 2 or more carbon atoms. In this scenario, self-assembling protocells harvesting the energy from hydrothermal vents gave rise to the most basic chemistry that is central to life. A chemical, self-replicating, cycle, the tri-

* Such a cliché! I know, but bear with me.

carboxylic acid cycle (TCA*), capable of synthesis of organic molecules would then lead to increasing complexity until it reached what we now consider life (Figure 2: Hydrothermal vents).

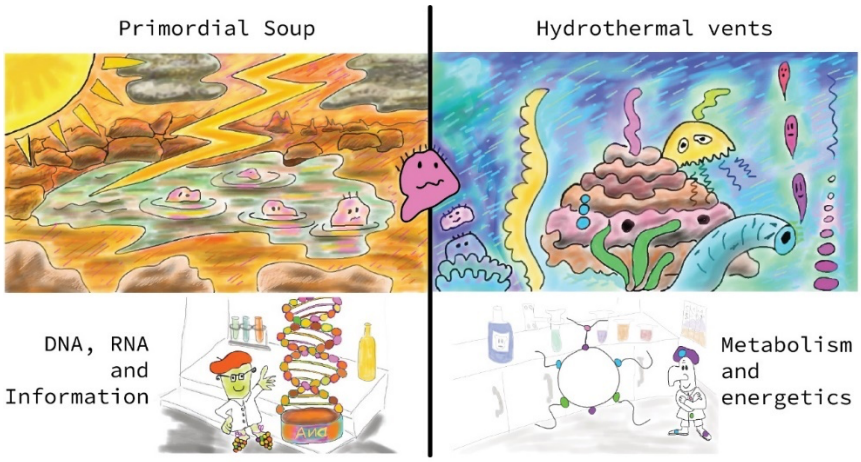


***Figure 2**. Fundamentally different theories on the origin of life. On the left, the theory of the primordial soup, in which life emerged in the form of complex biomolecules that formed with the help of UV radiation and lightning. Scientists which defend this theory hold DNA, RNA and information as fundamental to life. On the right, the theory of life emerging in hydrothermal vents by the self-assembling metabolic cycles powered by continuous energy emanating from the earth's crust. Scientists defending this theory hold metabolism and energy flows as fundamental to life.*

Does it matter which theory is correct? For now, it's less about determining the 'right' or the 'wrong'. It's about what each theory considers *fundamental* and how that affects our approach to research. The former theory places RNA and DNA – molecules that intrinsically carry information– at the beginning and centre of life. It suggests that information, in the form of molecules, came first and somehow invented metabolism. The latter theory places metabolism and the energy sustaining it at the centre of life, considering them to be *fundamental* [9].

In environmental microbiology, my current field of research, there is a heavy emphasis on collecting terabytes (tera = a trillion bytes) of data from DNA and RNA [10], with less focus on metabolic or energetic considerations. It is common to encounter research that concludes: "Microbe containing $Gene_X$, performs Function X" or "Microbe expressing high amounts of RNA_Y, has a significant

* TCA, one of the most used abbreviations in this thesis. Can you guess how many times I use 'TCA' in this whole thesis? Note that this footnote =+3 to the TCA-count.

enhanced Activity Y". These conclusions might be correct, but as always, the devil is in the details. They *might be true*; it really depends on the context in which the microbe exists. This possibility of doubt is not always stated. For me, that is troubling.

Studying molecules of information (DNA and RNA) is sometimes the most suitable method (if not the only one) to categorize and try to understand most microbes in this world today [11]. The analysis of such molecules is much simpler than that of proteins and metabolites. However, microbiologists need to contextualize this information from an energetic and metabolic-centric perspective. What really influences an organism to consume a given substrate, to move in a given way, to divide or to accumulate excessive amounts of minerals? DNA-based analysis may pave the way to answer these questions, but it is insufficient to travel through these paths. Conversely, we need an energetic/metabolic mindset to travel these paths.

Let us explore further this dichotomy with an example.

Cell size and simplicity: is it DNA or energetics based?

As a student, you've probably looked at cells under a microscope. Eukaryotes are cute and easy to spot. But bacteria? It takes a lot of training and patience to convincingly spot bacteria with a normal microscope.

It is not an exaggeration to say that bacteria are tiny –microscopically small (to state the obvious). They are at least 10 times smaller than eukaryotic cells. There must be a reason for this, and indeed, there is. But it's not hidden in DNA. On the contrary, the reason is rooted in energetic limitations that are independent of DNA, yet largely shape DNA content [12,13].

Cells require energy, which is primarily generated at the level of the membranes [14]. There is a difference in proton (H^+) concentration (*a.k.a.* pH) between the inside and the outside of a cell. This difference forces H^+ to cross through the membranes with such force that it generates energy. The more membrane space a cell has, the more energy it can generate. So, why then aren't bacteria much bigger? Larger cells will have more membrane space (surface area), which means greater capacity for energy generation. However, the cell itself (the volume) is also larger and will demand more energy. And the growth of both surface and volume isn't equal.

Consider a cell in the shape of a cube. The surface of this cube (*i.e.* the membranes) is calculated by multiplying the area of each face ($x * x = x^2$) by the number of faces (6). The volume, on the other hand, is calculated by multiplying the length, width and height ($x * x * x = x^3$). Now, if this cell doubles in size, the surface only grows at a quadratic rate (which is to say, the exponent of x^2), while the volume grows at a cubic rate (exponent of x^3). If you keep doubling the size, the difference between surface and volume gets wider, with the volume growing faster than the surface. Larger cells are simply not feasible because the amount of energy that a unit of membrane needs to generate becomes unrealistically high.

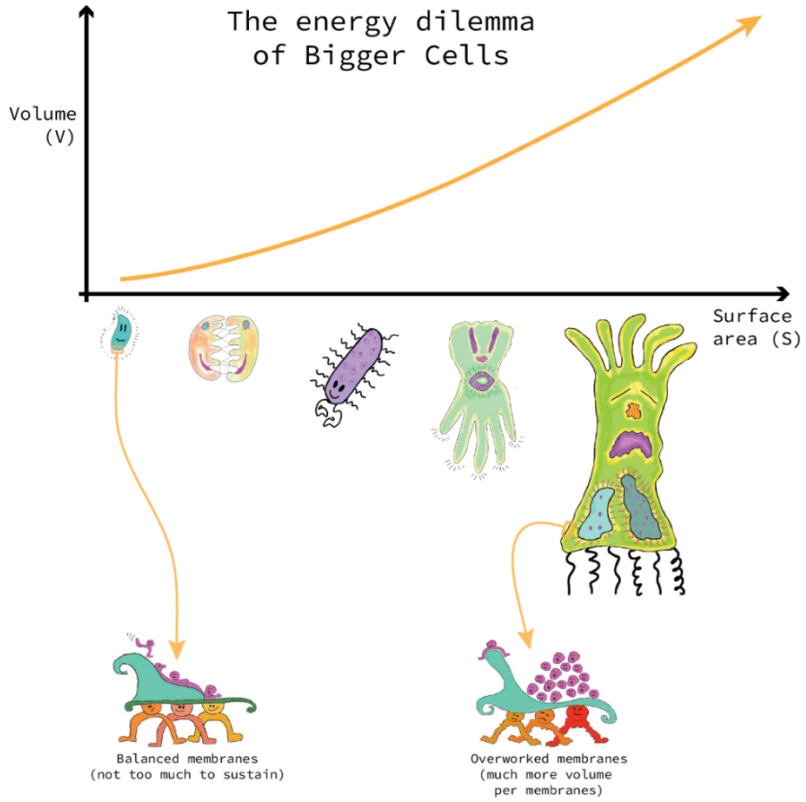


***Figure 3**. The energy dilemma of big cells. The volume (V) of a cell grows faster than its surface (S) as shown in the figure. This means that bigger cells have less membranes available per volume. Because membranes generate energy, bigger cells must either generate more energy per membranes, reduce their energy demands or just remain small.*

Because of this physical, geometric limit, bacteria on average are no larger than 10 µm in diameter. Simple! No DNA involved –at least at the very fundamental aspects. Actually, having, maintaining, and using DNA costs energy too. The very same reason that makes bacteria small also limits the amount of DNA a cell can

sustain. Thus, bacterial genomes are roughly 5 Mbp in size, harbouring no more than 3000 to 4000 genes [15]. For comparison, a human cell contains up to 3000 Mbp in their genome, harbouring around 25 000 genes [16]. That's nearly 500 times bigger! How? Because we (eukaryotes) harbour *energy generating slaves* within our cells called mitochondria, so we don't depend on our surface for energy generation. But that's a complicated story we won't deal with here.

In the following section, I will dig deeper into a type of environment where energy flows are also crucial in shaping bacterial function. This is where the environment of the unique microbe I mentioned at the very beginning. The one that indiscriminately stores fats, sugars and stones.

Intermittent fasting: how bacteria adapt to sudden periods of starvation

Almost every environment on this planet is dynamic. We experience monthly changes in weather as the Earth orbits the Sun and daily shifts in light as the planet rotates. We sleep, wake up, eat, get hungry, and exercise. We are never truly static. The same applies to microorganisms, except perhaps in the controlled confines of high-tech labs [17]. Take oceans as an example. Sunlight stimulates phototrophic growth near the surface, temporarily increasing oxygen levels. At night, this process reverses, and the detritus from daytime growth descends, nourishing deeper layers of the ocean. Over millions of years, life has evolved countless strategies to not just survive, but to exploit these recurring cycles [18].

In wastewater treatment plants, engineers mimic these cyclic environments, forcing bacteria to adopt a remarkable survival strategy. First, bacteria enter a tank with abundant food but with no oxygen (*i.e.* anaerobic feast). After a few hours, oxygen is injected into these tanks, though food is no longer abundant (*i.e.* aerobic famine). Bacteria are then cycled through this process repeatedly. This anaerobic feast, aerobic famine cycling is ingenious, as it forces bacteria to employ a survival strategy that cleans our water by removing organic carbon and phosphates –two of the main contaminants [19]. This survival strategy is truly remarkable and is worth exploring in detail.

Previously I mentioned that bacteria generate energy by means of a pH difference across their cell membranes. This pH difference doesn't appear out of thin air; it's established through a process that, paradoxically, requires energy—specifically, a type of energy that cannot be directly used for survival: combustion. Yes, like fire. Exactly the same.

During a fire, oxygen strips electrons from compounds in a process known as oxidation. You may have painfully felt this when blowing air too close to a fire[*]. A similar process occurs in all living cells, but instead of indiscriminately generating heat, cells use a controlled machinery to sequentially strip electrons from food and pass them to oxygen at the membrane. The energy released is carefully channelled into maintaining the pH difference between the inside and outside of the cell. This pH difference, in turn, powers energy generation that cells can use [14].

Now, let's return to our cyclic environment—this anaerobic-feast, aerobic-famine regime. A keen observer might spot a dilemma: if bacteria are separated from oxygen while consuming food, how do they continue generating energy? Well, they have a few options.

1. **Resist the urge to eat:** Some bacteria patiently wait until oxygen is reintroduced before consuming their food as normal. This strategy, often used by heterotrophic bacteria, is the most energy-efficient in terms of yield. However, it is vulnerable to competition from bacteria that sacrifice this yield to consume food anaerobically, leaving very little behind.
2. **Consume food anaerobically**: In the absence of oxygen, bacteria can use alternative molecules to accept electrons, generating energy. However, these molecules are not called *oxygen.* These substitutes are less efficient than oxygen, resulting in much slower growth (a painfully slow process for lab researchers).
3. **The best of both worlds:** The most effective strategy is a compromise. Bacteria consume as much food as possible during the anaerobic feast but refrain from stripping away its electrons. Instead, they store the food as fats, patiently waiting for oxygen to arrive. When oxygen becomes available, they use it to burn the stored fats, efficiently generating energy. Voilà—an elegant balance between speed and efficiency.

The latter strategy is the best fit for these cyclic environments. Many bacteria can employ this trick, but one type in particular stands out: the Polyphosphate Accumulating Organisms (PAOs[†]). As its name suggests, they are expert accumulators, not just of polyphosphate but of other polymers too. These organisms accumulates fats, sugars and polyphosphates – polymers that look like

[*] This happens a lot during a barbeque. At least for those that enjoy getting their hands (or more likely, their faces) dirty.

[†] PAOs. Surely the most used abbreviation of this thesis. By how much more than TCA? Place your bets!

huge stones! The following pages will explore how this unique group benefits from overaccumulation and how this affects their physiology.

PAOs: master of fats, sugars and stones

PAOs thrive in the cyclic environments imposed during wastewater treatment plants, specifically during Enhanced Biological Phosphorous Removal (EBPR) [19]. This process consists of repetitive anaerobic-feast, aerobic-famine regimes that allow PAOs to accumulate excessive amounts of polyphosphate inside their cells and, this way, to clean our water (see Figure 4 for a schematic of a typical cycle of PAOs). They are crucial to EBPR and decades of research have been devoted to cultivating them, disturbing them, breaking them apart to measure their pieces, and much more. We understand a great deal of these organisms, yet we still don't grasp what makes PAOs accumulate so much inside their cells.

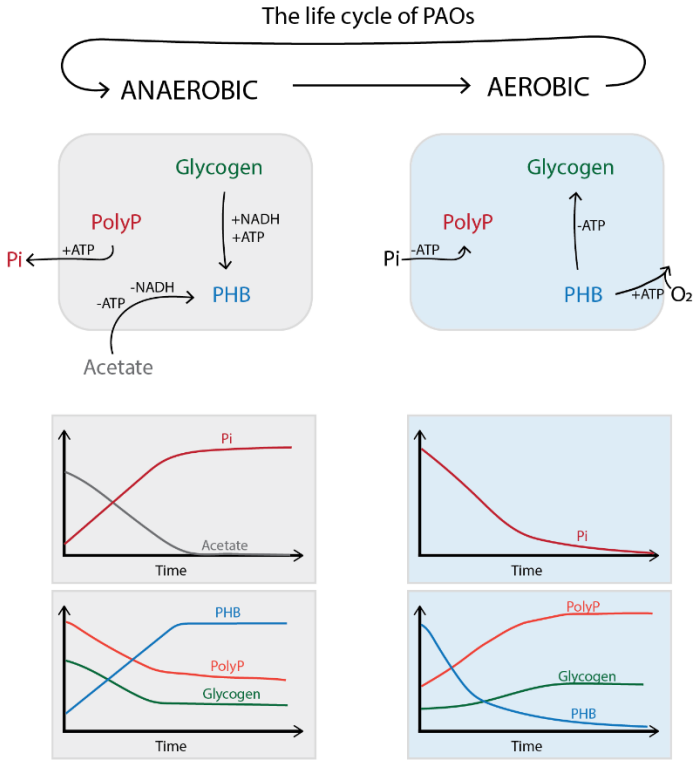


***Figure 4**. A typical cycle of PAOs in EBPR systems. In this schematic, the dynamic changes of extracellular and intracellular components that a typical PAO cell experiences during an EBPR cycle are represented.*

To help understand PAOs, we will look deeper into the history since their discovery. I make the argument that this history should be seen as marked by *epochs*. Each characterized by their discoverers: pragmatic engineers, microbiologists and process engineers, and finally molecular biologists. After this Chapter, you can find an image and a table representing a timeline of the most important discoveries made in the research of PAOs since their discovery (See the next section).

Pragmatic engineers (1959 – 1974)

The first report of excess phosphate being removed by *sludge* from wastewater was in 1959 as part of a problem: rice plants suspended on the effluent water grew many big leaves but very little rice grains because the effluent contained excess nitrogen but weirdly not excess phosphorus [20]. The decade that followed was the battleground of minds trying to prove or disprove the biological root of this removal [21,22], the physical conditions required to enhance the removal [23] and, finally in 1974, the establishment of a technology that can effectively remove phosphorus by means of bacterial polyphosphate accumulation [24].

Microbiologists and process engineers (1974 – 1998)

Towards the end of the 1970s it became apparent that engineers required a deeper understanding of the biology underlying phosphorus removal to enable system optimization. Thus, process engineers worked together with microbiologists to understand the mechanisms resulting in phosphorus removal. By 1987 a descriptive biochemical model for the metabolism of PAOs had been established [25,26]. Not only was the system qualitatively described, but great steps towards quantitative models to explain the biochemical changes in the anaerobic or in the aerobic period of EBPR were also made [27,28]. These quantitative models are still used today when describing the biochemistry of PAOs enrichments.

Mark van Loosdrecht reflected on this period in 1997:

> *"It might be considered remarkable that the biochemical model was developed by engineers, but all had personal contacts with or knowledge of the microbiological research field. Possibly engineers were less hampered by a traditional biochemical and microbial approach and could therefore more easily come up with new concepts in microbial ecophysiology […]"* [19]

At the end of the 90s, the process was well described and implemented effectively worldwide. Still, some biochemical details were still unclear such as the source of energy and reducing potential (ATP and NADH respectively) for the accumulation of fats [25], the varying stoichiometry observed across different treatment plants

and labs [27], the actual biochemical operation for glycogen degradation [29,30], amongst others.

The complexity of metabolic networks, which can have a large number of possible solutions, together with limited computational tools of the time was partly responsible for hampering this progress. But in the late 1990s, a revolution was underway—the sequencing of the human genome—that would transform biology and PAO research till the present day.

Molecular biologists (1999 – present day)

At the end of the 90s and beginning of the 2000s, researchers started implementing molecular techniques to characterize the bacterial community behind the umbrella term PAOs. For example, "*Candidatus* Accumulibacter phosphatis" was proposed as the main responsible PAO in 1999 [31], and several studies followed the exploration of this novel species [32,33].

In 2006, Martín, et al. [34] publish the first metagenomic study of "*Candidatus* Accumulibacter phosphatis", paving the ground for an extensive field of metagenomics in wastewater research. In their research they noted that genes for Entner-Doudoroff (ED) glycolysis were not present in this species; that the operation of a split TCA cycle could explain different stoichiometries observed over the years; and that genes related to carbon fixation might indicate potential autotrophic behaviour. These findings redefined how researchers understood PAO physiology and paved the way for an era dominated by metagenomics and molecular techniques.

Since then, the field has focused heavily on characterizing PAO genetics. For instance, researchers identified that two main variants of the gene *ppk1* diverged early in the evolution of "*Ca.* Accumulibacter" and could be used to distinguish between groups called clades [32,35]. However, such a separation showed to be confusing in the following decades since many functions showed to be randomly distributed amongst clades[36-38]. To solve this issue, it became common practice to publish the metagenome of the specific species associated to functional observations [39,40]. Finally in 2022, Petriglieri, et al. [41] performed a systematic re-evaluation of the phylogeny of "*Ca.* Accumulibacter" and today over 20 species of "*Ca.* Accumulibacter" are recognized*.

* Quite crazy to keep up with. For example, if we name some of the species (alphabetically), we have: *aalborogensis, adiacens, adjuntus, affinus, appositus, cognatus, conexus, contigus, delftensis,…* amongst many others. We didn't even reach *phosphatis* yet!

Regarding the current state-of-art in PAOs research, one cannot overlook the importance of the publication of the first metagenome of "*Ca.* Accumulibacter phosphatis"[34]. It was truly transformative for the field of PAOs research. This is the branching point at which most research moved to focus almost exclusively on molecular techniques and seemed to have forgotten biochemistry and systems biology*. This not only happened in PAOs, but in almost every research field in the world. This was eloquently captured by Sydney Brenner in 1995 in a letter to *Current Biology*, where he addressed the (then) challenges of understanding biology when too much emphasis is placed on genetics:

> *"The only way out is through biochemistry of one kind or another. In 1990, I made the remark that biochemistry and communism seemed to have disappeared in that year. Most people thought I said this with glee, but in fact it was with regret, at least for biochemistry.*
>
> *[…] what we now need to do is integrative biology; that we are very good at working out how simple systems with few components work but very bad at putting the parts of multicomponent systems together."* [42]

It's sobering to realize that almost 30 years later, this statement remains just as relevant. In the case of PAOs, the need for integrative biology approaches is more critical than ever to unravel the complexities of these marvellous organisms. To get the knowledge we have gathered from our genetics approach and apply it with a metabolic and energetics mindset. To use systems biology approaches to answer some of the questions that have been buried under mountains of data.

Aims and scope of this thesis

In this thesis, I aimed at deepening our understanding of how bacteria thrive in feast-famine cycles with a mindset of integrative biology. Specifically, using "*Ca.* Accumulibacter" as a model organism of PAOs, I aimed to answer some of the knowledge gaps that limit our ability to obtain a systems-biology level understanding of microbial communities living in EBPR systems. Specifically, I attempted to shed light on the following knowledge gaps that exist in the field:

* Don't let me convince you. Have a look at the timeline plot of PAOs history in the next section. There you might get an idea of how strong the focus on molecular biology has been in the last 20 years.

- The extent to which *genetic information and bioinformatics tools can predictive the physiology of PAOs.* With the advent of rich genomic data, attempts at predicting complex physiological structures and behaviours are being made. In Chapter 2 we studied extracellular polymeric substances (EPS) of PAOs to underscore how limited and inconclusive the use of genomic data alone can be.

- Genome-based metabolic models are used to perform simulations in which often *direction and reversibility of reactions is pre-defined*. Little effort has been made to apply systematic analysis to these models and to limit their solutions to given environmental conditions. Chapter 3 explores a metabolic model based on genomic data of *"Ca.* Accumulibacter" and highlights the diversity and complexity of network solutions. Further it explores the use of thermodynamic calculations in identifying feasible solutions within a given environmental context.

- Modelling frameworks that allow for the prediction of complex metabolic strategies in cyclic environments with resource allocation *are complex, lack clear documentation and are not available on open-access software*. This limits the application of systems biology on organisms living in dynamic cycles such as PAOs. Chapter 4 describes the development of an open-source Python toolbox to research optimal metabolic strategies under cyclic conditions.

- Resource allocation theory has been applied to many microorganisms to enable better understanding of their metabolism. However, little effort has been done in slow growing organisms such as PAOs. Furthermore, the consequences of a historical assumption that biosynthesis is restricted to aerobic phases has not been tested. In Chapter 5 we applied the modelling toolbox from Chapter 4 to explore the consequences of temporal separation of biosynthesis in "*Ca.* Accumulibacter" and explore the metabolic consequences of this separation. This was possible thanks to the implementation of temperature as an external variable affecting fluxes, which inevitably leads to tighter control on resource use at lower temperatures.

- The metabolic mechanisms of "Ca. Accumulibacter" under mixed substrate conditions remain poorly understood, with most studies focusing on single substrates. While some reports have identified potential synergies in co-substrate uptake, the specific metabolic interactions and their implications

for resource use and EBPR optimization have not been fully elucidated. Chapter 6 identifies hidden synergetic interactions between metabolic model operations by means of combining experimental tests on highly enriched PAOs cultures with metabolic modelling predictions.

Finally I integrate all these findings in Chapter 7 and give my personal outlook into the research field.

References

1 Kurzweil, R. *The Singularity Is Nearer: When We Merge with AI*. (Random House, 2024).

2 Walker, S. I. Life as no one knows it: The physics of life's emergence., (Hachette UK, 2024).

3 Rutherford, E. LXXIX. The scattering of α and β particles by matter and the structure of the atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **21**, 669-688 (1911).

4 Campagnari, C. & Franklin, M. The discovery of the top quark. *Reviews of Modern Physics* **69**, 137 (1997).

5 Scherk, J. & Schwarz, J. H. Dual models and the geometry of space-time. *Physics Letters B* **52**, 347-350 (1974).

6 Cleaves, H. J., Chalmers, J. H., Lazcano, A., Miller, S. L. & Bada, J. L. A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. *Origins of Life and Evolution of Biospheres* **38**, 105-115 (2008).

7 Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528-529 (1953).

8 Martin, W. & Russell, M. J. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **358**, 59-85 (2003).

9 Lane, N. Transformer: the deep chemistry of life and death. (Profile books, 2022).

10 Singleton, C. M. *et al.* Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nature communications* **12**, 1-13 (2021).

11 Huse, S. M. *et al.* Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS genetics* **4**, e1000255 (2008).

12 Bonner, J. T. Why size matters: from bacteria to blue whales. (2024).

13 Lane, N. Power, sex, suicide: mitochondria and the meaning of life. (Oxford University Press, 2018).

14 Mitchell, P. & Moyle, J. Chemiosmotic hypothesis of oxidative phosphorylation. *Nature* **213**, 137-139 (1967).

15 Trevors, J. Genome size in bacteria. *Antonie van Leeuwenhoek* **69**, 293-303 (1996).

16 Venter, J. C. *et al.* The sequence of the human genome. *science* **291**, 1304-1351 (2001).

17 Novick, A. & Szilard, L. Description of the chemostat. *Science* **112**, 715-716 (1950).

18 Mougi, A. Diversity of biological rhythm and food web

stability. *Biology Letters* **17**, 20200673 (2021).

19 Van Loosdrecht, M., Smolders, G., Kuba, T. & Heijnen, J. Metabolism of micro-organisms responsible for enhanced biological phosphorus removal from wastewater, Use of dynamic enrichment cultures. *Antonie van Leeuwenhoek* **71**, 109-116 (1997).

20 Srinath, E., Sastry, C. & Pillai, S. Rapid removal of phosphorus from sewage by activated sludge. *Experientia* **15**, 339-340 (1959).

21 Levin, G. V. & Shapiro, J. Metabolic uptake of phosphorus by wastewater organisms. *Journal (Water Pollution Control Federation)*, 800-821 (1965).

22 Menar, A. B. & Jenkins, D. Fate of phosphorus in waste treatment processes: enhanced removal of phosphate by activated sludge. *Environmental Science & Technology* **4**, 1115-1121 (1970).

23 Nichols, H. Full scale experimentation on the new Johannesburg extended aeration plants. *Water SA* **1**, 121-132 (1975).

24 Barnard, J. L. Cut P and N without chemicals. (1974).

25 Comeau, Y., Hall, K. J., Hancock, R. E. & Oldham, W. K. Biochemical model for enhanced biological phosphorus removal. *Water Research* **20**, 1511-1521 (1986).

26 Mino, T., Arun, V., Tsuzuki, Y. & Matsuo, T. in *Biological phosphate removal from wastewaters* 27-38 (Elsevier, 1987).

27 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnology and bioengineering* **43**, 461-470 (1994).

28 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. Stoichiometric model of the aerobic metabolism of the biological phosphorus removal process. *Biotechnology and bioengineering* **44**, 837-848 (1994).

29 Maurer, M., Gujer, W., Hany, R. & Bachmann, S. Intracellular carbon flow in phosphorus accumulating organisms from activated sludge systems. *Water Research* **31**, 907-917 (1997).

30 Hesselmann, R., Von Rummell, R., Resnick, S. M., Hany, R. & Zehnder, A. Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Research* **34**, 3487-3494 (2000).

31 Hesselmann, R. P., Werlen, C., Hahn, D., van der Meer, J. R. & Zehnder, A. J. Enrichment, phylogenetic analysis and detection of a bacterium that performs enhanced biological phosphate removal in activated sludge. *Systematic and applied microbiology* **22**, 454-465 (1999).

32 McMahon, K. D., Dojka, M. A., Pace, N. R., Jenkins, D. & Keasling, J. D. Polyphosphate kinase from activated sludge performing enhanced biological phosphorus removal. *Applied and environmental microbiology* **68**, 4971-4978 (2002).

33 Kong, Y., Nielsen, J. L. & Nielsen, P. H. Microautoradiographic

study of Rhodocyclus-related polyphosphate-accumulating bacteria in full-scale enhanced biological phosphorus removal plants. *Applied and environmental microbiology* **70**, 5383-5390 (2004).

34 Martín, H. G. *et al.* Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature biotechnology* **24**, 1263-1269 (2006).

35 He, S., Gall, D. L. & McMahon, K. D. "Candidatus Accumulibacter" population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Applied and environmental microbiology* **73**, 5865-5874 (2007).

36 Carvalho, G., Lemos, P. C., Oehmen, A. & Reis, M. A. Denitrifying phosphorus removal: linking the process performance with the microbial community structure. *Water research* **41**, 4383-4396 (2007).

37 Flowers, J. J., He, S., Yilmaz, S., Noguera, D. R. & McMahon, K. D. Denitrification capabilities of two biological phosphorus removal sludges dominated by different 'Candidatus Accumulibacter'clades. *Environmental microbiology reports* **1**, 583-588 (2009).

38 Zeng, W., Zhang, L., Fan, P., Guo, J. & Peng, Y. Community structures and population dynamics of "Candidatus Accumulibacter" in activated sludges of wastewater treatment plants using ppk1 as phylogenetic marker. *Journal of Environmental Sciences* **67**, 237-248 (2018).

39 Rubio-Rincón, F. *et al.* "Candidatus Accumulibacter delftensis": A clade IC novel polyphosphate-accumulating organism without denitrifying activity on nitrate. *Water research* **161**, 136-151 (2019).

40 Qiu, G. *et al.* Metabolic traits of Candidatus Accumulibacter clade IIF strain SCELSE-1 using amino acids as carbon sources for enhanced biological phosphorus removal. *Environmental Science & Technology* **54**, 2448-2458 (2019).

41 Petriglieri, F. *et al.* Reevaluation of the Phylogenetic Diversity and Global Distribution of the Genus "Candidatus Accumulibacter". *mSystems* **7**, e00016-00022 (2022).

42 Brenner, S. Loose end. *Current Biology* **5**, 94 (1995).

A Brief History of PAO Research

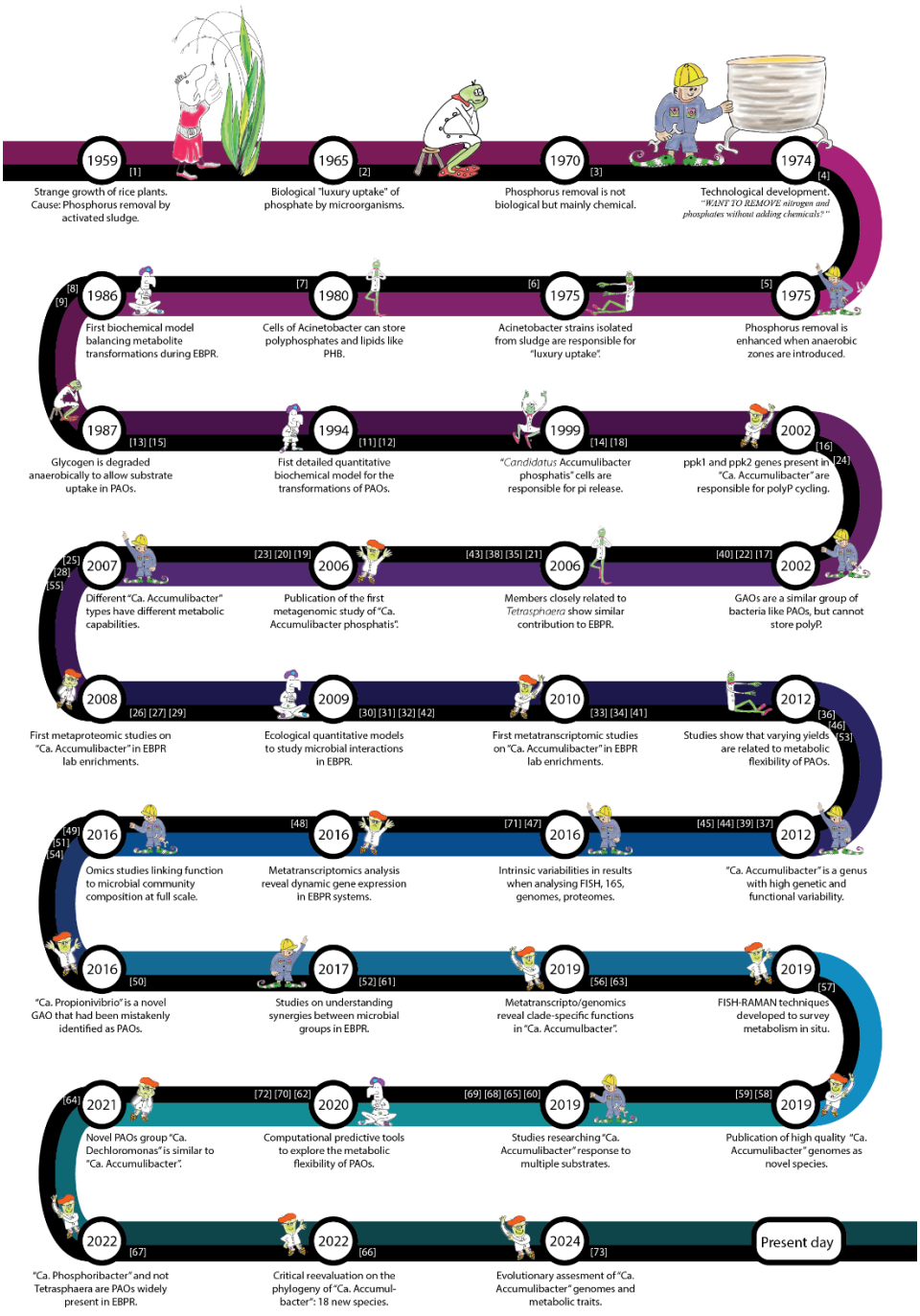


***Figure 1**. Timeline of significant discoveries in PAO-related research. Each node represents an initial observation, with subsequent references indicating ongoing research contributions over the years.*

Animated icons denote the primary focus of the research during each period: engineers, microbiologists, molecular biologists, and systems biologists. Corresponding references are detailed in Table 1 at the end of the introduction section.

The historic timeline depicted in Figure 1 shows the evolution over the years in the research of PAOs, starting with the very first description of phosphate removed by activated sludge. To put emphasis on the research type being performed, and especially on the popularity of research approaches over the years, an illustration was placed next to each node representing a scientist type (Figure 2 describes these types).

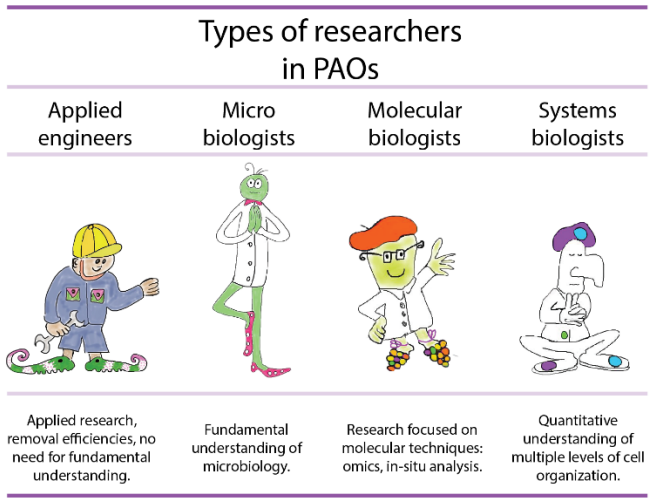


***Figure 2**. Type of researcher and approach in the study of PAOs. Each type is matched with the discoveries in the historical timeline from Figure 1.*

And finally, a not-so-brief summary of all the references from the timeline (Table 1). Hopefully, this serves as a handy PAO starter guide for curious young scientists! It's easy to get lost and overwhelmed in the gigantic world of publications.

***Table 1**. Summary of all the greatest discoveries in the history of PAO research. The number of each reference aligns with the reference in the Figure 1.*

Year	Discovery
1959	Srinath, et al. [1] observed that rice plants grown in activated sludge tanks exhibited excessive vegetative growth but poor grain formation, indicating phosphorus removal in the tanks.
1965	Levin and Shapiro [2] identified that phosphate uptake in treatment plants is biological, describing it as "luxury" uptake not linked to growth, and proposed designs for phosphate removal.

1970	Menar and Jenkins [3] concluded that phosphate removal is a chemical process rather than a biological one.
1974	Barnard [4] provided the first engineering-focused description of the enhanced biological phosphorus removal (EBPR) process, discussing mass balances and operational conditions.
1975	Nichols [5] tested Barnard's designs and found that anaerobic zones near the feed of plug flow reactors enhance phosphorus removal.
1975	Fuhs and Chen [6] Isolated *Acinetobacter* species and demonstrated their ability to accumulate polyphosphate, though the exact biochemical mechanisms were unclear.
1980	Deinema, et al. [7] found that EBPR sludge, dominated by *Acinetobacter*, accumulates polyphosphates and stores polyhydroxyalkanoates (PHAs), mainly polyhydroxybutyrate (PHB).
1986	Comeau, et al. [8] described the stoichiometry and kinetics of bacteria involved in phosphate removal, suggesting roles for polyphosphate and PHB in energy storage and substrate utilization.
1986	Wentzel, et al. [9] proposed a metabolic model for *Acinetobacter*, believed to be the main PAOs, highlighting the glyoxylate cycle's role in supplying intermediates to the TCA cycle.
1987	Mino, et al. [10] experimentally tested and validated that anaerobic glycogen degradation is involved in PAOs transformations.
1994	Smolders, et al. [11] and Smolders, et al. [12] developed the first quantitative metabolic models explaining the metabolism of PAOs during EBPR.
1997	Maurer, et al. [13] used labelled experiments to show that glycogen is potentially degraded via the Entner-Duodoroff (ED) pathway.
1999	Hesselmann, et al. [14] identified "*Candidatus* Accumulibacter phosphatis" as the main bacterium responsible for phosphorus removal in EBPR.
2000	Hesselmann, et al. [15] provided additional evidence that glycogen is degraded via the ED pathway with labelled substrate.
2002	McMahon, et al. [16] cloned and sequenced polyphosphate metabolism genes (ppk1 and ppk2) from "*Ca*. Accumulibacter," confirming their role in polyP synthesis
2002	Crocetti, et al. [17] discovered GAOs as competitors to PAOs, with similar metabolic behaviors but negative effects on phosphorus removal.
2004	Kong, et al. [18] used MAR-FISH microscopy to link phosphate and acetate uptake activities to "Ca. Accumulibacter."
2004	Venter, et al. [19] showed that metagenomes can be obtained from environmental samples.

2004	Tyson, et al. [20] provided quantitative insights into microbial communities based on metagenomic data.
2005	Kong, et al. [21] revealed that gram-positive Actinobacteria, related to Tetrasphaera, also participate in phosphorus cycling in EBPR.
2006	Oehmen, et al. [22] further test GAOs in EBPR systems and describe detailed metabolic operations with propionate.
2006	Martín, et al. [23] published the first metagenomic study on PAOs, resolving controversies in EBPR models and proposing new metabolic pathways.
2007	He, et al. [24] used the ppk1 gene as a marker to separate "*Ca*. Accumulibacter" into distinct clades
2007	Carvalho, et al. [25] provided evidence of different "*Ca.* Accumulibacter" morphotypes (rods and cocci) that might indicate varying metabolic capabilities, including denitrification.
2008	Wilmes, et al. [26] conducted the first meta-proteomics study of PAOs, linking identified proteins to EBPR processes using metagenomic sequences.
2008	Wilmes, et al. [27] performed high-resolution metaproteomics combined with a metagenomic database, revealing strain differences among "*Ca.* Accumulibacter" members.
2009	Flowers, et al. [28] showed that different "*Ca.* Accumulibacter" clades have varying denitrification abilities, with clade IA capable of using nitrate as an electron acceptor.
2009	Wexler, et al. [29] studied protein dynamics in EBPR cycles, confirming glycolysis through the EMP pathway and noting protein carry-over between anaerobic and aerobic phases.
2009	Lopez-Vazquez, et al. [30] developed a kinetic model to identify conditions where PAOs outcompete GAOs, calibrating parameters to fit experimental results.
2010	Nielsen, et al. [31] called for improved quantitative models that integrate ecological and metabolic principles for better understanding of EBPR communities.
2010	Oehmen, et al. [32] showcase ecological principles in the incorporation of microbial metabolic models of PAOs and GAOs in EBPR.
2010	He, et al. [33] performed the first metatranscriptomics analysis of PAOs, highlighting gene expression patterns between anaerobic and aerobic stages.
2011	He and McMahon [34] used transcriptomic analysis with RT-qPCR to show dynamic gene expression changes during EBPR cycles, responding to acetate, oxygen, and phosphate levels.

2011	Nguyen, et al. [35] demonstrated that EBPR systems include other key microbial players, such as the diverse *Tetrasphaera* group, beyond *Accumulibacter*.
2012	Acevedo, et al. [36] demonstrated metabolic flexibility in "Ca. Accumulibacter," showing varying stoichiometries with polyP and glycogen and resulting changes in clade composition.
2012	Albertsen, et al. [37] performed a quantitative metagenomic study of a full-scale EBPR system, highlighting the lack of reference genomes and showing low abundance of "Ca. Accumulibacter" in full-scale systems.
2013	Kristiansen, et al. [38] developed the first metabolic model for *Tetrasphaera* in EBPR using metagenomics data.
2013	Flowers, et al. [39] retrieved the MAG of a new clade of "Ca. Accumulibacter," showing that clades with high 16S rRNA sequence similarity can have distinct genomic and functional potentials.
2014	McIlroy, et al. [40] retrieved two MAGs for "*Ca.* Competibacter" (GAOs) and explored genomic differences underlying their distinct functions.
2014	Mao, et al. [41] conducted the first study combining metagenomics and metatranscriptomics for PAO enrichment but with limited biological insights.
2014	Lanham, et al. [42] developed a "black-box" parameterized kinetic model to explain experimental data on PAOs and GAOs in full-scale treatment plants.
2015	Nguyen, et al. [43] showed intracellular glycine accumulation in PAO enrichments dominated by *Tetrasphaera*.
2015	Skennerton, et al. [44] published eight new MAGs for "Ca. Accumulibacter," identifying pan-genomes, variable genomes, and differences in metabolic potential for electron acceptors.
2015	Mao, et al. [45] conducted a global study on the abundance and diversity of "*Ca*. Accumulibacter" clades in WWTPs, highlighting their diversity and prevalence worldwide.
2015	Welles, et al. [46] experimentally showed that PAOs use polyphosphate and glycogen as redundant ATP sources and revealed physiological differences between "*Ca*. Accumulibacter" clades.
2016	Barr, et al. [47] demonstrated that metaproteomics aligns better with FISH results than metagenomics for quantifying "*Ca*. Accumulibacter" and highlighted proteomic differences in granular vs. floccular sludge.
2016	Oyserman, et al. [48] identified dynamic temporal changes in the transcriptome of "*Ca*. Accumulibacter" during EBPR cycles, revealing novel metabolic routes such as hydrogen production and glycine consumption.

2016	Law, et al. [49] used molecular techniques to track "*Ca*. Accumulibacter" dynamics in full-scale WWTPs at high temperatures, showing GAO competition is less severe than previously thought.
2016	Albertsen, et al. [50] identified "*Ca.* Propionivibrio" as a GAO closely related to "Ca. Accumulibacter," often misidentified as PAOs due to probe overlap.
2017	Stokholm-Bjerregaard, et al. [51] surveyed 18 Danish EBPR plants over 9 years and found *Tetrasphaera* to be the most abundant PAO and *Micropruina* the most abundant GAO.
2017	Rubio-Rincón, et al. [52] showed that GAOs and PAOs can synergize during EBPR through electron sink interactions with the nitrogen cycle.
2017	Welles, et al. [53] characterized PAOs with varying polyphosphate levels, showcasing metabolic flexibility and capturing the first electron microscope photo of "*Ca*. Accumulibacter."
2017	Guo, et al. [54] used metagenomics to reveal microbial diversity in activated sludge and assigned functional roles, with "*Ca*. Accumulibacter" being highly prevalent.
2018	Zeng, et al. [55] monitored seasonal changes in the abundance and clade diversity of "*Ca*. Accumulibacter" in WWTPs using *ppk1* as a marker.
2019	Camejo, et al. [56] used metagenomics and metatranscriptomics in an EBPR reactor performing denitrification to show that gene expression indicates activity potential.
2019	Fernando, et al. [57] developed a FISH-RAMAN technique to identify and study the physiology of cells accumulating polyP, glycogen, and PHAs in situ.
2019	Rubio-Rincón, et al. [58] proposed "*Ca*. Accumulibacter delftensis" (from clade IC) to address clade diversity issues, showing it cannot denitrify.
2019	Arumugam, et al. [59] reported the first closed genome of "Ca. Accumulibacter," retrieved with high quality and coverage.
2019	Qiu, et al. [60] identified a MAG of "Ca. Accumulibacter" capable of using acetate and amino acids, linking their metabolism to phosphate cycling.
2019	Rubio-Rincón, et al. [61] using lactate as substrate for EBPR, showed that *Tetrasphaera* decouples polyP hydrolysis from carbon substrate uptake leading to loss of EBPR activity.
2020	da Silva, et al. [62] used stoichiometric metabolic modeling to show that PAO flexibility in EBPR stages is rooted in network structure and metabolic fluidity.

2021	McDaniel, et al. [63] compared gene expression across "Ca. Accumulibacter" MAGs, showing different genomes express genes differently during EBPR cycles.
2021	Petriglieri, et al. [64] identified two new PAO species, "Ca. Dechloromonas phosphoritropha" and "phosphorivorans," capable of cycling PHAs and glycogen.
2022	Chen, et al. [65] tested uptake mechanisms of acetate, propionate, glutamate, and aspartate in PAOs and GAOs, revealing microbial group-specific physiological differences.
2022	Petriglieri, et al. [66] reevaluated "Ca. Accumulibacter" phylogeny, proposing 18 novel species names and emphasizing MAGs for functional diversity analysis.
2022	Singleton, et al. [67] categorized the previously known *Tetrasphaera* members as new genus "*Ca.* Phosphoribacter", one of most abundant PAO present worldwide.
2023	Chen, et al. [68] showed that "*Ca*. Accumulibacter cognatus" can metabolize certain fermentation products, leading to phosphate release
2023	Ziliani, et al. [69] suggested that glucose is consumed in "Ca. Accumulibacter" enrichments, supported by 67% FISH evidence.
2023	Páez-Watson, et al. [70] employed quantitative models to predict the metabolic strategies of PAOs during a whole EBPR cycle.
2023	Kleikamp, et al. [71] showed that metagenomics, metaproteomics, and 16S rRNA provide differing quantitative insights into microbial community structures.
2024	Páez-Watson, et al. [72] demonstrated that computational tools assess metabolic capacity and diversity within simple models of individual MAGs.
2024	Xie, et al. [73] conducted evolutionary analysis of genes related to polyphosphate cycling in "Ca. Accumulibacter," highlighting activity during EBPR cycles.

References

1 Srinath, E., Sastry, C. & Pillai, S. Rapid removal of phosphorus from sewage by activated sludge. *Experientia* **15**, 339-340 (1959).

2 Levin, G. V. & Shapiro, J. Metabolic uptake of phosphorus by wastewater organisms. *Journal*

3 Menar, A. B. & Jenkins, D. Fate of phosphorus in waste treatment processes: enhanced removal of phosphate by activated sludge. *Environmental Science & Technology* **4**, 1115-1121 (1970).

4 Barnard, J. L. Cut P and N without chemicals. (1974).

5 Nichols, H. Full scale experimentation on the new Johannesburg extended aeration plants. *Water SA* **1**, 121-132 (1975).

6 Fuhs, G. W. & Chen, M. Microbiological basis of phosphate removal in the activated sludge process for the treatment of wastewater. *Microbial Ecology* **2**, 119-138 (1975).

7 Deinema, M., Habets, L., Scholten, J., Turkstra, E. & Webers, H. The accumulation of polyphosphate in Acinetobacter spp. *FEMS Microbiology Letters* **9**, 275-279 (1980).

8 Comeau, Y., Hall, K. J., Hancock, R. E. & Oldham, W. K. Biochemical model for enhanced biological phosphorus removal. *Water Research* **20**, 1511-1521 (1986).

9 Wentzel, M., Lotter, L., Loewenthal, R. & Marais, G. Metabolic behaviour of Acinetobacter spp. in enhanced biological phosphorus removal-a biochemical model. *Water Sa* **12**, 209-224 (1986).

10 Mino, T., Arun, V., Tsuzuki, Y. & Matsuo, T. in *Biological phosphate removal from wastewaters* 27-38 (Elsevier, 1987).

11 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnology and bioengineering* **43**, 461-470 (1994).

12 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. Stoichiometric model of the aerobic metabolism of the biological phosphorus removal process. *Biotechnology and bioengineering* **44**, 837-848 (1994).

13 Maurer, M., Gujer, W., Hany, R. & Bachmann, S. Intracellular carbon flow in phosphorus accumulating organisms from activated sludge systems. *Water Research* **31**, 907-917 (1997).

14 Hesselmann, R. P., Werlen, C., Hahn, D., van der Meer, J. R. & Zehnder, A. J. Enrichment, phylogenetic analysis and detection of a bacterium that performs enhanced biological phosphate removal in activated sludge. *Systematic and applied microbiology* **22**, 454-465 (1999).

15 Hesselmann, R., Von Rummell, R., Resnick, S. M., Hany, R. & Zehnder, A. Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Research* **34**, 3487-3494 (2000).

16 McMahon, K. D., Dojka, M. A., Pace, N. R., Jenkins, D. & Keasling, J. D. Polyphosphate kinase from activated sludge performing enhanced biological phosphorus removal. *Applied and environmental microbiology* **68**, 4971-4978 (2002).

17 Crocetti, G. R., Banfield, J. F., Keller, J. r., Bond, P. L. & Blackall, L. L. Glycogen-accumulating organisms in laboratory-scale and full-scale wastewater treatment processes. *Microbiology* **148**, 3353-3364 (2002).

18 Kong, Y., Nielsen, J. L. & Nielsen, P. H. Microautoradiographic study of Rhodocyclus-related

19 Venter, J. C. *et al*. Environmental genome shotgun sequencing of the Sargasso Sea. *science* **304**, 66-74 (2004).

20 Tyson, G. W. *et al*. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43 (2004).

21 Kong, Y., Nielsen, J. L. & Nielsen, P. H. Identity and ecophysiology of uncultured actinobacterial polyphosphate-accumulating organisms in full-scale enhanced biological phosphorus removal plants. *Applied and environmental microbiology* **71**, 4076-4085 (2005).

22 Oehmen, A. *et al*. Anaerobic and aerobic metabolism of glycogen-accumulating organisms selected with propionate as the sole carbon source. *Microbiology* **152**, 2767-2778 (2006).

23 Martín, H. G. *et al*. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature biotechnology* **24**, 1263-1269 (2006).

24 He, S., Gall, D. L. & McMahon, K. D. "Candidatus Accumulibacter" population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Applied and environmental microbiology* **73**, 5865-5874 (2007).

25 Carvalho, G., Lemos, P. C., Oehmen, A. & Reis, M. A. Denitrifying phosphorus removal: linking the process performance with the microbial community structure. *Water research* **41**, 4383-4396 (2007).

26 Wilmes, P. *et al*. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *The ISME journal* **2**, 853-864 (2008).

27 Wilmes, P., Wexler, M. & Bond, P. L. Metaproteomics provides functional insight into activated sludge wastewater treatment. *PloS one* **3**, e1778 (2008).

28 Flowers, J. J., He, S., Yilmaz, S., Noguera, D. R. & McMahon, K. D. Denitrification capabilities of two biological phosphorus removal sludges dominated by different 'Candidatus Accumulibacter'clades. *Environmental microbiology reports* **1**, 583-588 (2009).

29 Wexler, M., Richardson, D. J. & Bond, P. L. Radiolabelled proteomics to determine differential functioning of Accumulibacter during the anaerobic and aerobic phases of a bioreactor operating for enhanced biological phosphorus removal. *Environmental microbiology* **11**, 3029-3044 (2009).

30 Lopez-Vazquez, C. M. *et al*. Modeling the PAO–GAO competition: effects of carbon source, pH and temperature. *Water Research* **43**, 450-462 (2009).

31 Nielsen, P. H. *et al.* A conceptual ecosystem model of microbial communities in enhanced biological phosphorus removal

32 Oehmen, A., Carvalho, G., Lopez-Vazquez, C., Van Loosdrecht, M. & Reis, M. Incorporating microbial ecology into the metabolic modelling of polyphosphate accumulating organisms and glycogen accumulating organisms. *Water research* **44**, 4992-5004 (2010).

33 He, S. *et al.* Metatranscriptomic array analysis of 'Candidatus Accumulibacter phosphatis'-enriched enhanced biological phosphorus removal sludge. *Environmental microbiology* **12**, 1205-1217 (2010).

34 He, S. & McMahon, K. D. J. M. b. Microbiology of 'Candidatus Accumulibacter'in activated sludge. *Microbial biotechnology* **4**, 603-619 (2011).

35 Nguyen, H. T. T., Le, V. Q., Hansen, A. A., Nielsen, J. L. & Nielsen, P. H. High diversity and abundance of putative polyphosphate-accumulating Tetrasphaera-related bacteria in activated sludge systems. *FEMS microbiology ecology* **76**, 256-267 (2011).

36 Acevedo, B. *et al.* Metabolic shift of polyphosphate-accumulating organisms with different levels of polyphosphate storage. *Water Research* **46**, 1889-1900 (2012).

37 Albertsen, M., Hansen, L. B. S., Saunders, A. M., Nielsen, P. H. & Nielsen, K. L. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *The ISME journal* **6**, 1094-1106 (2012).

38 Kristiansen, R. *et al.* A metabolic model for members of the genus Tetrasphaera involved in enhanced biological phosphorus removal. *The ISME journal* **7**, 543-554 (2013).

39 Flowers, J. J. *et al.* Comparative genomics of two 'Candidatus Accumulibacter'clades performing biological phosphorus removal. *The ISME journal* **7**, 2301-2314 (2013).

40 McIlroy, S. J. *et al.* 'Candidatus Competibacter'-lineage genomes retrieved from metagenomes reveal functional metabolic diversity. *The ISME journal* **8**, 613-624 (2014).

41 Mao, Y., Yu, K., Xia, Y., Chao, Y. & Zhang, T. Genome reconstruction and gene expression of "Candidatus Accumulibacter phosphatis" Clade IB performing biological phosphorus removal. *Environmental science & technology* **48**, 10363-10371 (2014).

42 Lanham, A. B. *et al.* Metabolic modelling of full-scale enhanced biological phosphorus removal sludge. *Water research* **66**, 283-295 (2014).

43 Nguyen, H. T. T., Kristiansen, R., Vestergaard, M., Wimmer, R. & Nielsen, P. H. Intracellular accumulation of glycine in polyphosphate-accumulating organisms in activated sludge, a novel storage mechanism under dynamic anaerobic-aerobic conditions. *Applied and Environmental Microbiology* **81**, 4809-4818 (2015).

44 Skennerton, C. T., Barr, J. J., Slater, F. R., Bond, P. L. & Tyson, G. W. Expanding our view of genomic

45 Mao, Y., Graham, D. W., Tamaki, H. & Zhang, T. Dominant and novel clades of Candidatus Accumulibacter phosphatis in 18 globally distributed full-scale wastewater treatment plants. *Scientific reports* **5**, 11857 (2015).

46 Welles, L. *et al*. Accumulibacter clades Type I and II performing kinetically different glycogen-accumulating organisms metabolisms for anaerobic substrate uptake. *Water research* **83**, 354-366 (2015).

47 Barr, J. J. *et al*. Metagenomic and metaproteomic analyses of Accumulibacter phosphatis-enriched floccular and granular biofilm. *Environmental microbiology* **18**, 273-287 (2016).

48 Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. *The ISME journal* **10**, 810-822 (2016).

49 Law, Y. *et al*. Integrative microbial community analysis reveals full-scale enhanced biological phosphorus removal under tropical conditions. *Scientific Reports* **6**, 25719 (2016).

50 Albertsen, M., McIlroy, S. J., Stokholm-Bjerregaard, M., Karst, S. M. & Nielsen, P. H. "Candidatus Propionivibrio aalborgensis": a novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal

51 Stokholm-Bjerregaard, M. *et al*. A critical assessment of the microorganisms proposed to be important to enhanced biological phosphorus removal in full-scale wastewater treatment systems. *Frontiers in microbiology* **8**, 718 (2017).

52 Rubio-Rincón, F., Lopez-Vazquez, C., Welles, L., Van Loosdrecht, M. & Brdjanovic, D. Cooperation between Candidatus Competibacter and Candidatus Accumulibacter clade I, in denitrification and phosphate removal processes. *Water Research* **120**, 156-164 (2017).

53 Welles, L. *et al*. Metabolic response of "Candidatus Accumulibacter Phosphatis" clade II C to changes in influent P/C ratio. *Frontiers in microbiology* **7**, 2121 (2017).

54 Guo, J. *et al*. Unraveling microbial structure and diversity of activated sludge in a full-scale simultaneous nitrogen and phosphorus removal plant using metagenomic sequencing. *Enzyme and Microbial Technology* **102**, 16-25 (2017).

55 Zeng, W., Zhang, L., Fan, P., Guo, J. & Peng, Y. Community structures and population dynamics of "Candidatus Accumulibacter" in activated sludges of wastewater treatment plants using ppk1 as phylogenetic marker. *Journal of Environmental Sciences* **67**, 237-248 (2018).

56 Camejo, P. Y., Oyserman, B. O., McMahon, K. D. & Noguera, D. R.

57 Fernando, E. Y. *et al*. Resolving the individual contribution of key microbial populations to enhanced biological phosphorus removal with Raman–FISH. *The ISME journal* **13**, 1933-1946 (2019).

58 Rubio-Rincón, F. *et al.* "Candidatus Accumulibacter delftensis": A clade IC novel polyphosphate-accumulating organism without denitrifying activity on nitrate. *Water research* **161**, 136-151 (2019).

59 Arumugam, K. *et al.* Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data. *Microbiome* **7**, 1-13 (2019).

60 Qiu, G. *et al.* Metabolic traits of Candidatus Accumulibacter clade IIF strain SCELSE-1 using amino acids as carbon sources for enhanced biological phosphorus removal. *Environmental Science & Technology* **54**, 2448-2458 (2019).

61 Rubio-Rincón, F. J. *et al.* Effect of lactate on the microbial community and process performance of an EBPR system. *Frontiers in microbiology* **10**, 125 (2019).

62 da Silva, L. G. *et al.* Revealing the Metabolic Flexibility of "Candidatus Accumulibacter phosphatis" through Redox Cofactor Analysis and Metabolic Network Modeling. *Applied Environmental Microbiology* **86** (2020).

63 McDaniel, E. A. *et al.* Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics. *Msystems* **6**, e00474-00421 (2021).

64 Petriglieri, F. *et al.* "Candidatus Dechloromonas phosphoritropha" and "Ca. D. phosphorivorans", novel polyphosphate accumulating organisms abundant in wastewater treatment systems. *The ISME Journal* **15**, 3605-3614 (2021).

65 Chen, L. *et al.* Carbon uptake bioenergetics of PAOs and GAOs in full-scale enhanced biological phosphorus removal systems. *Water Research* **216**, 118258 (2022).

66 Petriglieri, F. *et al.* Reevaluation of the Phylogenetic Diversity and Global Distribution of the Genus "Candidatus Accumulibacter". *mSystems* **7**, e00016-00022 (2022).

67 Singleton, C. *et al.* The novel genus,'Candidatus Phosphoribacter', previously identified as Tetrasphaera, is the dominant polyphosphate accumulating lineage in EBPR wastewater treatment plants worldwide. *The ISME Journal* **16**, 1605-1616 (2022).

68 Chen, L. *et al.* Candidatus Accumulibacter use fermentation products for enhanced biological phosphorus removal. *Water Research* **246**, 120713 (2023).

69 Ziliani, A. *et al.* Putative metabolism of Ca. Accumulibacter via the utilization of glucose. *Water Research* **229**, 119446 (2023).

70 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: Application to polyphosphate accumulating organisms. *Water Research* **228**, 119365 (2023).

71 Kleikamp, H. B. *et al.* Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. *Water research* **246**, 120700 (2023).

72 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. From metagenomes to metabolism: Systematically assessing the metabolic flux feasibilities for “Candidatus Accumulibacter” species during anaerobic substrate uptake. *Water Research* **250**, 121028 (2024).

73 Xie, X. *et al*. Integrated genomics provides insights into the evolution of the polyphosphate accumulation trait of Ca. Accumulibacter. *Environmental Science and Ecotechnology* **20**, 100353 (2024).

2

SWEET SECRETS.

EXPLORING NOVEL GLYCANS AND GLYCOCONJUGATES IN THE EXTRACELLULAR POLYMERIC SUBSTANCES OF "CANDIDATUS ACCUMULIBACTER"

Published as: Páez-Watson, T., Tomás-Martinez, S., de Wit, R., Keisham, S., Tateno, H., van Loosdrecht, M. C. M., & Lin, Y.l. Sweet Secrets: Exploring Novel Glycans and Glycoconjugates in the Extracellular Polymeric Substances of "Candidatus Accumulibacter". ACS Es&t Water 4, 3391-3399 (2024).

Supplementary figures and tables are available in the online publica]on

Abstract

Biological wastewater treatment relies on microorganisms that grow as flocs, biofilms or granules for efficient separation of biomass from cleaned water. This biofilm structure emerges from the interactions between microbes which produce, and are embedded in, extracellular polymeric substances (EPS). The true composition and structure of the EPS responsible for dense biofilm formation is still obscure. We conducted a *bottom-up* approach utilizing advanced glycomic techniques to explore the glycan diversity in the EPS from a highly enriched "*Candidatus* Accumulibacter" granular sludge. Rare novel sugar monomers such as N-Acetylquinovosamine (QuiNAc) and 2-O-Methylrhamnose (2-OMe-Rha) were identified to be present in the EPS of both enrichments. Further, a high diversity in the glycoprotein structures of said EPS was identified by means of lectin-based microarrays. We explored the genetic potential of "*Ca.* Accumulibacter" high quality metagenome assembled genomes (MAGs) to showcase the shortcoming of *top-down* bioinformatics-based approaches at predicting EPS composition and structure, especially when dealing with glycans and glycoconjugates. This work suggests that more *bottom-up* research is necessary to understand the composition and complex structure of EPS in biofilms, since genome-based inference cannot directly predict glycan structures and glycoconjugate diversity.

Introduction

Biological wastewater treatment relies on microbial communities that form aggregates called biofilms, flocs or granules, which play a pivotal role in the separation of biomass from treated water [1,2]. These structures house microorganisms embedded within a complex mixture of extracellular polymeric substances (EPS), which are produced by the microorganisms themselves [3]. Despite the intricate nature of EPS, significant progress has been made by focusing research on community members that are easily controllable in lab reactors. Among these organisms, "*Candidatus* Accumulibacter", a well-studied gram-negative bacterium, emerges as a dominant member in most aerobic granular sludge (AGS) systems [4] and is believed to play a major role in EPS formation. Despite not being isolated as a pure culture, "*Ca.* Accumulibacter" can be highly enriched in open lab cultures while maintaining the desired biofilm granular structure. Consequently, "*Ca.* Accumulibacter" has become a valuable model organism to study not only EPS formation but also the functioning, relationships, and assembly of microbial aggregates more broadly.

EPS plays a pivotal role in biofilm formation [5], provides protection against predation and environmental stress [6], facilitates nutrient cycling [7], and shapes overall microbial community structure [8]. Their composition, exceedingly complex, emerges from active secretion, cell decay and sorption from the environment [9]. Thus they comprise of sugars, proteins, nucleic acids and lipids, although the reported composition is strongly dependent on the method employed for its extraction [10] and analysis [11]. Focusing on isolating EPS into their individual molecular components overlooks the potential existence of combinations of these molecules. In this context, "glycans", which denote sugar chains, can be found as free molecules or linked to other macromolecules, particularly proteins and lipids.

Glycans are one of the most complex macromolecules in nature. Not only are their basic components diverse (typically ranging from 3 to 7 carbons) but the types of linkages (*i.e.* glycosidic bonds) that can occur at each individual carbon leads to different degrees of branching resulting in a nearly unlimited range of structures [8,12]. In addition, in a microbial community each individual member could contribute to a unique set of glycan molecules which further hinders the understanding of the EPS's glycome. Thus, developing systematic methods to better understand the sugar component that makes the EPS of a biofilm is of paramount importance. One such method is a '*top down'* approach in which the

genetic makeup of microbial communities can be analysed and the potential for production of glycans predicted [13]. This method however is limited to a set of well-studied polysaccharides and lacks the discovery of novel or unknown structures. We propose a different method which involves a '*bottom up*' approach to start with examining the glycan composition (*i.e. what is there)* to guide the further analysis on a species-based proteomic or genomic analysis.

Recent advances in next-generation mass spectrometry and an ever-growing resolution have revolutionized our ability to explore the composition of glycans from environmental samples [14-16]. The high precision and sensitivity allow for the identification of novel glycans. By employing these cutting-edge techniques, researchers are now equipped to identify and characterize new glycan structures within the EPS of "*Ca*. Accumulibacter", this way expanding our knowledge of the glycan diversity in these bacteria. High throughput techniques such as lectin microarrays [17] exploit the natural selectivity of lectins to recognize specific glycan structures. Recently, the use of this technique was combined with protein identification opening the possibility to study glycoconjugates such as glycoproteins.

Glycoproteins in bacteria have only recently gained scientific attention, as glycosylation was long believed to be exclusive to eukaryotic organisms [18,19]. However, pathogenic bacteria have been found to contain multiple glycoproteins that play significant roles in various processes, for example the bacterial adhesion to host mucosal membranes [20]. In addition, an array of glycoproteins were recently discovered in bacteria from environmental samples, e.g. from an enrichment of anaerobic ammonium oxidizing (ANAMMOX) bacteria [14,21] indicating not only their presence but also their high variety. Consequently, it is crucial to continue investigating the presence of glycoproteins in environmental bacteria and explore their potential connections to the formation and function of EPS.

Glycoproteins result from protein glycosylation, a post-translational modification that influences protein structure, stability, and functionality. Two primary protein glycosylation systems have been identified in bacteria: *en-bloc* and *sequential* glycosylation [19]. *En-bloc* glycosylation involves the assembly of a lipid-oligosaccharide in the cytoplasmic membrane, followed by the export and transfer to a protein in the extracellular space [22,23]. Conversely, sequential glycosylation entails the stepwise transfer of sugar moieties (mono or oligosaccharides) onto proteins [24]. While extensive information exists regarding these processes in model organisms like *Campylobacter jejuni* and *Haemophilus influenzae*, limited knowledge is available concerning protein glycosylation

mechanisms in microorganisms commonly found in wastewater treatment plants, such as "*Ca*. Accumulibacter".

In this paper, we aim to uncover the functional significance of glycoproteins and their associated glycans within the EPS of *"Ca.* Accumulibacter". For this, we adopted a comprehensive bottom-up approach to investigate the diversity of the glycome within the EPS of "*Ca*. Accumulibacter." Utilizing advanced glycomic techniques, we identified previously elusive novel glycan structures and explored the variety of glycoproteins present in two highly enriched "*Ca.* Accumulibacter" granular cultures. Guided by these results, we examined the genetic potential of available genomes of "*Ca.* Accumulibacter" to produce novel glycans and glycoproteins. This work highlights the importance of a thorough analysis of structural components of EPS rather than relying solely on functional roles from genomic-only inferred components.

Results

Reactors performance and microbial community

Two reactor enrichments were operated under the same conditions except for the rotational speed of the impeller (reactor 1: 400 rpm; reactor 2: 800 rpm). Both enrichments achieved a steady state in which the cyclic profiles of phosphate and acetate concentrations were typical of a polyphosphate accumulating organisms (PAOs) enrichment (Figure 1: Activity) [25]. Biomass concentrations in both reactors were relatively comparable at 4.61 ± 0.05 and 4.68 ± 0.08 g/L total suspended solids (TSS) for reactors 1 and 2 respectively. Further reactor characterization revealed a closely related microbial community based on 16S rRNA gene analysis (Figure 1: 16S), in both cases dominated by the genus *"Ca.* Accumulibacter". These findings align with the FISH results indicating a strong dominance of "*Ca*. Accumulibacter" in both enrichments (95.8 ± 4.4 % and 97.9 ± 2.3 % of biovolume in reactor 1 and 2 respectively).

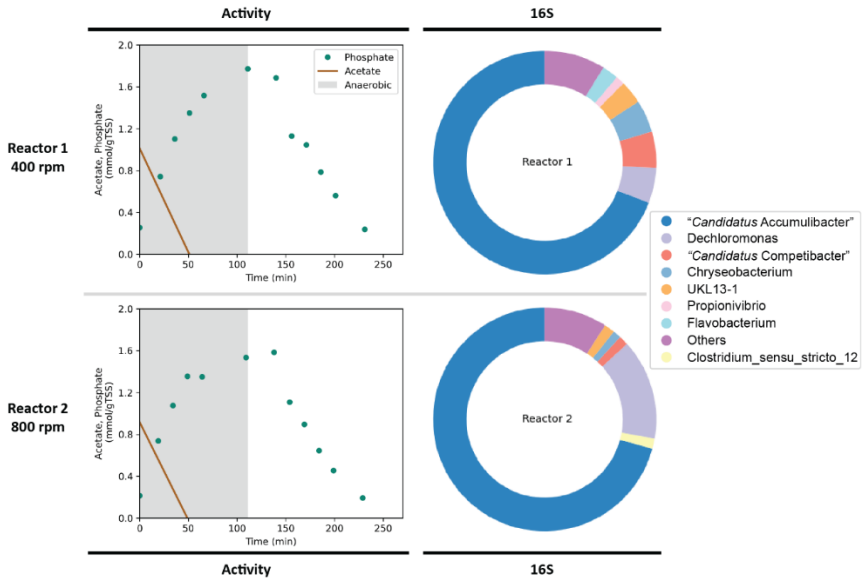


***Figure 1**. Reactor characteristics for the enrichments with the impeller rotating at 400 (top panel) and 800 (bottom panel) rpm at steady state. Each panel presents the activity test of a cycle by showing the concentrations of phosphate and acetate (mmol/gTSS) (left) and the microbial community abundance based on 16S rRNA amplicon sequencing (right). Both activity tests indicate that acetate was taken up during the anaerobic phase with the concurrent release of phosphate, typical for PAOs enrichments. For 16S rRNA results, the resolution at genus level indicates ≥ 1 % abundance, otherwise genus with < 1 % abundance were clustered into the category "Others".*

EPS yield and characterization

To characterize the "glycans" in the EPS of the "*Ca*. Accumulibacter" enrichment, the biomass from each reactor was collected at the end of the aerobic phase and the EPS were extracted. The total carbohydrate content was determined as 50.6 $mg_{eq\_glucose}/g_{EPS}$ for reactor 1 and 64.1 $mg_{eq_glucose}/g_{EPS}$for reactor 2. Additionally, the total protein content was determined as 288.9 mg_{eq_BSA}/g_{EPS} for reactor 1 and 398.6 mg_{eq_BSA} / g_{EPS} for reactor 2. Analysis of the specific glycosyl composition of the EPS (Figure 2) revealed a similar glycan profile, with the presence of both common carbohydrate monomers such as glucose (Glc), rhamnose (Rha), 2-O-Methylrhamnose (2-OMe-Rha), mannose (Man), galactose (Gal), ribose (Rub), N-Acetylglucosamine (GlcNAc), and relative uncommon monomer N-Acetylquinovosamine (QuiNAc).

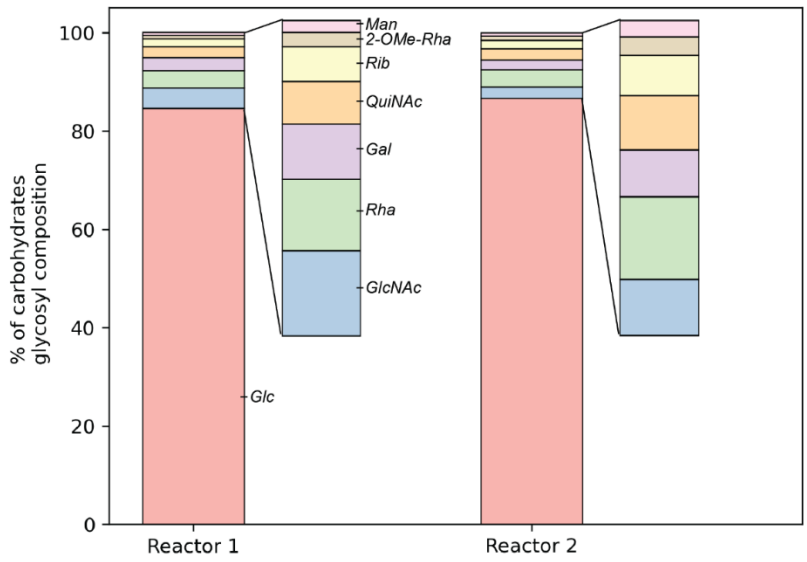


***Figure 2**. Glycosyl composition of the extracted EPS as relative mole abundance from the total amount of carbohydrate monomers determined by GC-MS. Carbohydrate monomers detected: glucose (Glc), Rhamnose (Rha), Mannose (Man), Galactose (Gal), Ribose (Rib,) N-Acetylglucosamine (GlcNAc), N-Acetylquinovosamine (QuiNAc) and 2-O-Methylrhamnose (2-OMe-Rha).*

Potential for biosynthesis of the carbohydrate monomer QuiNAc

Guided by the identification of the rare monomer QuiNAc in the EPS of our highly enriched reactors, it is interesting to investigate the genetic potential for its biosynthesis in "*Ca.* Accumulibacter" species. The pathway for QuiNAc synthesis identified in *Pseudomonas aureginosa* (also present in *Rhizobium elti* and *Bacilus cereus*) is shown in Figure 3A. The first steps involve the biochemical conversions from Fructose-6-Phosphate (a glycolytic intermediate) towards UDP-GlcNAc catalyzed by the enzymes coded by *GlmS, GlmM* and *GlmU*. Next, UDP-GlcNAc is dehydrated and further oxidized by two distinct enzymes (coded by *wbpM* and *wbpV* respectively) to generate UDP-QuiNAc. Analysis of "*Ca.* Accumulibacter" MAGs indicated that all assessed species harvested the complete gene set for synthesis up to UDP-GlcNAc (Figure 3.B). For the further conversion of this sugar towards UDP-QuiNAc, several MAGs contained the *wbpM* gene, but none of the MAGs were annotated to harvest *wbpV*. Nevertheless, several MAGs contained coding sequences that were matched to *wbpV* with over 40 % identity but had been annotated only as '*SDR family oxidoreductases'* (Figure 3). These could represent genes carrying the function of *wbpV* and thus represent the potential for QuiNAc synthesis in *"Ca.* Accumulibacter" species.

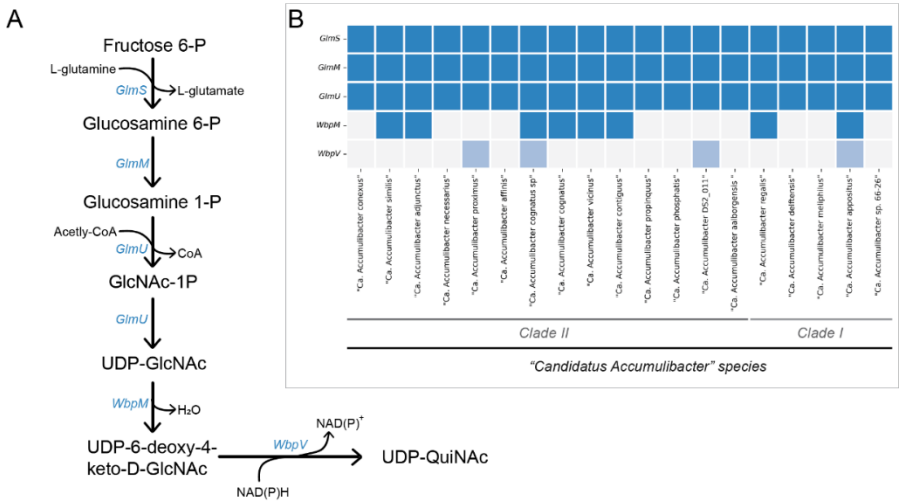


***Figure 3**. Genetic potential for the biosynthesis of QuiNAc in "Ca. Accumulibacter". (A) Biosynthetic pathway of the glycan QuiNAc in bacteria indicating the genes coding for each reaction step enzyme. (B) Presence (filled with blue) or absence (empty) of the genes involved in this biosynthetic pathway in multiple metagenome-assembled genomes (MAGs) of "Ca. Accumulibacter" species. Genes with BLAST hit > 40 % identity but not annotated as such are filled with lighter blue.*

Glycoprotein analysis with lectin microarrays

Glycans include both free carbohydrates and glycoconjugates (glycoproteins and glycolipids). Since protein glycosylation is a key posttranslational modification to proteins, the possible presence of glycoproteins in the EPS was studied. A lectin microarray was used to analyse the protein glycosylation within the EPS of "*Ca.* Accumulibacter". In this assay, proteins in the extracted EPS were initially labelled with Cy3. If a protein was glycosylated, the glycan part would bind to the specific lectin present on the array and would emit a fluorescent signal due to the presence of Cy3 at the protein part. Therefore, from this essay it is possible to evaluate the presence of glycoproteins and identify the glycan profile based on the lectin specificity. In brief, a fluorescent signal signifies two things: first, the attached proteins are glycoproteins, and second, their glycan profile matches the pattern recognized by the lectin.

Among the 96 lectins tested, 63 and 52 emitted a detectable fluorescent signal for the extracted EPS from reactors 1 and 2 respectively. To focus on the strongest signals, a filter (fluorescence intensity > 200) was applied, sorting out 17 lectins that bound significantly to the EPS (Figure 4). It was found that, the fluorescence intensity profiles were similar for both reactors 1 and 2. Notably, lectins binding glycans containing specific sugar monomers such as *rGRFT* (*mannose* containing glycans), *rCGL2* and *rGal3C* (*galactose* containing glycans), along with *PVL* (Sialic

acids containing glycans), *HEA* (O-glycans) and *FLAG-EW29Ch-E200K (6-sulfo-galactose glycan),* exhibited the highest fluorescence in both cases.

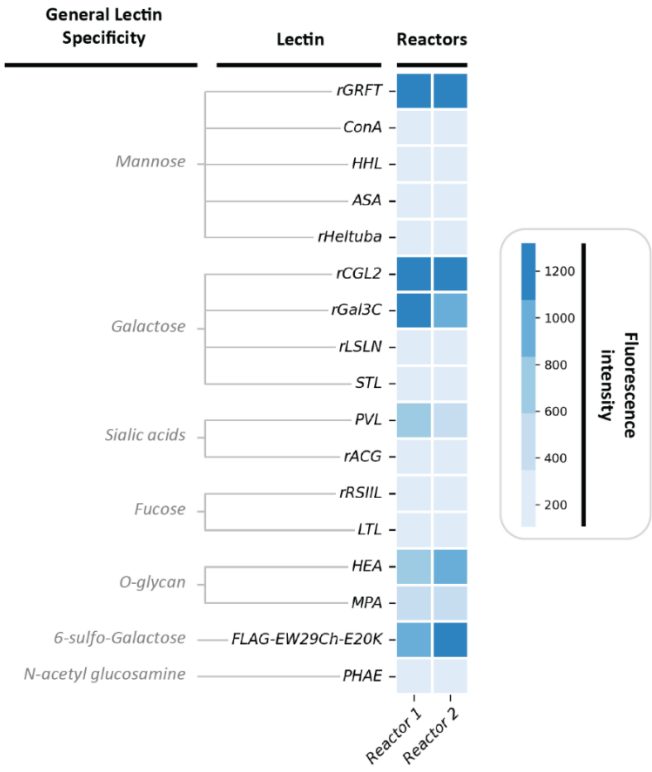


***Figure 4**. Lectin microarray profile indicating the fluorescence intensity for binding of glycoproteins in the EPS to each individual lectin. The broad specificity of each lectin is showed, more specific structural specificity is included in the published article online.*

Potential for protein glycosylation in "*Ca.* Accumulibacter"

The lectin microarrays results revealed a diverse array of glycoprotein structures in the EPS of both reactors. Glycoproteins assembly typically involves the transfer of an oligosaccharide from a lipid-oligosaccharide to a protein, and the diversity stems from variations in the oligosaccharide assembly. To investigate the genetic potential responsible for lipid-oligosaccharide assembly, MAGs of "*Ca.* Accumulibacter" were compared to the well-described lipid-oligosaccharide assembly system of *Campylobacter jejuni* (Figure 5.A*)*. The analysis of gene presence and absence in diverse species of "*Ca.* Accumulibacter" revealed significant variations in the assembly system for oligosaccharides linked to glycoprotein synthesis. While some species exhibit 2 or 3 related genes, others possess near-complete systems akin to *C. jejuni* such as "*Ca*. Accumulibacter regalis" (Figure 5.B). These differences imply potential species-related diversity in

glycoprotein structures, as evidenced by the wide array of glycoprotein structures observed in lectin microarrays.

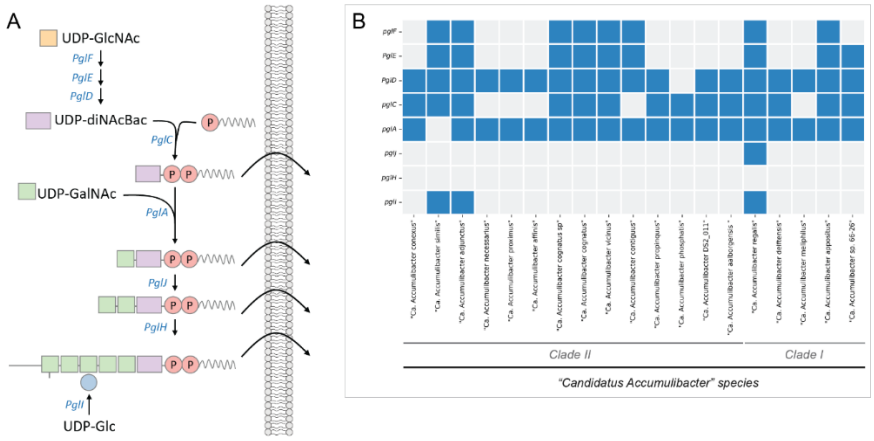


***Figure 5**. (A) Protein glycosylation mechanism present in C. jejuni (figure adapted from Nothaft and Szymanski [19]) (B) Presence (filled) or absence (empty) of the genes involved in this biosynthetic pathway in multiple metagenome assembled genomes (MAGs) of "Ca. Accumulibacter" species.*

Discussion

In this research we operated two lab-scale reactors with conditions to enrich for "*Ca.* Accumulibacter" to allow a deep understanding of the glycans and associated macromolecules produced in the EPS by members of these species. We obtained two highly enriched reactors (16S rRNA resulted in ~70 % "*Ca.* Accumulibacter" for both reactors – FISH indicated ~95 % of the biovolume) with remarkable similarities in their reactor performance and more importantly in the EPS glycans and glycoprotein profiles.

Identification of previously undescribed sugar monomers in the EPS of "*Ca.* Accumulibacter" enrichment

Bacteria coat themselves with a dense array of cell envelope glycans that enhance bacterial fitness and promote survival [26]. Within a microbial aggregate, this sweet coat may end up as a component of the EPS. Additionally, there are glycans specifically produced within the extracellular space. As bacterial glycans play a critical role in cell-cell and cell-environment interaction, it is significantly important to study the glycan profile of "*Ca.* Accumulibacter", which is one of the dominant microorganisms in EBPR systems. In the current research, lab-scale

reactors and various analytical methods were used to conduct this study. GC-MS analysis revealed the presence of novel glycans previously undocumented in EPS from "*Ca.* Accumulibacter": i.e. QuiNAc and 2-OMe-Rha.

QuiNAc has been reported in bacterial species of *Pseudomonas and Rhizobium* associated to lipopolysaccharides (LPS) [27] yet its function is not yet fully understood. QuiNAc-deficient mutants of *R. elti,* for example, exhibit LPS with significant reduction in the O-antigen content compared to the wild type. Such mutants fail to aggregate and colonize nodules in the roots of their legume hosts [28] even when the O-antigen content is increased by genetic engineering [29]. Thus, QuiNAc is proposed to serve as the bridging glycan between lipids and oligosaccharides in LPS [30-32]. It is worth pointing out the similarities between *R. elti* and "*Ca.* Accumulibacter" since both bacterial species appear to grow as densely aggregated microcolonies. In this respect, the role and the exact location of QuiNAc in "*Ca.* Accumulibacter" requires further research, which may shed light on maintaining a stable population of "*Ca.* Accumulibacter" in the EBPR system at wastewater treatment plant.

Besides QuiNAc, 2-O-methyl-rhamnose is another uncommon sugar monomer detected in the EPS of "Ca. Accumulibacter". 2-O-methyl-rhamnose has been reported on the S-layer glycoprotein glycan of *Geobacillus stearothermophilus* [33], It has also been reported as part of the repeating unit of the lipopolysaccharide from *Thiocapsa roseopersicina* [34] and as a spore-specific constituent of *Bacillus cereus* [35]. The role of 2-O-methyl-rhamnose is not clearly described in literature. It was hypothesized that 2-O-methylation of the terminal rhamnose residue on the S-layer glycoprotein glycan of *G. stearothermophilus* might function as a termination signal for chain elongation (Schäffer et al., 2002). Why it is produced by "*Ca.* Accumulibacter" enrichment and where it is located are interesting topics to be investigated.

Glycoproteins are present and highly diverse in the EPS of "*Ca.* Accumulibacter"

Within the glycans, besides free polysaccharides, there are glycoconjugates such as glycoproteins and glycolipids. To further investigate the potential existence of glycoproteins and its glycan profile, a lectin microarray analysis was performed. The existence of glycoproteins with diverse glycosylation patterns were observed. Protein glycosylation has profound effects on protein function and stability. For example, the surface layer proteins, which envelop almost all bacteria, have glycosylation patterns that significantly influence properties like water retention, surface roughness and fluidity [36]. In environmental microorganisms such as "*Ca.*

Accumulibacter", both the presence and strong diversity of glycoproteins in the EPS may be crucial for the functioning and assembly of the microbial community. This has significant implications for comprehending the role of EPS proteins since their functionality and structure can be fundamentally different depending on the type and diversity of the associated glycans [37].

Typically, approaches for studying glycoproteins in environmental samples involve identifying individual glycan structures and further characterizing the proteins with mass spectrometry [21,38,39]. Recently, Pabst, et al. [14] introduced a systematic glycoproteomics method, revealing a wide array of glycoproteins in an enrichment culture of anaerobic ammonium-oxidizing bacteria, aligning with our findings for similar environmental bacteria. While the described glycoproteomics approach effectively identifies specific proteins and glycan compositions, lectin microarrays, such as the method applied in this study, offer a high throughput examination of the glycans on the protein surfaces, enabling a broader screening of possible protein glycosylation pattern. Combining both approaches can provide a comprehensive understanding of glycoproteins, bridging the gap between structural characterization and functional implications.

In EPS research, identifying novel glycans and glycoconjugates needs a *bottom-up* approach

Bacteria produce a tremendous variety of unusual sugars and sugar linkages as well as modifications of sugars. The study of bacterial glycans is further complicated by their enormous structural diversity. In comparison, mammalian cells construct their cell surface glycans using only nine monosaccharide building blocks, plants use twelve monosaccharides, whereas >700 monosaccharides have been found in bacterial glycans [26].

Moreover, unlike DNA replication or protein translation, glycan biosynthesis is not directed by a pre-existing template molecule. Instead, the production of glycans is decided by a few factors together: the biosynthetic machinery, the available nucleotide sugars (which serve as monosaccharide donors), and signals from the intracellular and extracellular environment. Thus, the presence of glycans is dynamic and is influenced by both genetic and environmental factors [40]. Therefore, if the factors influencing glycans production and the remarkable variety of monosaccharides that can be produced by bacteria are added up, it is tremendously challenging to study the glycan composition in EPS.

Currently, *top-down* approaches are widely used. They predict the glycans composition in the EPS based on metagenomes [13], resulting in *theoretical*

polymeric substances that require experimental validation [41]. In addition, bioinformatic approaches based on DNA sequence are limited to only discovering biosynthetic pathways that have been very well described (*e.g.* cellulose biosynthesis [42]), shadowing the identification of unknown biomolecules and new glycans. As an example, in this research, the sugar monomer QuiNAc was detected in highly enriched "*Ca.* Accumulibacter" cultures, while its complete biosynthetic pathway could not be obtained in the available high quality genomes of *"Ca.* Accumulibacter". This indicates that with the top-down approach, the existence of QuiNAc can hardly be predicted.

Findings of the current research, together with those of Pabst, et al. [14], highlight the huge diversity encountered in the glycans in the EPS. Homology modelling of enzymes involved in carbohydrate synthesis and transfer rarely provides information on the type of monosaccharide involved in the process [43] which further hinders a complete description from metagenome information only. We identified that the genomes of "*Ca.* Accumulibacter" species harbour different sets of genes in the described system for oligosaccharides-lipid assembly, showcasing that even on a genus level there is a high potential for varying glycan compositions. Due to the special property of glycan synthesis, it is significantly necessary to use more *bottom-up* approaches for the chemical description of EPS components which could guide further genetic analysis and generalizations.

Conclusions

- Novel glycans containing QuiNAc and 2-OMe-Rha were identified for the first time in the EPS of "*Ca.* Accumulibacter" enrichments.
- Glycoproteins in the EPS from "*Ca.* Accumulibacter" are present and exhibit a high variation in the glycan structures that make them.
- The complexity in the EPS of environmental bacteria hinder the *top-down* approaches to only discover well-known polymeric substances. More *bottom-up* research is required to fill the knowledge gap that is required for genomic modelling approaches to understanding EPS of environmental bacteria.

Materials and methods

Reactor operation

Two reactor conditions were tested for this research. Both reactors were operated under the exact same conditions except for a change in the stirring speed of the reactor impeller (400 rpm vs 800 rpm) to enrich for different sized granules. The "*Ca.* Accumulibacter" enrichment was obtained in a 2 L (1.5 L working volume) sequencing batch reactor (SBR), following conditions similar to the one described by Guedes da Silva *et al.* (2020) with some adaptations. The reactor was inoculated using activated sludge from a municipal wastewater treatment plant (Harnaschpolder, The Netherlands). Each SBR cycle lasted 6 hours, consisting of 30 minutes of settling, 50 minutes of effluent removal, 10 minutes of N_2 sparging, 5 minutes of feeding, 130 minutes of anaerobic phase and 135 minutes of aerobic phase. The hydraulic retention time (HRT) was 12 hours (removal of 750 mL of broth per cycle). The average solids retention time (SRT) was controlled to 8 days by the removal of effluent at the end of the mixed aerobic phase. The pH was controlled at 7.0 ± 0.1 by dosing 1 M HCl or 1 M NaOH. The temperature was maintained at 20 ± 1 °C.

The reactor was fed with two separate media: a concentrated COD medium (400 mg COD/L) of acetate (17 g/L $NaAc \times 3H_2O$) and a concentrated mineral medium (1.53 g/L NH_4Cl, 1.59 g/L $MgSO_4 \times 7H_2O$, 0.40 g/L $CaCl_2 \times 2H_2O$, 0.48 KCl, 0.04 g/L N-allylthiourea (ATU), 2.22 g/L $NaH_2PO_4 \times H_2O$, 0.04 g/L yeast extract and 6 mL/L of trace element solution prepared following Smolders et al. (1994). In each cycle, 75 mL of each medium was added to the reactor, together with 600 mL of demineralized water. The final feed contained 400 mg COD/L of acetate. Extracellular concentrations of phosphate and ammonium were measured with a Gallery Discrete Analyzer (Thermo Fisher Scientific, Waltham, MA). Acetate was measured by high performance liquid chromatography (HPLC) with an Aminex HPX-87H column (Bio-Rad, Hercules, CA), coupled to RI and UV detectors (Waters, Milford, MA), using 0.0015 M phosphoric acid as eluent supplied at a flowrate of 1 mL/min.

Microbial community analysis

The microbial community of each reactor condition was characterized after a minimal of 4 residence times was reached (approximately 35 days of operation). Two orthogonal approaches were used for the community characterization: 16S amplicon sequencing and Fluoresence In Situ Hybdirization (FISH).

For 16S RNA amplicon sequencing, DNA was extracted from the granules using the DNeasy UltraClean Microbial kit (Qiagen, Venlo, The Netherlands), using the manufacturer's protocol. The extracted DNA was quantified using a Qubit 4 (Thermo Fisher Scientific, Waltham, MA). Samples were sent to Novogene Ltd. (Hong Kong, China) for amplicon sequencing of the V3-4 hypervariable region of the 16S rRNA gene (position 341-806) on a MiSeq desktop sequencing platform (Illumina, San Diego, CA) operated under paired-end mode. The raw sequencing reads were processed by Novogene Ltd. (Hong Kong, China) and quality filtered using the QIIME software [44]. Chimeric sequences were removed using UCHIME [45] and sequences with ≥97% identity were assigned to the same operational taxonomic units (OTUs) using UPARSE [46]. Each OTU was taxonomically annotated using the Mothur software against the SSU rRNA database of the SILVA Database [47]. Sequences obtained are deposited under the Bioproject accession number PRJNA1084229 in the NCBI database.

For FISH, samples underwent the procedures outlined by [48] for handling, fixation, and staining. Bacteria were selectively identified using a blend of EUB338, EUB338-II, and EUB338-III probes [49,50]. "*Ca*. Accumulibacter" was visualized employing a mixture of PAO462, PAO651, and PAO846 probes (referred to as PAOmix) [51]. Hybridized samples were subsequently examined utilizing the Axio Imager 2 fluorescence microscope (Zeiss, Oberkochen, Germany). To quantify and analyse the fluorescent pixels in the microscopic images, a custom image analysis tool was developed. The tool employs algorithms to identify and quantify different colour categories, including blue (Eubacteria only), purple (PAOmix + Eubacteria) and green (GAOmix + Eubacteria) providing a comprehensive analysis of the microbial composition. The tool is available on GitHub [https://github.com/TP-Watson/FISH-quantification-PaezWatson].

EPS extraction and characterization

EPS extraction from the biomass

Biomass samples collected at the end of the aerobic phase were freeze-dried prior to EPS extraction. EPS were extracted in alkaline conditions at high temperature, using a method adapted from Felz et al. (2016). Freeze-dried biomass were stirred in of 0.1 M NaOH (1 % w/v of volatile solids) at 80 °C for 30 min. Extraction mixtures were centrifuged at 4000xg at 4 °C for 20 min. Supernatants were collected and dialyzed overnight in dialysis tubing with a molecular cut-off of 3.5 kDa, frozen at-80 °C and freeze-dried. The freeze-dried extracted EPS samples were stored for further analysis.

Determination of the total protein and carbohydrate contents of the extracted EPS

The total protein content was estimated using the bicinchoninic acid (BCA) assay [52] with bovine serum albumin (BSA) as standard. The total carbohydrate content was determined using the phenol–sulfuric acid assay [53] with glucose as standard. Both analyses were performed as described by [11].

Glycosyl composition and detection of glycoproteins in the EPS

Glycosyl composition analysis of the extracted EPS was performed at the Complex Carbohydrate research Center (CCRC, the University of Georgia) by combined GC/MS of the O-trimethylsilyl (TMS) derivatives of the monosaccharide methyl glycosides produced from the sample by acidic methanolysis. These procedures were carried out as previously described in Santander, et al. [15]. In brief, lyophilized EPS aliquots of 300 µg were added to separate tubes with 20 µg inositol as the internal standard. Methyl glycosides were then prepared from the dry sample following the mild acid treatment by methanolysis in 1 M HCl in methanol at 80 °C (16 h). The samples were re-N-acetylated with 10 drops of methanol, 5 drops of pyridine, and 5 drops of acetic anhydride, and were kept at room temperature for 30 minutes (for detection of amino sugars). The sample was then per-o-trimethylsilyated by treatment with Tri-Sil (Pierce) at 80 °C (30 min). These procedures were carried out as described by Merkle and Poppe [54]. GC/MS analysis of the per-o-trimethylsilyl methyl glycosides was performed on an AT 7890A gas chromatograph interfaced to a 5975B MSD mass spectrometer, using a Supelco EC-1 fused silica capillary column (30 m × 0.25 mm ID) and the temperature gradient shown in Table 1.

***Table 1**. Temperature program for the GC-MS analysis for the TMS method.*

	Rate (ºC/min)	Value (ºC)	Hold Time (min)	Run Time (min)
Initial		80	2	2
Ramp 1	20	140	2	7
Ramp 2	2	200	0	37
Ramp 3	30	250	5	43.7

Identification of methylated sugar by alditol acetates

Identification of methylated sugar was performed by GC-MS of the alditol acetates as described Peña, et al. [16]. The analysis was performed on 400 mg of

the sample. The sample was hydrolyzed in 2 M trifluoroacetic acid (TFA) for 2 h in a sealed tube at 120 °C, reduced with $NaBD_4$, and acetylated using acetic anhydride/TFA. The resulting alditol acetates were analyzed on an Agilent 7890A GC (Table 2) interfaced to a 5975C MSD, electron impact ionization mode. A SP2331 fused silica capillary was used as column.

***Table 2**: Temperature program for GC-MS analysis by alditol acetates*

	Rate (ºC/min)	Value (ºC)	Hold Time (min)	Run Time (min)
Initial		60	1	1
Ramp 1	27.5	170	0	5
Ramp 2	4	235	2	23.5
Ramp 3	3	240	12	36.9

Glycoproteins detection by Lectin microarray

The high-density lectin microarray was constructed based on the procedure outlined by Tateno, et al. [17]. To label EPS, 0.4 µg of it was mixed with Cy3-N-hydroxysuccinimide ester (GE Healthcare). Excess Cy3 was removed using Sephadex G-25 desalting columns (GE Healthcare). The Cy3-labelled EPS was then diluted to a concentration of 0.5 µg/ml with probing buffer, which contained 25 mM Tris-HCl (pH 7.5), 140 mM NaCl, 2.7 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, and 1% Triton X-100. The mixture was incubated with the lectin microarray overnight at 20°C. The lectin microarray was washed three times with probing buffer, and the resulting fluorescence images were acquired using a Bio-Rex scan 200 evanescent-field-activated fluorescence scanner (Rexxam Co. Ltd., Kagawa, Japan).

Gene identity analysis

Genomic analysis was undertaken to explore the existence of genes within various "*Ca*. Accumulibacter" species that are associated with potential glycan synthesis and protein glycosylation machinery. We acquired MAG (Metagenome-Assembled Genome) sequences for 19 "*Ca*. Accumulibacter" species from the European Nucleotide Archive as described in Páez-Watson, et al. [55]. BLAST analysis was executed on the coding sequences of these genomes to identify the presence of (or potential for) specific genes in a reference set (reference genes

used in table S2). Sequence alignment was employed to evaluate conservation and recognize potential orthologs or homologs (min_identity 30 %, evalue e-12).

References

1 Beun, J. *et al.* Aerobic granulation in a sequencing batch reactor. *Water research* **33**, 2283-2290 (1999).

2 Seviour, R. & Nielsen, P. H. *Microbial ecology of activated sludge*. (IWA publishing, 2010).

3 Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nature reviews microbiology* **8**, 623-633 (2010).

4 Kleikamp, H. B. *et al.* Comparative metaproteomics demonstrates different views on the complex granular sludge microbiome. *bioRxiv*, 2022.2003. 2007.483319 (2022).

5 Frølund, B., Palmgren, R., Keiding, K. & Nielsen, P. H. Extraction of extracellular polymers from activated sludge using a cation exchange resin. *Water research* **30**, 1749-1758 (1996).

6 Whitton, B. A. *Ecology of cyanobacteria II: their diversity in space and time*. (Springer Science & Business Media, 2012).

7 Pinchuk, G. E. *et al.* Utilization of DNA as a sole source of phosphorus, carbon, and energy by Shewanella spp.: ecological and physiological implications for dissimilatory metal reduction. *Applied and environmental microbiology* **74**, 1198-1208 (2008).

8 Flemming, H.-C. *et al.* The biofilm matrix: Multitasking in a shared space. *Nature Reviews Microbiology* **21**, 70-86 (2023).

9 Liu, Y. & Fang, H. H. Influences of extracellular polymeric substances (EPS) on flocculation, settling, and dewatering of activated sludge. (2003).

10 Felz, S., Al-Zuhairy, S., Aarstad, O. A., van Loosdrecht, M. C. & Lin, Y. M. Extraction of structural extracellular polymeric substances from aerobic granular sludge. *JoVE (Journal of Visualized Experiments)*, e54534 (2016).

11 Felz, S., Vermeulen, P., van Loosdrecht, M. C. & Lin, Y. M. Chemical characterization methods for the analysis of structural extracellular polymeric substances (EPS). *Water Research* **157**, 201-208 (2019).

12 Lens, P., O'Flaherty, V., Moran, A., Stoodley, P. & Mahony, T. *Biofilms in medicine, industry and environmental biotechnology*. (IWA publishing, 2003).

13 Dueholm, M. K. D. *et al.* Genetic potential for exopolysaccharide synthesis in activated sludge bacteria uncovered by genome-resolved metagenomics. *Water Research* **229**, 119485 (2023).

14 Pabst, M. *et al.* A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. *The ISME Journal* **16**, 346-357 (2022).

15 Santander, J. *et al.* Mechanisms of intrinsic resistance to antimicrobial peptides of Edwardsiella ictaluri and its influence on fish gut inflammation and virulence. *Microbiology* **159**, 1471 (2013).

16 Peña, M. J., Tuomivaara, S. T., Urbanowicz, B. R., O'Neill, M. A. & York, W. S. in *Methods in enzymology* Vol. 510 121-139 (Elsevier, 2012).

17 Tateno, H. *et al*. Glycome diagnosis of human induced pluripotent stem cells using lectin microarray. *Journal of Biological Chemistry* **286**, 20345-20353 (2011).

18 Mescher, M. F. & Strominger, J. L. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of Halobacterium salinarium. *Journal of Biological Chemistry* **251**, 2005-2014 (1976).

19 Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology* **8**, 765-778 (2010).

20 Formosa-Dague, C. *et al*. The role of glycans in bacterial adhesion to mucosal surfaces: how can single-molecule techniques advance our understanding? *Microorganisms* **6**, 39 (2018).

21 Boleij, M., Pabst, M., Neu, T. R., Van Loosdrecht, M. C. & Lin, Y. Identification of glycoproteins isolated from extracellular polymeric substances of full-scale anammox granular sludge. *Environmental science & technology* **52**, 13127-13135 (2018).

22 Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Molecular microbiology* **32**, 1022-1030 (1999).

23 Wacker, M. *et al*. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* **298**, 1790-1793 (2002).

24 Gross, J. *et al*. The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *Journal of Biological Chemistry* **283**, 26010-26015 (2008).

25 Welles, L. *et al*. Accumulibacter clades Type I and II performing kinetically different glycogen-accumulating organisms metabolisms for anaerobic substrate uptake. *Water research* **83**, 354-366 (2015).

26 Barrett, K. & Dube, D. H. Chemical Tools to Study Bacterial Glycans: A Tale from Discovery of Glycoproteins to Disruption of Their Function. *Israel Journal of Chemistry* **63**, e202200050 (2023).

27 Li, T., Simonds, L., Kovrigin, E. L. & Noel, K. D. In vitro biosynthesis and chemical identification of UDP-N-acetyl-d-quinovosamine (UDP-d-QuiNAc). *Journal of Biological Chemistry* **289**, 18110-18120 (2014).

28 Cava, J. R., Elias, P. M., Turowski, D. A. & Noel, K. D. Rhizobium leguminosarum CFN42 genetic regions encoding lipopolysaccharide structures essential for complete nodule development on bean plants. *Journal of bacteriology* **171**, 8-15 (1989).

29 Noel, K. D., Forsberg, L. S. & Carlson, R. W. Varying the abundance of O antigen in Rhizobium etli and its effect on symbiosis with Phaseolus vulgaris. *Journal of bacteriology* **182**, 5317-5324 (2000).

30 Forsberg, L. S., Bhat, U. R. & Carlson, R. W. Structural characterization of the o-antigenic polysaccharide of the lipopolysaccharide from Rhizobium

etli strain CE3: a unique O-acetylated glycan of discrete size, containing 3-O-methyl-6-deoxy-L-talose and 2, 3, 4-tri-O-methyl-L-fucose. *Journal of Biological Chemistry* **275**, 18851-18863 (2000).

31 Ojeda, K. J., Simonds, L. & Noel, K. D. Roles of predicted glycosyltransferases in the biosynthesis of the Rhizobium etli CE3 O antigen. *Journal of bacteriology* **195**, 1949-1958 (2013).

32 Li, T. & Noel, K. D. Synthesis of N-acetyl-d-quinovosamine in Rhizobium etli CE3 is completed after its 4-keto-precursor is linked to a carrier lipid. *Microbiology* **163**, 1890 (2017).

33 Schäffer, C. *et al*. The surface layer (S-layer) glycoprotein of Geobacillus stearothermophilus NRS 2004/3a: analysis of its glycosylation. *Journal of Biological Chemistry* **277**, 6230-6239 (2002).

34 Hurlbert, R. E., WECKESSER, J., THARANATHAN, R. N. & MAYER, H. Isolation and characterization of the lipopolysaccharide of Thiocapsa roseopersicina. *European Journal of Biochemistry* **90**, 241-246 (1978).

35 Fox, A., Black, G. E., Fox, K. & Rostovtseva, S. Determination of carbohydrate profiles of Bacillus anthracis and Bacillus cereus including identification of O-methyl methylpentoses by using gas chromatography-mass spectrometry. *Journal of clinical microbiology* **31**, 887-894 (1993).

36 Schuster, B. & Sleytr, U. B. Relevance of glycosylation of S-layer proteins for cell surface properties. *Acta biomaterialia* **19**, 149-157 (2015).

37 Lepenies, B. & Seeberger, P. H. Simply better glycoproteins. *Nature biotechnology* **32**, 443-445 (2014).

38 Fulton, K. M., Li, J., Tomas, J. M., Smith, J. C. & Twine, S. M. Characterizing bacterial glycoproteins with LC-MS. *Expert Review of Proteomics* **15**, 203-216 (2018).

39 Van Teeseling, M. C. *et al*. The S-layer protein of the anammox bacterium Kuenenia stuttgartiensis is heavily O-glycosylated. *Frontiers in microbiology* **7**, 1721 (2016).

40 Chen, L. M. *et al*. Sialylation and Sulfation of Anionic Glycoconjugates Are Common in the Extracellular Polymeric Substances of Both Aerobic and Anaerobic Granular Sludges. *Environmental Science & Technology* **57**, 13217-13225 (2023).

41 Seviour, T. *et al*. Extracellular polymeric substances of biofilms: Suffering from an identity crisis. *Water research* **151**, 1-7 (2019).

42 McNamara, J. T., Morgan, J. L. & Zimmer, J. A molecular description of cellulose biosynthesis. *Annual review of biochemistry* **84**, 895-921 (2015).

43 Szymanski, C. M. Bacterial glycosylation, it's complicated. *Frontiers in Molecular Biosciences* **9**, 1015771 (2022).

44 Caporaso, J. G. *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**, 335-336 (2010).

45 Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and

speed of chimera detection. *Bioinformatics* **27**, 2194-2200 (2011).

46 Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **10**, 996-998 (2013).

47 Quast, C. *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-D596 (2012).

48 Winkler, M.-K. *et al*. Selective sludge removal in a segregated aerobic granular biomass system as a strategy to control PAO–GAO competition at high temperatures. *Water research* **45**, 3291-3299 (2011).

49 Amann, R. I. *et al*. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and environmental microbiology* **56**, 1919-1925 (1990).

50 Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Systematic and applied microbiology* **22**, 434-444 (1999).

51 Crocetti, G. R. *et al*. Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantitation. *Applied and environmental microbiology* **66**, 1175-1182 (2000).

52 Smith, P. e. *et al*. Measurement of protein using bicinchoninic acid. *Analytical biochemistry* **150**, 76-85 (1985).

53 DuBois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. t. & Smith, F. Colorimetric method for determination of sugars and related substances. *Analytical chemistry* **28**, 350-356 (1956).

54 Merkle, R. K. & Poppe, I. in *Methods in enzymology* Vol. 230 1-15 (Elsevier, 1994).

55 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. From metagenomes to metabolism: Systematically assessing the metabolic flux feasibilities for “Candidatus Accumulibacter” species during anaerobic substrate uptake. *Water Research* **250**, 121028 (2024).

3

FROM METAGENOMES TO METABOLISM.

SYSTEMATICALLY ASSESSING THE METABOLIC FLUX FEASIBILITIES FOR "CANDIDATUS ACCUMULIBACTER" SPECIES DURING ANAEROBIC SUBSTRATE UPTAKE

Published as: Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. From metagenomes to metabolism: Systematically assessing the metabolic flux feasibili]es for "Candidatus Accumulibacter" species during anaerobic substrate uptake. Water Research 250, 121028 (2024).

Supplementary figures and tables are available in the online publication

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.

Introduction

Microbial ecology research strongly relies on cultivation independent approaches since most bacterial species are, to date, unculturable [1]. 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. [2]). 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 [3,4]. Particularly in the context of water engineering, MAG's metabolic potentials encompass functions like exopolysaccharide synthesis [5], nitrogen, phosphorus and iron removal [6] and even mutualistic interactions amongst species [7].

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 do within the community. Constraint based methods like flux balance analysis (FBA), provide tools to predict metabolic functions and have been successfully applied to study monocultures [8,9]. Efforts have been made to extend their applications towards understanding metabolic interactions [10,11], resource allocation [12], microbial biosynthesis [13], or even inferring functional guilds [14] from microbial communities in ecology (Dillard, et al. [15] 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 [16]. 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 [17-20]. 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 [21]. 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. [22] proposed that the complete tricarboxylic acid (TCA) cycle is active anaerobically as a source of NADH for the accumulation of polyhydroxybutyrate (PHB) (Figure 1.A). Mino, et al. [17] demonstrated anaerobic depletion of glycogen and suggested it as the source of NADH by means of glycolysis (Figure 1.B). Pereira, et al. [23] 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 (Figure 1.C). Conversely, Hesselmann, et al. [24] suggested a model including glycolysis and a split TCA cycle (Figure 1.D), introducing the use of one of the anaplerotic reactions (catalysing the conversion from pyruvate to oxaloacetate) while Yagci, et al. [25] introduced the concept of the glyoxylate shunt to bypass certain reactions within the TCA cycle (Figure 1.E). Burow, et al. [26] 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. [25]. Conversely, Zhou, et al. [27] 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 [28].

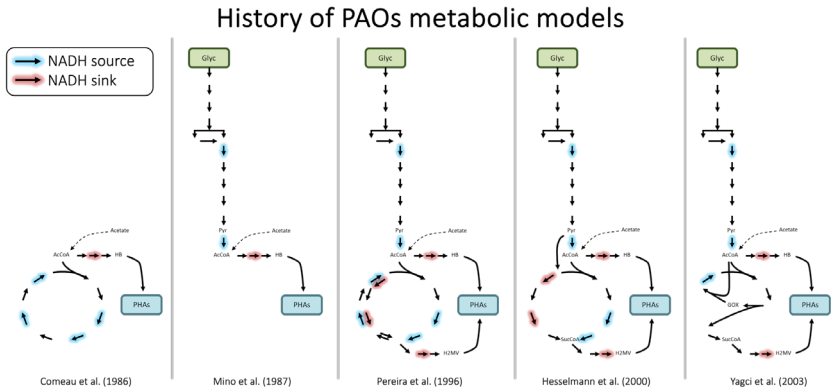


***Figure 1**. Schematic representation of the proposed metabolic transformations in PAOs during the anaerobic uptake of acetate over time [17,22-25].*

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" [29] 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 [30], 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 [31,32]. 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 [33]. The high number of flux 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 [34,35]. 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 [36] 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.

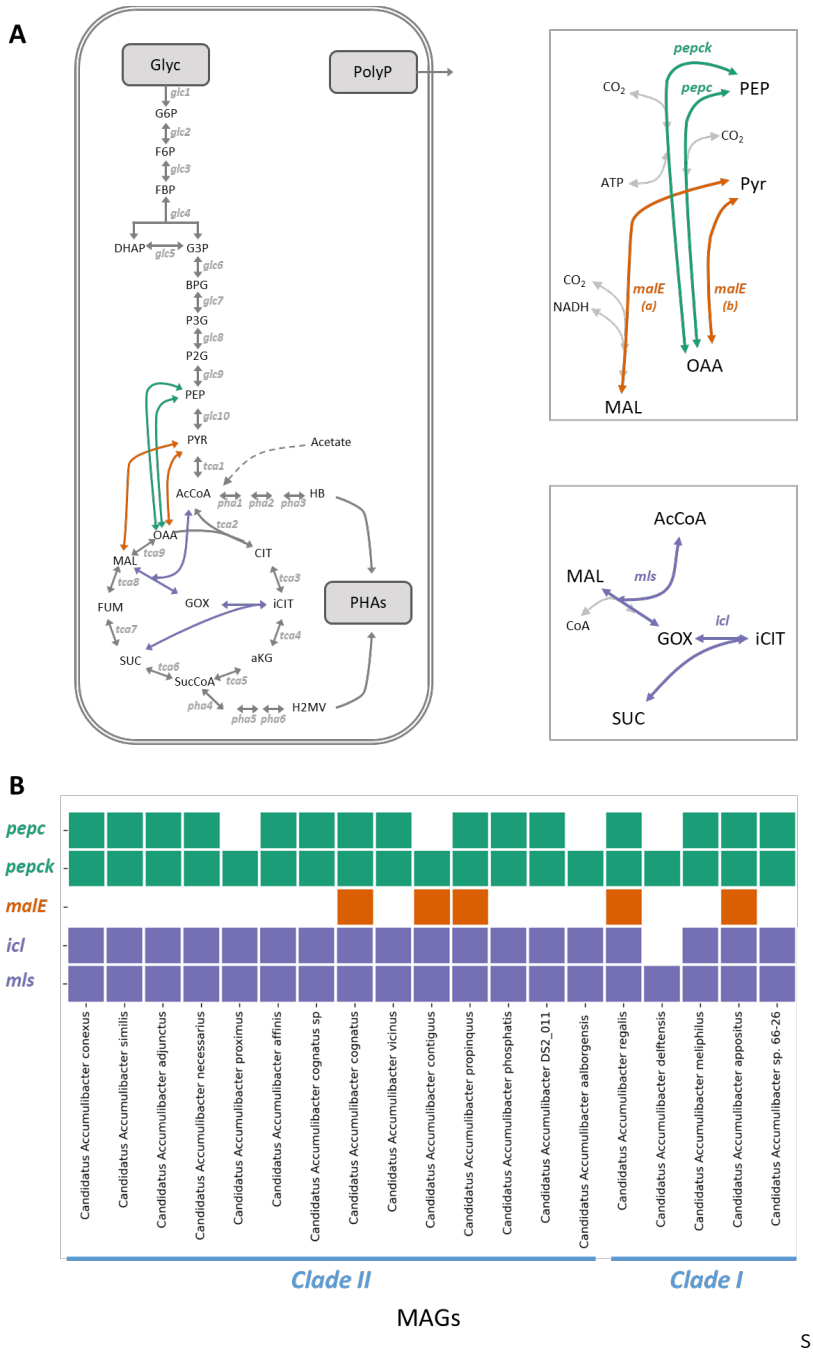
In this work, we analysed the anaerobic metabolism of PAOs to illustrate a systematic methodology for deriving consistent metabolic insights from MAGs. We assess the metabolic potential in central carbon metabolism of nineteen high quality MAGs of "*Ca.* Accumulibacter" species by employing EFMA and 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.

Results

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 (Figure 2.A). 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 (Figure 2.B.). 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*.



S

Figure 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).*

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 (Figure 3).

Among these solutions, we identified flux distributions that have been previously proposed in the literature, including studies by Comeau, et al. [22], Mino, et al. [21], Hesselmann, et al. [24] and Yagci, et al. [25] (highlighted in Figure 3). The solution corresponding to the model proposed by Pereira, et al. [23] 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.

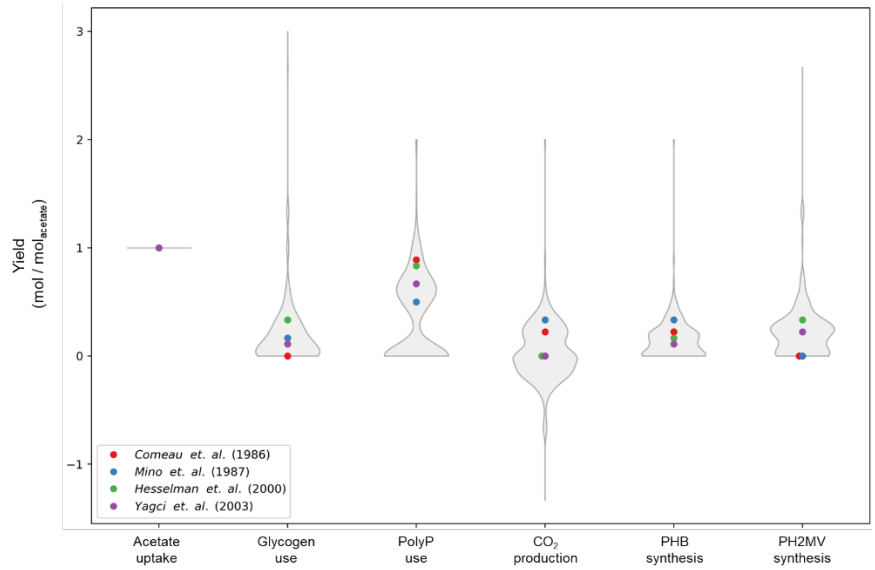


***Figure 3**. Model yields of glycogen and polyphosphate use, CO_2 production and PHB and PH2MV synthesis (mole) per mole of acetate consumed in the metabolic model of PAOs. The violin plots illustrate the distribution of stoichiometric values based on 700 unique solutions. Published solutions from previous studies by Comeau, et al. [22], Mino, et al. [21], Hesselmann, et al. [24] and Yagci, et al. [25] 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.*

A small subset of the metabolic model solutions is 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- Δ_rG' for each reaction) of a given metabolic model solution [36] (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 Δ_rG' (Figure 4A). The models proposed by Mino, et al. [21], Hesselmann, et al. [24] and Yagci, et al. [25] 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. [22] 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$_{acetate}$ (Figure 4.A and Figure 4B 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.

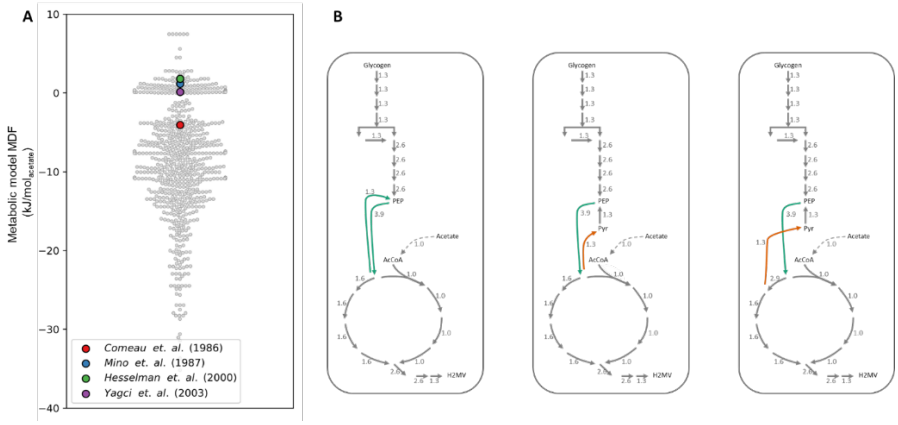


Figure 4. (A) *A swarm plot displaying the distribution of the Minimal Driving Force (MDF) achieved by all 700 metabolic model solutions. The MDF results based on the models proposed by Comeau, et al. [22], Mino, et al. [21], Hesselmann, et al. [24] and Yagci, et al. [25] are highlighted in distinct 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.*

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. Figure 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 (Figure 5A) and *glc10, tca1, tca2, tca3, tca4, tca5, pepck, icl, mls* and *malE* in the reverse direction

(Figure 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 [37] 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 (Figure 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 (Figure 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.

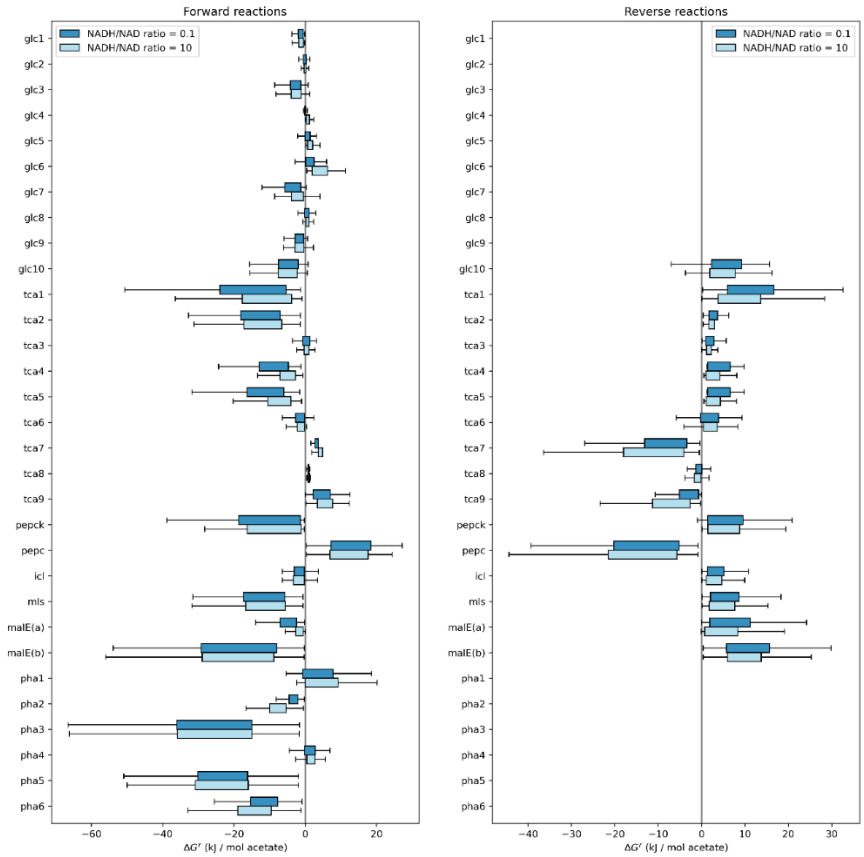


Figure 5. *Boxplot distribution of the calculated $\Delta_r G'$ for each reaction based on the 700 Elementary Flux Modes (EMF) solutions and respective Minimal Driving Force (MDF) optimization. Dark and light blue indicate different NADH/NAD⁺ ratios that were fixed for the MDF analysis simulating changes 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.*

Discussion

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 as a crucial step towards transitioning from descriptive studies in microbiome research towards more mechanistic and predictive studies, also called eco-systems frameworks [38,39]. With the increasing availability of high quality MAGs, research has rapidly focused on

assigning metabolic functions to given species based on gene presence/absence [2,6,40-43] and, in limited cases, integrating multi-omic data [44-48]. 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 [49-51]. 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 [52] 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 *Escherichia coli* the function of 35 % of coding sequences is unknown [53]). 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. [2]). 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.

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 [29]. 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. [28] 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 (Figure 2.B). We found that the presence of these reactions (especially their combination) contributed to the highest driving forces of the metabolic models (Figure 4.B). 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 [54]. 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 [55] and substantial reduction of growth yields [56]. Similarly, the *pepc/pepck* pair has been shown to participate in futile cycles in *Escherichia coli* [57]. Overall, a better understanding of the absence of these genes through the complete EBPR anaerobic/aerobic cycle is needed.

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 (Figure 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. [22], Mino, et al. [21], Hesselmann, et al. [24] and Yagci, et al. [25] are indeed stoichiometrically possible. However, the model proposed by [23] 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 Figure 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* (Figure 4.B). 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 (Figure 5), we outline the metabolic model for "*Ca.* Accumulibacter" in Figure 6, highlighting reactions that represent strong thermodynamic infeasibilities. Below we discuss on the implications of specific reactions on the model for "*Ca.* Accumulibacter".

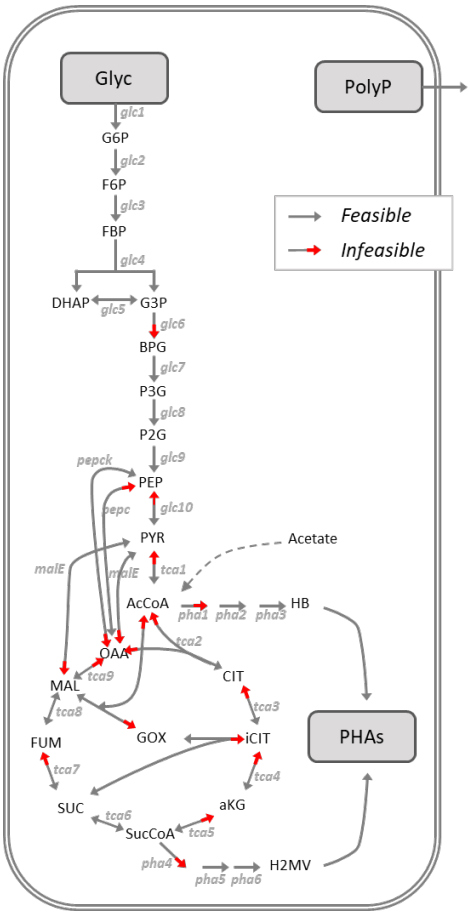


Figure 6. *Redefinition of the PAOs metabolic model for anaerobic uptake of acetate and accumulation of PHAs. Scheme represents the reactions (grey arrows) connecting metabolites participating in central carbon metabolism. Based on thermodynamic analysis, each reaction of the pathway has been defined 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).*

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 (Figure 5). Some theories suggest

succinate-to-fumarate conversion could occur through a novel cytochrome b/b6 fusion protein [58] or electron bifurcation mechanisms [59]. However, attempts at assessing this activity have failed [60]. 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 anaerobic 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 [61] 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 [36]. 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 [37,62,63], which would pull this reaction towards the infeasibility region. Further, van der Rest, et al. [64] 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. [28] 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 anaerobic conditions in the opposite reaction, a phenomenon previously observed in *Escherichia coli* cells [65].

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 [36]. 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 [58], proteomic [66] and enzymatic [67] 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 (Figure 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. [68] 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.

Trade-offs between stoichiometric yield and thermodynamic feasibility

The metabolic model solutions that resulted in the highest MDF (Figure 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 [69,70], 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 [42]) 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.

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.

Materials and methods

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. 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 Figure 1). For building more complex metabolic models, we recommend the user to apply available methods for model construction such as gapseq [50], CarveMe [51], 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:

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

where *v* represents a vector containing the fluxes of each reaction (column in **S**) [30]. 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 [13].

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. 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.

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 [31]. 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. [22], Mino, et al. [17], Pereira, et al. [23], Hesselmann, et al. [24] and Yagci, et al. [25]) were manually filtered based on the presence and direction of their active reactions within the EFMs.

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) [36]. 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 each 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 $-\Delta_rG'$). If this MDF ($-\Delta_rG'$) 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:

Maximize **B**

$$\boldsymbol{x}, \boldsymbol{B}$$

Subject to $-(\Delta_r G'^{o} + RT \cdot \boldsymbol{S}^T \boldsymbol{x}) \geq \boldsymbol{B}$

$$ln\,(\boldsymbol{C_{min}}) \leq \boldsymbol{x} \leq ln\,(\boldsymbol{C_{max}})$$

where **B** represents the minimum driving force of all reactions in the EFM, ***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 $\Delta_f G^{\circ\prime}$ 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. [36] 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 $\Delta_r G'$ 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 thermodynamically feasible. Feasible reactions were manually selected, and a stoichiometry scheme derived. The achieved optimized $\Delta_r G'$ values for each reaction in all the solutions were used to create a bar plot in Python, representing the distribution of $\Delta_r G'$.

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 [37]. MDF optimization was then performed on all EFMs with these conditions and resulting MDF and $\Delta_r G'$ where analysed in a similar fashion using Python.

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: 48899967 (https://gitlab.com/delft_paos/from-metagenomes-to-metabolism-paos).

REFERENCES

1 Steen, A. D. *et al.* High proportions of bacteria and archaea across most biomes remain uncultured. *The ISME journal* **13**, 3126-3130 (2019).

2 Singleton, C. M. *et al.* Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nature communications* **12**, 1-13 (2021).

3 Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661-1665 (2012).

4 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature communications* **7**, 13219 (2016).

5 Dueholm, M. K. D. *et al.* Genetic potential for exopolysaccharide synthesis in activated sludge bacteria uncovered by genome-resolved metagenomics. *Water Research* **229**, 119485 (2023).

6 Kristensen, J. M., Singleton, C., Clegg, L.-A., Petriglieri, F. & Nielsen, P. H. High diversity and functional potential of undescribed "Acidobacteriota" in Danish wastewater treatment plants. *Frontiers in microbiology* **12**, 643950 (2021).

7 Fujii, N. *et al.* Metabolic potential of the superphylum Patescibacteria reconstructed from activated sludge samples from a municipal wastewater treatment plant. *Microbes and environments* **37**, ME22012 (2022).

8 Famili, I., Förster, J., Nielsen, J. & Palsson, B. O. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences* **100**, 13134-13139 (2003).

9 O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using genome-scale models to predict biological capabilities. *Cell* **161**, 971-987 (2015).

10 Du, H. *et al.* Microbial active functional modules derived from network analysis and metabolic interactions decipher the complex microbiome assembly in mangrove sediments. *Microbiome* **10**, 1-17 (2022).

11 Kuppa Baskaran, D. K., Umale, S., Zhou, Z., Raman, K. & Anantharaman, K. Metagenome-based metabolic modelling predicts unique microbial interactions in deep-sea hydrothermal plume microbiomes. *ISME communications* **3**, 42 (2023).

12 Sharma, S. & Steuer, R. Modelling microbial communities using biochemical resource allocation analysis. *Journal of the Royal Society Interface* **16**, 20190474 (2019).

13 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A.

Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: Application to polyphosphate accumulating organisms. *Water Research* **228**, 119365 (2023).

14 Muller, E. E. *et al*. Using metabolic networks to resolve ecological properties of microbiomes. *Current Opinion in Systems Biology* **8**, 73-80 (2018).

15 Dillard, L. R., Payne, D. D. & Papin, J. A. Mechanistic models of microbial community metabolism. *Molecular Omics* **17**, 365-375 (2021).

16 Kim, M., Sung, J. & Chia, N. Resource-allocation constraint governs structure and function of microbial communities in metabolic modeling. *Metabolic Engineering* **70**, 12-22 (2022).

17 Mino, T., Arun, V., Tsuzuki, Y. & Matsuo, T. in *Biological phosphate removal from wastewaters* 27-38 (Elsevier, 1987).

18 Van Loosdrecht, M., Smolders, G., Kuba, T. & Heijnen, J. Metabolism of micro-organisms responsible for enhanced biological phosphorus removal from wastewater, Use of dynamic enrichment cultures. *Antonie van Leeuwenhoek* **71**, 109-116 (1997).

19 Oehmen, A. *et al*. Advances in enhanced biological phosphorus removal: from micro to macro scale. *Water Research* **41**, 2271-2300 (2007).

20 Nielsen, P. H., McIlroy, S. J., Albertsen, M. & Nierychlo, M. Re-evaluating the microbiology of the enhanced biological phosphorus removal process. *Current opinion in biotechnology* **57**, 111-118 (2019).

21 Mino, T., Van Loosdrecht, M. & Heijnen, J. Microbiology and biochemistry of the enhanced biological phosphate removal process. *Water Research* **32**, 3193-3207 (1998).

22 Comeau, Y., Hall, K. J., Hancock, R. E. & Oldham, W. K. Biochemical model for enhanced biological phosphorus removal. *Water Research* **20**, 1511-1521 (1986).

23 Pereira, H. *et al*. Model for carbon metabolism in biological phosphorus removal processes based on in vivo13C-NMR labelling experiments. *Water Research* **30**, 2128-2138 (1996).

24 Hesselmann, R., Von Rummell, R., Resnick, S. M., Hany, R. & Zehnder, A. Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Research* **34**, 3487-3494 (2000).

25 Yagci, N., Artan, N., Çokgör, E. U., Randall, C. W. & Orhon, D. Metabolic model for acetate uptake by a mixed culture of phosphate-and glycogen-accumulating organisms under anaerobic conditions. *Biotechnology bioengineering* **84**, 359-373 (2003).

26 Burow, L. C., Mabbett, A. N. & Blackall, L. L. Anaerobic glyoxylate cycle activity during simultaneous utilization of glycogen and acetate in uncultured Accumulibacter enriched in enhanced biological phosphorus removal communities. *The ISME Journal* **2**, 1040-1051 (2008).

27 Zhou, Y., Pijuan, M., Zeng, R. J. & Yuan, Z. Involvement of the TCA cycle in the anaerobic metabolism of polyphosphate accumulating organisms (PAOs). *Water research* **43**, 1330-1340 (2009).

28 da Silva, L. G. *et al*. Revealing the Metabolic Flexibility of "Candidatus Accumulibacter phosphatis" through Redox Cofactor Analysis and Metabolic Network Modeling. *Applied Environmental Microbiology* **86** (2020).

29 Petriglieri, F. *et al*. Reevaluation of the Phylogenetic Diversity and Global Distribution of the Genus "Candidatus Accumulibacter". *mSystems* **7**, e00016-00022 (2022).

30 Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nature biotechnology* **28**, 245-248 (2010).

31 Terzer, M. & Stelling, J. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* **24**, 2229-2235 (2008).

32 Maarleveld, T. R., Khandelwal, R. A., Olivier, B. G., Teusink, B. & Bruggeman, F. J. Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology journal* **8**, 997-1008 (2013).

33 Sweetlove, L. J., Beard, K. F., Nunes-Nesi, A., Fernie, A. R. & Ratcliffe, R. G. Not just a circle: flux modes in the plant TCA cycle. *Trends in plant science* **15**, 462-470 (2010).

34 Gerstl, M. P., Jungreuthmayer, C., Müller, S. & Zanghellini, J. Which sets of elementary flux modes form thermodynamically feasible flux distributions? *The FEBS journal* **283**, 1782-1794 (2016).

35 Peres, S., Jolicœur, M., Moulin, C., Dague, P. & Schuster, S. How important is thermodynamics for identifying elementary flux modes? *PLoS One* **12**, e0171440 (2017).

36 Noor, E. *et al*. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS computational biology* **10**, e1003483 (2014).

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

38 Lawson, C. E. *et al*. Common principles and best practices for engineering microbiomes. *Nature Reviews Microbiology* **17**, 725-741 (2019).

39 McDaniel, E. A. *et al*. Prospects for multi-omics in the microbial ecology of water engineering. *Water Research* **205**, 117608 (2021).

40 Woodcroft, B. J. *et al*. Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49-54 (2018).

41 Aalismail, N. A. *et al*. Functional metagenomic analysis of dust-associated microbiomes above the Red Sea. *Scientific reports* **9**, 13741 (2019).

42 Qiu, G. *et al*. Metabolic traits of Candidatus Accumulibacter clade IIF strain SCELSE-1 using amino acids

as carbon sources for enhanced biological phosphorus removal. *Environmental Science & Technology* **54**, 2448-2458 (2019).

43 Singleton, C. *et al*. The novel genus,'Candidatus Phosphoribacter', previously identified as Tetrasphaera, is the dominant polyphosphate accumulating lineage in EBPR wastewater treatment plants worldwide. *The ISME Journal* **16**, 1605-1616 (2022).

44 McDaniel, E. A. *et al*. Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics. *Msystems* **6**, e00474-00421 (2021).

45 Ishii, S. i. *et al*. A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. *Nature communications* **4**, 1601 (2013).

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

47 Roume, H. *et al*. Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. *npj Biofilms and Microbiomes* **1**, 1-11 (2015).

48 Scarborough, M. J., Myers, K. S., Donohue, T. J. & Noguera, D. R. Medium-Chain Fatty Acid Synthesis by "Candidatus Weimeria bifida" gen. nov., sp. nov., and "Candidatus Pseudoramibacter fermentans" sp. nov. *Applied and Environmental Microbiology* **86**, e02242-02219 (2020).

49 Bekiaris, P. S. & Klamt, S. Automatic construction of metabolic models with enzyme constraints. *BMC bioinformatics* **21**, 1-13 (2020).

50 Zimmermann, J., Kaleta, C. & Waschina, S. gapseq: Informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome biology* **22**, 1-35 (2021).

51 Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic acids research* **46**, 7542-7553 (2018).

52 Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B. & Lercher, M. J. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS computational biology* **13**, e1005494 (2017).

53 Ghatak, S., King, Z. A., Sastry, A. & Palsson, B. O. The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. *Nucleic acids research* **47**, 2446-2454 (2019).

54 Sauer, U. & Eikmanns, B. J. 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 microbiology reviews* **29**, 765-794 (2005).

55 Zhang, Y., Smallbone, L. A., DiCenzo, G. C., Morton, R. & Finan, T. M. Loss of malic enzymes leads to metabolic imbalance and altered levels of trehalose and putrescine in the bacterium Sinorhizobium meliloti. *BMC microbiology* **16**, 1-13 (2016).

56 Netzer, R. *et al*. Roles of pyruvate kinase and malic enzyme in Corynebacterium glutamicum for growth on carbon sources requiring gluconeogenesis. *Archives of microbiology* **182**, 354-363 (2004).

57 Hua, Q., Yang, C., Baba, T., Mori, H. & Shimizu, K. Responses of the Central Metabolism in Escherichia coli to Phosphoglucose Isomerase and Glucose-6-Phosphate Dehydrogenase Knockouts. *Journal of Bacteriology* **185**, 7053-7067 (2003).

58 Martín, H. G. *et al*. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature biotechnology* **24**, 1263-1269 (2006).

59 Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. *The ISME journal* **10**, 810-822 (2016).

60 da Silva, L. G. G. Life in changing environments: The intriguing cycles of Polyphosphate Accumulating Organisms PhD thesis, Delft University of Technology, (2021).

61 Miller, S. L. & Smith-Magowan, D. The thermodynamics of the Krebs cycle and related compounds. *Journal of Physical and Chemical Reference Data* **19**, 1049-1073 (1990).

62 Wimpenny, J. W. & Firth, A. Levels of nicotinamide adenine dinucleotide and reduced nicotinamide adenine dinucleotide in facultative bacteria and the effect of oxygen. *Journal of Bacteriology* **111**, 24-32 (1972).

63 Liu, Y., Landick, R. & Raman, S. A regulatory NADH/NAD+ redox biosensor for bacteria. *ACS Synthetic Biology* **8**, 264-273 (2019).

64 van der Rest, M. E., Frank, C. & Molenaar, D. Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of Escherichia coli. *Journal of Bacteriology* **182**, 6892-6899 (2000).

65 Chen, X., Alonso, A. P., Allen, D. K., Reed, J. L. & Shachar-Hill, Y. Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in E. coli. *Metabolic engineering* **13**, 38-48 (2011).

66 Wilmes, P. *et al*. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *The ISME journal* **2**, 853-864 (2008).

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

68 Vögeli, B. *et al*. Archaeal acetoacetyl-CoA thiolase/HMG-CoA

synthase complex channels the intermediate via a fused CoA-binding site. *Proceedings of the National Academy of Sciences* **115**, 3380-3385 (2018).

69 Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**, 504-507 (2001).

70 Lipson, D. A. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. *Frontiers in microbiology* **6**, 615 (2015).

4

CONDITIONAL FLUX BALANCE ANALYSIS (CFBA) TOOLBOX FOR PYTHON:

APPLICATION TO RESEARCH METABOLISM IN CYCLIC ENVIRONMENTS

Published as: Páez-Watson, T., Hernández Medina, R., Vellekoop, L., van Loosdrecht, M. C. & Wahl, S. A. Conditional Flux Balance Analysis (cFBA) Toolbox for python: application to research metabolism in cyclic environments. Bioinformatics Advances, vbae174 (2024).

Supplementary figures and tables are available in the online publication

Abstract

We present py_cFBA, a Python-based toolbox for conditional flux balance analysis (cFBA). Our toolbox allows for an easy implementation of cFBA models using a well-documented and modular approach and supports the generation of Systems Biology Markup Language (SBML) models. The toolbox is designed to be user-friendly, versatile, and freely available to non-commercial users, serving as a valuable resource for researchers predicting metabolic behaviour with resource allocation in dynamic-cyclic environments.

Availability: Extensive documentation, installation steps, tutorials and examples are available at https://tp-watson-python-cfba.readthedocs.io/en/ and the py_cFBA python package is available at https://pypi.org/project/py-cfba/

Introduction

Optimal resource allocation is a widespread theory used to study evolutionary trade-offs—inherent in metabolic processes [1-4]. The prevailing literature predominantly focuses on microorganisms thriving under stationary conditions [1-4]. While such conditions lend themselves to laboratory validation, in reality, microbial habitats in nature and most environmental biotechnology applications are far from static. Over evolutionary timescales, microorganisms have evolved diverse metabolic strategies to face a diverse array of dynamic environmental fluctuations [5,6]. These fluctuating environments and the metabolic strategies of organisms living therein can be studied with computational models. Modelling is key to establish fundamental principles governing evolutionary fitness.

Rügen et. al., [7] introduced a mathematical framework named conditional flux balance analysis (cFBA), designed to predict optimal resource allocation dynamics under fluctuating conditions. This framework has been applied to cyanobacteria [8] and polyphosphate accumulating organisms [9]. In both cases, temporal synthesis of storage polymers (*e.g.* glycogen, polyphosphate and polyhydroxyalkanoates) resulted as an emergent property of resource optimization in dynamic-cyclic scenarios. Nevertheless, reports on dynamic conditions remain sparse, primarily confined to these exemplar cases.

The cFBA method integrates stoichiometric modelling, dynamic Flux Balance Analysis (dFBA) with a final optimization through the whole simulation time, and resource allocation to study metabolic dynamics within cyclic environments. As such, it serves as a potent predictive tool for unveiling optimal metabolic strategies in ecosystems such as diurnal cycles, feast-famine dynamics, and aerobic-anaerobic transitions. Given the prevalence of such environmental conditions in nature, the field of microbial ecology could benefit from the application of cFBA. To date, there is no cFBA tool based on open-source platforms. The current implementation relies on adoption of complex metabolic models in MATLAB [7,8] with little documentation or simple examples for its application. Here, we present an easy-to-use Python toolbox for the application of cFBA (see Table 1 for a comparison of py_cFBA with other published models). This toolbox allows users to explore the boundaries of metabolic behaviour given a stoichiometric model, enzyme capacities, and a set of environmental conditions.

Table 1. *A comparison of cFBA with similar methods to research dynamic metabolism.*

Feature	**RBApy** [10]	Cycle Sync	cFBA	py_cFBA (this research)
Method	Resource Balance Analysis [11]	Cycle Sync [12]	Conditional FBA [7]	Conditional FBA [7]
Language	Python	Python	MATLAB*	Python
Toolbox for user implementation	+	-	-	+
Comprehensive documentation, tutorials and examples	+	-	-	+
Temporal flux analysis	+	+	+	+
Enzyme constraints	+	-	+	+
Study of organisms in cyclic environments	-	+	+	+
Capture temporal use of storage polymers	-	+	+	+

*Requires a license for its use.

Toolkit usage and user input

The cFBA toolkit is accessible as a Python package [https://pypi.org/project/py-cfba/]. It comprises a suite of functions enabling the construction of the cFBA model architecture and subsequent simulations. The required user input varies depending on the desired complexity of the analysis. For instance, to model dynamic cycling without any catalytic information, a stoichiometric matrix suffices. However, if enzyme capacities are to be included, numerical constants describing the relationships between reactions and their respective catalytic efficiencies (k_{cat}) are needed. The first step in this toolkit's pipeline is to generate a basic cFBA model structure, which is encoded into SBML. Subsequently, the SBML model is parsed into a linear programming problem. Detailed instructions on model generation and Systems Biology Markup Language (SBML) file creation can be found at [https://tp-watson-python-cfba.readthedocs.io/en/].

Methods and Implementation

Minimal Set of Constraints: Unlimited Catalytic Activity

The underlying metabolic model in cFBA is represented by a stoichiometric matrix (S), which represents the interplay of metabolites and reactions in a metabolic network. From the participating species (metabolites enzymes and biomass components), a subset is expressed as imbalanced (M). These species—typically enzymes, ribosomes, membranes, storage polymers, and substrates—exhibit explicit concentration changes over time, which are explicitly modelled as done with dFBA [13]. Conversely, the remainder of metabolites ($\bar{M}$) are presumed to remain in quasi-steady state, because their turnover rate is significantly faster than that of imbalanced species [7,8,14]. Users can simulate dynamic environmental changes (such as variations in light, substrate and oxygen) by constraining reactions with upper and lower bounds.

Each cFBA simulation is normalized to an initial amount of biomass (typically 1 gram dry weight (g_{DW})). Biomass is not modelled as an independent metabolite with its corresponding biosynthesis reaction, but rather defined as the weighed sum of all components in M at each time point (all imbalanced metabolites). To normalize the initial time point of the simulation to 1 g_{DW}, Eq. 1 is employed.

$$w^T M^{t=0} = 1 \qquad \text{(Eq. 1)}^7$$

Here, w^T represents the transpose of a matrix containing the molecular weights of each imbalanced metabolite in *M*. The cyclic behaviour of cFBA is achieved by enforcing identical relative amounts of imbalanced metabolites at both the beginning and end of the simulation (Eq. 2).

$$M^{t_{end}} = \mu\, M^{t=0} \qquad \text{(Eq. 2)}\ ^{7}$$

μ represents the balanced growth of the system. These constraints represent a quadratic programming problem which becomes linear for each value of μ. The objective of the cFBA model is to achieve the highest multiplication factor (μ) using a binary search algorithm. Numerically-stable solvers with high numeric precision, such as Gurobi, are recommended since complex models may lead to ill-conditioned problems [8]. The implementation of this method uses the OPTLANG library in Python and the solvers supported and their limitations have been described [15].

Cellular Limits and Requirements on Metabolites: Quotas

By default, the synthesis of imbalanced metabolites is not enforced (apart from maintaining the relation in Eq. 1). Minimal cellular requirements can be enforced by setting quota definitions (minimal concentration constraints) For instance, Rügen, et al. [7] employed quota compounds to establish minimal thresholds for inorganic ions, cell wall constituents, lipids, DNA, and non-catalytic proteins, relative to biomass. Expanding upon the quota definitions utilized by Rügen, et al. [7] and Reimers, et al. [8], our toolkit enables users to define exact, minimum, and maximum quota constraints at any time point during the simulation. This facilitates the capture of dynamic behaviours in simulated environmental conditions.

Coupling metabolism to protein allocation: Enzyme activity based on enzyme amounts

Imbalanced metabolites can also act as catalysts of specific reactions. The relation between the metabolite and the reaction it catalyses is indicated in the capacity matrices (A and B in Eq. 3) which denote the associations between catalysed reactions and the k_{cat} values of each catalysed reaction.

$$A_{cap}{}^{t} v_r{}^{t} \leq B_{cap} \cdot M^{t} \qquad \text{(Eq. 3)}\ ^{7}$$

$$v_r{}^{t} \leq M_e{}^{t} \cdot k^{r}_{cat_e} \qquad \text{(Eq. 4)}\ ^{7}$$

Equation 3 sets a reaction *r* catalysed by an imbalanced metabolite M_e to be constrained by its upper limit following the relation in Eq. 4. It is noteworthy that this defines an upper boundary to the reaction, not an exact value. Additionally, if a reversible reaction is catalysed by an imbalanced specie, both directions of the reaction must be accounted for in the S matrix.

Standard genome-reconstructed metabolic models typically include storage polymers such as glycogen as part of biomass components. Following the approach of Ofaim et al., [16] we allow for the explicit separation of storage metabolites from that of biomass (also referred to as *lean* biomass). This allows the independent accumulation and utilization of said polymers in various simulations irrespective of biomass composition (independent from Eq. 2).

An illustrative example of cFBA implementation for a toy model of a minimal cell is presented in Figure 1. The system comprises one balanced metabolite (named 'intermediate') and three imbalanced species: storage, enzymes and biomass. The reactions for substrate uptake and biomass synthesis are catalysed by the species 'enzymes' each with a distinct k_{cat} value. The simulation incorporates a dynamic component wherein substrate is only available until 2 hours in the simulation. No quota compounds are defined, and the metabolite '*biomass*' solely contributes to w^T. The cFBA simulation results in an early use of resources (substrate) into enzyme biosynthesis to reach the maximum catalytic capacity at the third time-point (1 hour). After this point, the system produces biomass at a balanced rate (optimizing enzyme usage), making temporal use of storage to allow this steady rate of biomass synthesis. Variations and step-by-step examples of this model implementation are available at [https://tp-watson-python-cfba.readthedocs.io/en].

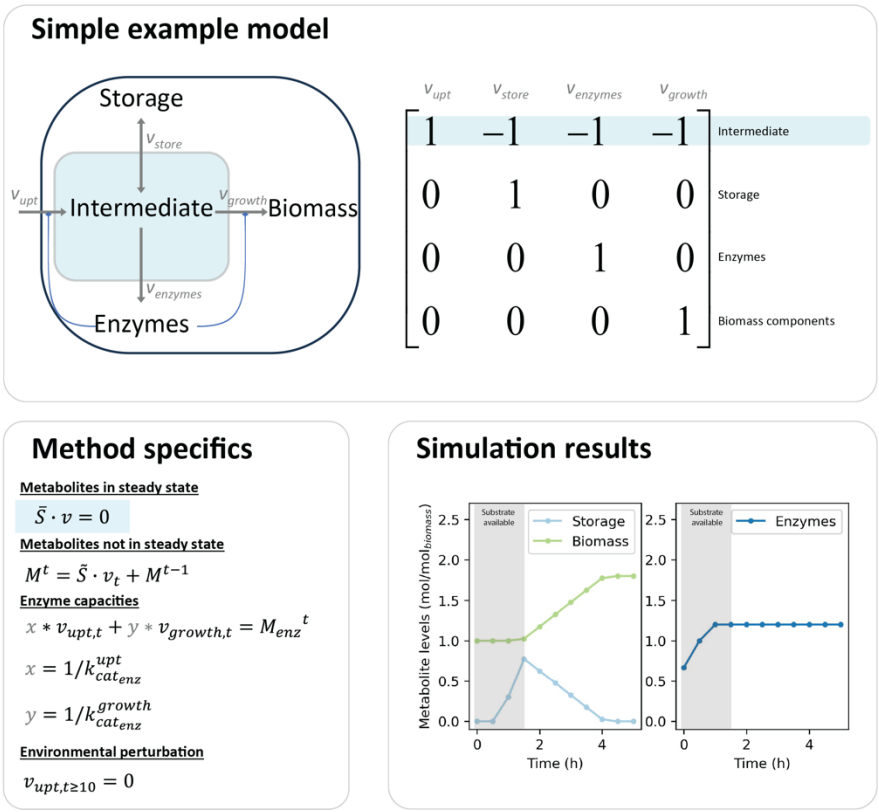


Figure 1. *Basic simulation of a toy-model using the cFBA Python Toolbox. Inputs required for this model are a stoichiometric matrix, definition of balanced and imbalanced metabolites, and enzyme capacities. The simulation includes an active feed during the first 10 units of time after which there is no longer substrate simulating a feast-famine condition. Specifics and step-by-step implementation of this model in the cFBA python toolbox are available at [https://tp-watson-python-cfba.readthedocs.io/en].*

Conclusion

The Python cFBA toolkit facilitates the study of metabolic dynamics in cyclic environments. We included clear documentation and examples for a fast familiarization to resource allocation strategies in dynamic conditions. Two considerations are of note: numerical challenges may require specialized solvers, and further developments are needed to address complex biological systems such as non-optimal balanced growth strategies or microbial communities. Notwithstanding, the toolkit represents a significant advancement in systems biology, offering researchers a powerful tool to explore metabolic behaviour in dynamic-cyclic environments.

References

1 Molenaar, D., Van Berlo, R., De Ridder, D. & Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology* **5**, 323 (2009).

2 Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences* **110**, 10039-10044 (2013).

3 Vázquez-Laslop, N. & Mankin, A. S. Protein accounting in the cellular economy. *Cell* **157**, 529-531 (2014).

4 Elsemman, I. E. *et al*. Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nature communications* **13**, 801 (2022).

5 Bajic, D. & Sanchez, A. The ecology and evolution of microbial metabolic strategies. *Current opinion in biotechnology* **62**, 123-128 (2020).

6 Caetano, R., Ispolatov, Y. & Doebeli, M. Evolution of diversity in metabolic strategies. *Elife* **10**, e67764 (2021).

7 Rügen, M., Bockmayr, A. & Steuer, R. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Scientific reports* **5**, 15247 (2015).

8 Reimers, A.-M., Knoop, H., Bockmayr, A. & Steuer, R. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences* **114**, E6457-E6465 (2017).

9 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: Application to polyphosphate accumulating organisms. *Water Research* **228**, 119365 (2023).

10 Bulović, A. *et al*. Automated generation of bacterial resource allocation models. *Metabolic engineering* **55**, 12-22 (2019).

11 Goelzer, A. *et al*. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic engineering* **32**, 232-243 (2015).

12 Sarkar, D., Mueller, T. J., Liu, D., Pakrasi, H. B. & Maranas, C. D. A diurnal flux balance model of Synechocystis sp. PCC 6803 metabolism. *PLoS computational biology* **15**, e1006692 (2019).

13 Mahadevan, R., Edwards, J. S. & Doyle III, F. J. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical journal* **83**, 1331-1340 (2002).

14 Waldherr, S., Oyarzún, D. A. & Bockmayr, A. Dynamic optimization of metabolic networks coupled with gene expression. *Journal of theoretical biology* **365**, 469-485 (2015).

15 Sonnenschein, N. *et al*. (ed Zenodo) (2018).

16 Ofaim, S., Sulheim, S., Almaas, E., Sher, D. & Segrè, D. Dynamic allocation of carbon storage and nutrient-dependent exudation in

a revised genome-scale model of Prochlorococcus. *Frontiers in Genetics* **12**, 586293 (2021).

5

PREDICTING THE IMPACT OF TEMPERATURE ON METABOLIC FLUXES USING RESOURCE ALLOCATION MODELLING:

APPLICATION TO POLYPHOSPHATE ACCUMULATING ORGANISMS.

Published as: Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A.
Predicting the impact of temperature on metabolic fluxes
using resource allocation modelling:
Application to polyphosphate accumulating organisms.
Water Research 228, 119365 (2023)

Supplementary figures and tables are available
in the online publication

Abstract

The understanding of microbial communities and the biological regulation of its members is crucial for implementation of novel technologies using microbial ecology. One poorly understood metabolic principle of microbial communities is resource allocation and biosynthesis. Resource allocation theory in polyphosphate accumulating organisms (PAOs) is limited as a result of their slow imposed growth rate (typical sludge retention times of at least 4 days) and limitations to quantify changes in biomass components over a 6 hour cycle (less than 10 % of their growth). As a result, there is no direct evidence supporting that biosynthesis is an exclusive aerobic process in PAOs that alternate continuously between anaerobic and aerobic phases. Here, we apply resource allocation metabolic flux analysis to study the optimal phenotype of PAOs over a temperature range of 4 °C to 20 °C. The model applied in this research allowed to identify optimal metabolic strategies in a core metabolic model with limited constraints based on biological principles. The addition of a constraint limiting biomass synthesis to be an exclusive aerobic process changed the metabolic behaviour and improved the predictability of the model over the studied temperature range by closing the gap between prediction and experimental findings. The results validate the assumption of limited anaerobic biosynthesis in PAOs, specifically "*Candidatus* Accumulibacter" related species. Interestingly, the predicted growth yield was lower, suggesting that there are mechanistic barriers for anaerobic growth not yet understood nor reflected in the current models of PAOs. Moreover, we identified strategies of resource allocation applied by PAOs at different temperatures because of the decreased catalytic efficiencies of their biochemical reactions. Understanding resource allocation is paramount in the study of PAOs and their currently unknown complex metabolic regulation, and metabolic modelling based on biological first principles provide a useful tool to develop a mechanistic understanding.

Introduction

Biotechnological applications, especially in environmental engineering, strongly depend on the function and stability of microorganisms that interact with each other and with the environment in dynamically changing communities [1]. Further development of tools to study and predict these microbial communities holds the key to improving and expanding the plethora of their applications [2]. Wastewater treatment is a field of application that, for decades, has relied on the use of microbial communities and modelling to better predict and control the emerging dynamics of such environments [3]. Extensive experimental knowledge on these systems has been obtained using long-term lab-scale experiments [4-7]. While significant experimental progress was achieved and there is a growing availability of high-throughput data generating techniques (i.e. genomics [8], transcriptomics [9,10] and proteomics [11]), a full characterization and especially mechanistic understanding of these communities remains a challenging task in ecology [12]. Modelling enables to test different hypothesis of mechanisms and especially whole cell models are crucial to integrate observations, unravel biological principles and predict functions or conditions that are currently inaccessible via experimental approaches only [13].

The enhanced biological phosphorous removal (EBPR) is one of the most studied microbial processes in wastewater treatment and decades of research have unravelled the main biological transformations of this complex engineered ecosystem [14-16]. In the EBPR system, microorganisms cycle between anaerobic and aerobic phases leading to the enrichment and proliferation of polyphosphate accumulating organisms (PAOs). In this community, one of the most studied bacteria is "*Candidatus* Accumulibacter" (hereafter referred to as Accumulibacter) [15]. Accumulibacter thrives in a typical EBPR cycle by a complex metabolic strategy involving the cycling of storage polymers: polyphosphate, glycogen and polyhydroxyalkanoates (PHAs) [16].

So far, metabolic modelling approaches for PAOs have been developed and are used with the aim to maintaining stable operations during EBPR processes and/or predicting better conditions favouring simultaneous carbon, phosphorous and nitrogen removal [3]. Early models such as the ASM3 or TUDP have shown to translate well with process observations, however, require extensive recalibration depending on plant properties and environmental conditions. For example, more recent model developments include the effect of putative PAO competitors [17,18], fermentative PAOs [19], storage levels' effects [20], etc. Nevertheless, most current

models rely on pre-defined metabolic strategies and consequently reflect the already existing, though not necessarily complete, scientific knowledge on metabolic functions of this ecosystem. Although useful, these features limit the mechanistic understanding of the biological principles governing microbial communities under dynamic conditions.

One of the pre-defined microbial strategies commonly applied in the EBPR models is the assumption of growth being limited to the aerobic phase. Already very early PAO studies suggested that biosynthesis occurs only in the aerobic phase of EBPR systems at expenses of PHA degradation [21,22]. Throughout the years, this assumption has been adopted into metabolic models. Interestingly, this assumption has not been verified in-depth besides the observations on the consumption of certain nutrients (e.g. $NH4^+$) linked to aerobic growth [9]. Many bacterial species that survive under both anaerobic and aerobic conditions grow anaerobically while fermenting substrate [23,24], hence the validity of this assumption (aerobic biosynthesis) in PAOs is yet to be confirmed.

Due to the slow growth of organisms in EBPR processes, it has been difficult to estimate protein turnover rates and to precisely calculate protein synthesis in different phases of a cycle. A typical EBPR lab-scale setting consist of daily cycles of 6 hours with an imposed SRT of 8 days. Thus, theoretically the biomass in the system should renew its proteome in the range of around 32 cycles, meaning that in one cycle the newly produced proteins would account for only 3 % of the proteome, and the putative contribution of anaerobic biosynthesis would be even lower. Measuring such small differences in microbial communities is technically challenging considering the inhomogeneity of the culture as well as current limitations such as number of mass spectra acquisition per time [25], biased extraction methods to soluble proteins [26,27], amongst others. Proteomics studies on EBPR sludge from Wilmes, et al. [28] and Wexler, et al. [29] fell short on quantitatively identifying these changes, exemplifying the complications of studying protein synthesis in slow growing systems such as PAOs.

Data from transcriptomic studies, however, seem to indicate a major trend towards aerobic protein biosynthesis. Time series meta transcriptomic data from a highly enriched Accumulibacter culture showed different clusters of expression throughout the EBPR cycle [10]. The largest number of transcripts from their study showed trends of transcription during the aerobic phase (identified as aerobic pattern, redox transition and low phosphate patterns). Although the link from transcription to protein synthesis is not always direct, these results highly suggest that there is a regulation favouring protein synthesis to occur in the aerobic phase. This hypothesis could be further explored experimentally with the use of

isotopically labelled acetate fed to a PAOs enrichment, such as the experiments done by [30], to identify the fate of ^{13}C anaerobically and aerobically over one or multiple cycles, however up to date such evidence is lacking.

With no final experimental evidence, model-based studies considering biosynthesis and resource allocation could be applied to develop an understanding of a putative synthesis of biomass components and furthermore quantify the putative benefit of anaerobic growth. Constraint-based models such as flux balance analysis (FBA) integrating principles from resource allocation represent an opportunity to test a hypothesis regarding biosynthesis during a cyclic, dynamic system [31,32]. Conditional flux balance analysis (cFBA) is a metabolic modelling tool that was originally developed to predict the metabolism of cyanobacteria under dynamic day/night cycles with strong dependency on resource allocation [32]. The characteristics of the cFBA framework and the similarities of these cyclic conditions with those of EBPR make cFBA a suitable method to apply in the context of studying PAOs. Recently, the cFBA method was applied to predict optimal strategies under dynamic environmental conditions encountered during EBPR [33]. Depending on the environmental constraints, different optimal strategies, i.e. organisms accumulating polyphosphate (PAOs), glycogen (GAOs), polyhydroxyalkanoates (PHA-Os) and heterotrophs were predicted. The optimizations resulted in metabolic strategies comparable to those typically observed for Accumulibacter. While the general behaviour was correctly predicted, there were quantitative mismatches, suggesting the need for further model development probably beyond parameter calibration. Because the model is strongly shaped by the relation between enzymatic activities and resource allocation, there is potential for expansion in this front.

Enzymatic activities are strongly influenced by temperature and consequently temperature plays a crucial role in shaping the metabolism of microorganisms [34]. Implementation of temperature dependencies on metabolic models have shed light on basic biological principles such as a linear relationship between growth rate and ribosome content [35,36] and optimal proteome allocation as a function of temperature [37]. However, these principles result from models for organisms such as *Escherichia coli* at steady-state and may not apply to microbial communities under dynamic conditions. On the other hand, there are extensive experimental studies on the effect of temperature on PAOs metabolism [38-40]. The current mechanistic understanding of PAOs metabolism with relation to temperatures could be combined with metabolic models such as cFBA to identify metabolic principles governing growth and resource allocation. Especially, at low

temperatures, the efficient use of available resources like enzyme capacity are assumed to be crucial for evolutionary competitiveness.

In the current study, we combined the cFBA modelling framework with temperature dependency for PAOs to identify metabolic principles regarding resource allocation. We compared the complete model with and without a constraint on anaerobic biomass synthesis. Our results further validate the assumption on biomass being limited to the aerobic sector of EBPR and shed light on previously unexplored putative regulation on protein biosynthesis.

Results

Reference simulation: EBPR cycle with *unconstrained growth* at 20°C

We first investigated the predicted metabolic response of the model with *unconstrained growth* (Figure 1.A) in a typical EBPR system at 20 °C and used it as the reference simulation for further comparisons. A total cycle length of 5 hours was applied: 1.5 hours anaerobic and 3.5 hours aerobic. Acetate was enforced to be consumed within 30 minutes of the anaerobic period to simulate the competition for substrate and obtain the competitive strategy of organisms enriched under this regime [33]. The minimal initial amounts (quota) of storage polymers (polyphosphate, glycogen and PHA) per biomass (mol / g_{DW}) were introduced in the model based on data published by Acevedo, et al. [7]. Using these conditions, the reference metabolic response of PAOs simulated was obtained and compared to the experimental profiles from Acevedo, et al. [7] (Figure 1B).

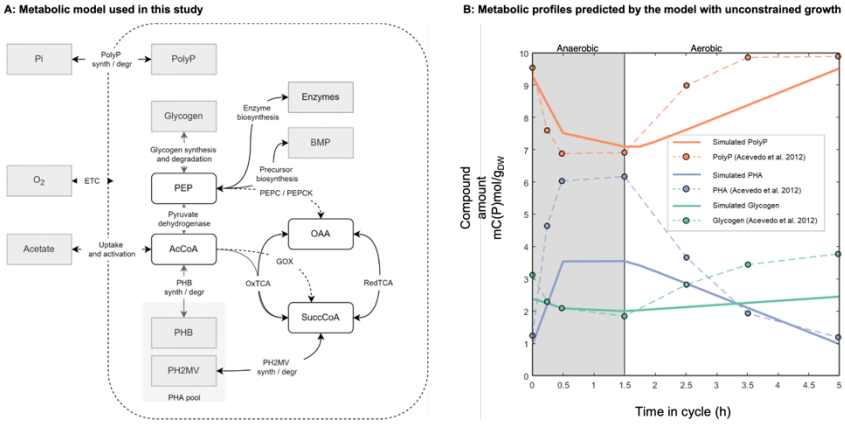


***Figure 1**. A) Schematic representation of the reaction network used for the cFBA simulations. Compounds enclosed by grey boxes are dynamic compounds, that can accumulate resp. decrease over time. All other compounds were assumed to reach quasi steady-state. Each reaction is catalysed by a specific enzyme available at a calculated, dynamic concentration. B) Solid lines: predicted metabolite concentrations of PAOs during an EBPR cycle at 20 °C with unconstrained growth [33]. Circles and dotted lines: experimental data from an EBPR system retrieved from Acevedo, et al. [7].*

The simulation reproduced complete acetate consumption and accumulation as PHA (Figure 1B), while polyphosphate was degraded to provide the required ATP for the uptake and activation of acetate and anaerobic maintenance. Reducing equivalents (modelled as NADH) for PHA accumulation were provided by glycogen degradation and a fraction of acetate being channelled through the TCA cycle. When oxygen became available (aerobic phase, with activated ETC), the intracellular PHAs were degraded and the resources from this degradation were used to restore the glycogen and polyphosphate pool. Such metabolic strategy observed in the cycling of polymers is typical of Accumulibacter [7,15,16].

Although the model with *unconstrained growth* predicted typical profiles for Accumulibacter, there are some quantitative mismatches with the compared experimental data set (Figure 1.B), mainly regarding the amount of glycogen and PHA used/accumulated within a cycle. This mismatch between prediction and experimental data results in part from a higher contribution of the NADH generated in the TCA cycle leading to lower required flux through glycolysis (Figure 3, explored in the following subsections). We additionally identified that biosynthesis was active anaerobically in this simulation (mainly synthesising enzymes). To evaluate if the observed mismatches were specific to the chosen condition (EBPR at 20 °C) or based on a systematic feature of our modelling approach, we explored a broad range of different temperatures.

Simulations for different temperatures

Reported, experimentally determined values of temperature coefficients (θ) in PAOs range from 1.03 for polyphosphate synthesis to up to 1.13 for PHA degradation (See Table S2 and Figure S1). For other reactions of the metabolic model (TCA cycle, ETC, protein biosynthesis, ribosome synthesis, etc.) no specific θ values for PAOs were found in literature. Therefore, these were assumed to be the same for all uncharacterized reactions and we analysed the impact of this value varying within a range of the previous reported values.

A typical EBPR cycle was simulated ranging between 4 °C to 20 °C. For all simulations, a shift in the metabolic strategy towards lower temperatures was observed (Figure S2). The tested θ value of uncharacterized reactions did not

affect the metabolic shift but affected the temperature at which the shift was observed (Figure S2). For example, the simulations indicated a metabolic shift of the system towards colder temperatures below either 13 °C or 18 °C when assuming θ values for uncharacterized reactions with θ=1.05 or 1.15 respectively. From here on, a value of 1.05 was used for the reactions where no experimental value was available.

Influence of the growth constraint on simulations

With temperature dependencies implemented in the metabolic model, we simulated a typical EBPR cycle of PAOs over a temperature range from 4 °C to 20 °C with no constraints on growth. The results show a decrease in growth yield with a decrease in temperature following two different exponential regions (Figure 2.A: *Unconstrained growth*). In these simulations, the model employed slightly a different metabolic strategy towards lower temperatures resulting in a larger use of glycogen and polyphosphate over the cycle (Figure 2.B: *Unconstrained growth*). We observed that in the studied temperature range, resources were destined for growth in both the anaerobic and aerobic phases. Interestingly, the lower the temperature, the higher contribution anaerobic growth had on the system, especially destined towards enzyme synthesis reactions (reaching up to 25 % of all biosynthesis at temperatures below 10 °C).

Although there is no conclusive proof that growth is an exclusive aerobic process in PAOs, we analysed the impact of constraining growth to the aerobic phase only. Particularly, a constraint was introduced to block biosynthesis reactions during anaerobic conditions. The model with these new constraints was used to simulate the previous conditions (typical EBPR cycle of PAOs over a temperature range from 4 °C to 20 °C). The resulting growth yields were lower than those of the *unconstrained growth* model and similarly decreased towards lower temperatures (Figure 2.A: *aerobic growth* model). This decrease, however, followed a distinct exponential Arrhenius-like curve. Additionally, these simulations resulted in higher levels of glycogen use at 20 °C then the model with no growth constraints (>3 fold higher) better replicating the results from Acevedo, et al. [7]. Towards lower temperatures, less glycogen and more polyphosphate was used in these simulations (Figure 2.B: *Aerobic growth*).

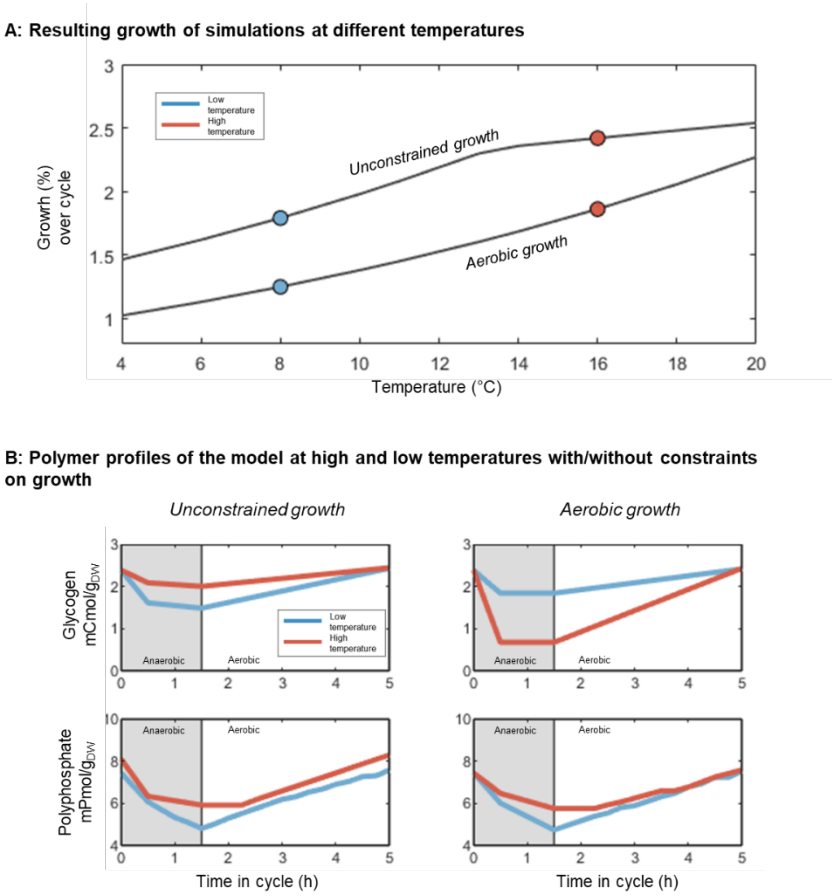


Figure 2. *Predicted biomass yields and intracellular storage pools during an EBPR cycle at different temperatures. (A) Biomass increase at the end of a cycle. A decrease in temperature results in a decrease of biomass synthesis during the cycle of both the unconstrained growth and the aerobic growth model. The red and blue dots indicate simulations that are displayed w.r.t. glycogen and polyphosphate profiles over time (B). (B) Glycogen and polyphosphate over time in an EBPR cycle at 8 °C (blue line) and 16 °C (red line). Left panel represents the profile for the unconstrained growth model and the right panel represents the profiles of the aerobic growth model. For a representation of polymer changes over the entire temperature range, see Figure S3.*

Comparison and analysis of resource allocation strategies

The different predictions obtained over the studied temperature range are compared in terms of their resource allocation. Especially, we focus on the allocation of electrons (in form of NADH) and energy (in the form of ATP) by analysing reactions using or generating these metabolites. For this, we analysed the generation and consumption of NADH (Figure 3) and ATP (Figure 4) during the anaerobic phase of each simulation with respect to the amount of acetate consumed. Note that the metabolism of PAOs is strongly constrained in the

anaerobic phase, hence we primarily focus on the anaerobic phase of each simulation.

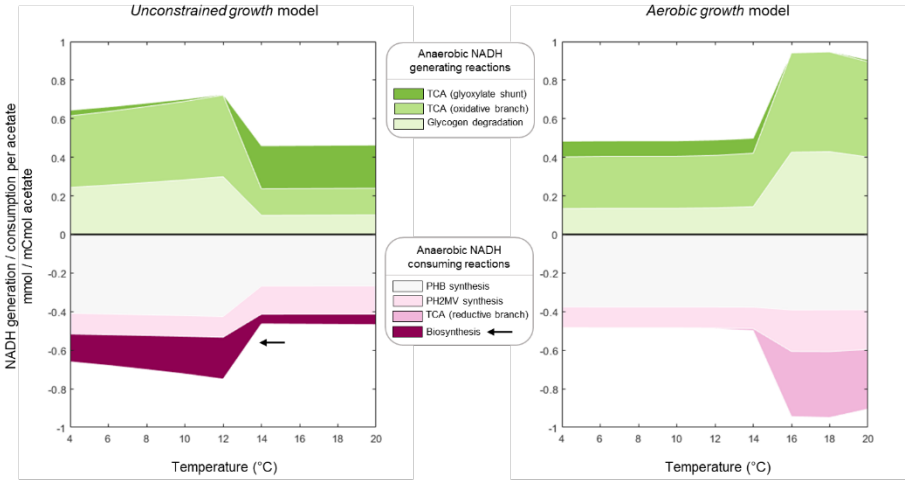


Figure 3. Balance and reactions involved in the NADH generation (green) and consumption (magenta) during the anaerobic phase at different temperatures of the unconstrained growth (left panel) model and the aerobic growth (right panel) model. Black arrow indicates the use of NADH for biosynthetic reactions showcasing the main difference in the structure between both models. Biosynthesis comprised all reactions synthetizing enzymes, ribosomes and biomass precursors. For the specific reaction stoichiometry, please see Table S1. Abbreviations: tricarboxylic acid cycle (TCA), poly-hydroxy-butyrate (PHB), poly-hydroxy-2-methylvalerate (PH2MV).

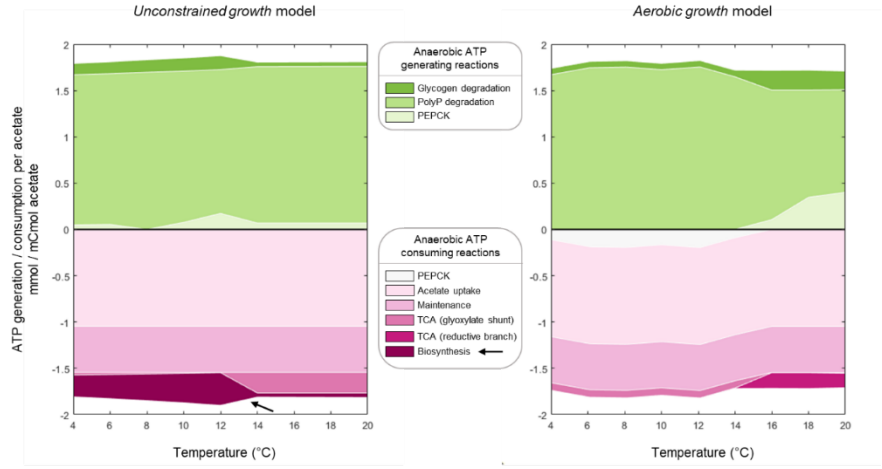


Figure 4. Balance and reactions involved in the ATP generation (green) and consumption (magenta) during the anaerobic phase of the PAOs simulations at different temperatures of the unconstrained growth (left panel) model and the aerobic growth(right panel) model. Black arrow indicates the use of ATP for biosynthetic reactions showcasing the main difference in the structure between both

models. Biosynthesis comprised all reactions synthetizing enzymes, ribosomes and biomass precursors. For the specific reaction stoichiometry, please see Table S1. Abbreviations: polyphosphate (PolyP), phosphoenolpyruvate carboxykinase (PEPCK), tricarboxylic acid cycle (TCA).

The model with no constraints on anaerobic growth predicted that, anaerobically, NADH and ATP were allocated towards biosynthesis of enzymes, and this allocation grew larger at lower temperatures (Figures 3 and 4: unconstrained growth model). The increased need for these resources for biosynthesis was met with an increased flux of glycogen degradation (up to 280 % more turnover) supplying both the required ATP and NADH in this model, which fits the observations of a larger turnover of glycogen towards lower temperatures (Figure 2: *Unconstrained growth*).

Predictions of the model allowing only aerobic growth indicated an opposite trend than that of the unconstrained model. Namely, the turnover of NADH was larger at higher temperatures (Figure 3) resulting in a larger turnover of glycogen as described in the previous section (Figure 2). Interestingly, the larger amounts of glycogen degraded at higher temperatures led to an overproduction of electrons than required only for acetate uptake, leading to a higher accumulation of PH2MV as a sink of these electrons. Additionally, the source of ATP produced anaerobically in this model shifted from a nearly full contribution of polyphosphate at lower temperatures towards a shared contribution of polyphosphate degradation, glycogen degradation and PEPCK at higher temperatures (Figure 4: *Aerobic growth* model*)*. Note that PEPCK is a reversible reaction, and as such acts as both an ATP sink and source at lower and higher temperatures respectively.

Discussion

The model with only aerobic growth predicts better polymer use of PAOs

The metabolic shift predicted by the *unconstrained growth* model indicated that at lower temperatures larger amounts of glycogen were being degraded than at higher temperatures (Figure 2.B: *unconstrained growth*). This result was surprising and opposite to what has been observed experimentally and described in literature. Brdjanovic, et al. [39] exposed PAOs enrichments at short term temperature changes and observed that at lower temperatures less glycogen was used overall, contradictory with the predictions of this model. Further, Brdjanovic, et al. [38] confirmed the same experimental observations in PAOs enrichments on

longer term temperature effects. This same behaviour has been reproduced by using kinetic models [17,41] and even proven to hold true for Glycogen Accumulating Organisms (GAOs) under similar conditions [42].

On the other hand, the constraint limiting biosynthesis to be an exclusive aerobic process resulted in predictions more in line with the described literature over the studied temperature range [17,38,39,43]. *I.e.* the amount of glycogen used anaerobically in a cycle increased with an increase in temperature (Figure 2, 3 and 4: *aerobic* growth). Further, the overall amount of glycogen used (and consequently PHA accumulated) at 20 °C was also larger than in the *unconstrained growth* model, resulting in improved model predictions when compared to the experimental dataset obtained by Acevedo, et al. [7] (Figure 1). Not only the predictions of the *aerobic growth* model approximated better the results from Acevedo, et al. [7], but also the anaerobic stoichiometric yields for glycogen, PHA and polyphosphate fit better within the observed yields from multiple PAOs enrichments at 20 °C (summarized in Welles, et al. [44]) (Figure 5) further supporting the validity of the added biological constraint of this model.

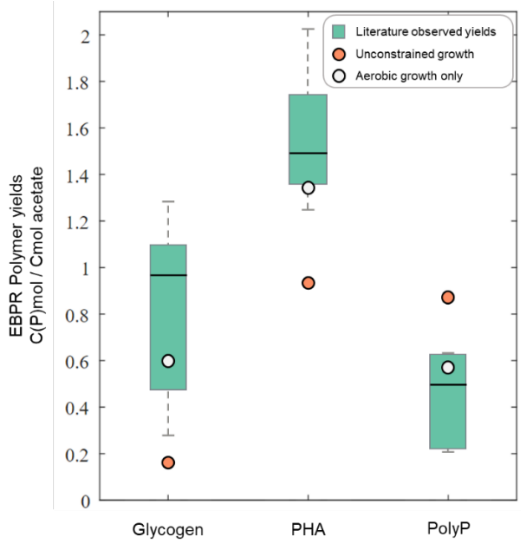


Figure 5. *Anaerobic yields of Glycogen (Cmol glycogen / Cmol acetate), PHA (Cmol PHA / Cmol acetate) and polyphosphate (Pmol / Cmol acetate) on acetate over a typical EBPR cycle at 20 °C. The figure indicates a (green) box plot sumarizing typical literature values for PAOs enrichments from several research groups (summarized in Welles, et al. [44]), and the predictions at 20 °C from the (orange) unconstrained growth model and (light grey) aerobic growth model.*

Other than the direct effect observed on polymer cycling, the model also improved in fulfilling a basic principle of systems biology: linear relationship between growth rate and ribosome content. Such a biological principle is

paramount for biological systems at balanced growth [35,36], and should not be an exception for PAOs. The simulations with no growth constraints at different temperatures showed no such linear relation (Figure S4 A & B). However, once the constraint on limiting anaerobic biosynthesis was introduced, a linear relation between growth rate and ribosome content emerged (Figure S4 C & D). We note that recently Mairet, et al. [37] showed that at higher temperatures, this linear relationship breaks, nevertheless the temperature range of this study is below such threshold.

The different predictions from both models arose as a consequence of resource limitation in a system that was tightly constrained by its catalytic capacities [32]. At lower temperatures, the decrease in catalytic efficiency of each metabolic process resulted in lower flux capacities that could only be resolved by either producing larger enzymatic levels or adopting a different metabolic strategy. Such limitations in energy metabolism are known to strongly shape the proteome and metabolic strategy of microorganisms [45]. This was clearly observed in the *unconstrained growth* model, when at lower temperatures there was an increased flux of biosynthesis during the anaerobic phase (*Figures 3* and *4*) in order to maintain metabolic fluxes high. These higher fluxes also resulted in higher growth yields of the unconstrained growth model as compared to the growth constrained model. Such biosynthetic fluxes were balanced with increased glycogen degradation that generated the required NADH, ATP and metabolic precursors. As these results strongly contradict what has been observed experimentally, we conclude that indeed anaerobic biosynthesis is severely limited in PAOs.

The aforementioned discussed results highlight the validity of the newly introduced constraint as a general biological principle that could apply to PAOs such as Accumulibacter, but might even be generalized to organisms that are adapted to live under dynamic anaerobic/aerobic environments such as those encountered in EBPR or estuary sediments. Next, we attempt to give a biological meaning to the introduced constraint on limiting biomass synthesis and hypothesise on the possible regulation behind it.

Putative regulation of biosynthesis under dynamic conditions

Biomass synthesis in PAOs has been commonly assumed to be limited to only the aerobic phase [21,22]. However, these assumptions have never been proven experimentally. Since anaerobic growth is very common among bacteria it cannot simply be assumed as non-existing in PAOs.

We suggest that the limit on anaerobic biosynthesis is likely caused due to a dynamic shift in the energetic and redox state over an EBPR cycle. The presence of an external electron acceptor (*i.e.* oxygen) has been proven to strongly affect the redox state of cells [46]. That is, anaerobically the $NADH/NAD^+$ ratio increases, limiting reactions that are near equilibrium in the cell. Although not studied in depth during EBPR we hypothesise that this ratio is dynamically changing over a cycle (Zhao, et al. [47] showed direct measurements of this ratio, although this methodology has not been extensively proven). Similarly, we hypothesise that polyphosphate degradation could be initiated anaerobically by thermodynamic control caused by a shift in the ATP/AMP balance in the cell (hypothesised in early PAOs research by Comeau et al. (1996)) as a combined consequence of fast acetate uptake and the changed redox state. However, the synthesis and polymerization of proteins requires a relatively low ATP/AMP ratio, being opposite to that required for polyphosphate degradation [48], making both physiological processes to be thermodynamically opposed. More research is required in the dynamics of the redox and energetic state of PAOs to further understand physiological mechanisms of organisms living under EBPR like conditions.

First principles modelling approach to predict PAOs metabolism

The dynamic resource allocation modelling approach applied resulted in typical polymeric profiles that PAOs exhibit in a EBPR cycle (Figure 1) without the need of parameter calibration or predefinition of metabolic strategies. Thus, the employed method could be used not only to understand the environmental selection on PAOs, GAOs and PHA-AOs as was done by Guedes da Silva, et al. [33], but to test basic metabolic principles shaping optimality in dynamic conditions. This modelling approach represents an alternative to traditional used modelling approaches that rely on experimental yields and kinetics [3,17-20], but is however not intended to be used as an indicator of EBPR process control or performance.

The applied model in this research could be used as a tool to expand our current understanding of redox and energetic state of bacterial cells under dynamic conditions. For example, here we identified a potential shift in the sources of NADH and ATP (Figure 3 and 4 respectively) as a function of temperature. Further studying these individual contributions, this model could explain the reason behind the large variation in P/C ratio obtained by different research groups when studying PAOs enrichments (summarized in da Silva, et al. [49]). This model is not intended to be used for monitoring waste water treatment plants, but rather to gather fundamental knowledge that could help improve our mechanistic understanding of organisms commonly encountered in said processes.

Conclusions

From this research, we can conclude that:

- Resource allocation theory delivers a strong framework to analyse metabolic processes in microbial communities typically found in wastewater treatment systems
- Integrating temperature into a FBA models of organisms living in dynamic conditions allows for deeper understanding of resource allocation limitations of cells.
- Based on the resource allocation theory results, the biosynthetic routes of Accumulibacter are limited to the aerobic phase of the EBPR cycle.

Materials and methods

Model description

For modelling the metabolism and energy allocation of PAOs, a constrained-based approach named conditional flux balance analysis (cFBA) was used [32]. The model and parameters for PAOs was obtained from Guedes da Silva, et al. [33].

The model consists of 29 metabolites connected by 36 reactions (Figure 1A) represented in a stoichiometric matrix **S**. All network reactions and stoichiometry are detailed in *Table S1*. Different to conventional approaches, only a subset of metabolites (7 metabolites) are considered in steady state by the relation:

$$\bar{\mathbf{S}} \,.\, v_t = 0\,,$$

where $\bar{\mathbf{S}}$ denotes the stoichiometric matrix subset of steady state metabolites and v_t denotes the fluxes of each reaction at time interval t. The remaining 22 metabolites are dynamic, i.e. are allowed to accumulate or deplete over time (polymers, enzymes and biomass precursors) and their molar amount is updated at each time point by the following relation:

$$M_t = \tilde{\mathbf{S}} \,.\, v_t + M_{t-1}$$

M_t is a vector that indicates the molar amount of the non-steady state metabolites at time interval *t*, and $\tilde{\mathbf{S}}$ denotes the stoichiometric matrix subset of the dynamic metabolites (not in steady state).

Biomass growth and composition

Growth of the system at the end of the simulation is described as the overall fold-change (α) of the initial metabolite composition (M_0):

$$M_{end} = \alpha * M_0$$

This relation introduces the possibility to simulate a system in a cycle, as the proportions of non-steady state metabolites need to be maintained only at the end of the simulation. The values of M_0 are an outcome of the simulation in order to achieve maximal growth but limited to the total sum of 1 gram dry weight (g_{DW}) by the relation:

$$w^T . M_0 = 1,$$

where w^T is a transposed vector containing the weights that different components have on 1 g_{DW} of biomass. The values in w^T were defined by Guedes da Silva, et al. [33] for glycogen, polyphosphate and PHA levels based on data published by

Acevedo, et al. [7]. Data published by Acevedo, et al. [7] belong to a reactor performing > 99 % phosphate removal with a defined population of > 85 % PAOs and show clearly identified cycling of polyphosphate, glycogen, PHB and PHV. Protein content was based on the work of Yücesoy, et al. [50].

Quota definition of compounds

The metabolites that are not in steady state consist of storage polymers (glycogen, polyphosphate, PHB and PH2MV), catalytic metabolites (enzymes and ribosomes) and non-catalytic metabolites (named here as biomass precursors and other proteins). Because some of these metabolites do not reinforce the autocatalytic behaviour of the system, their values in the vector M_0 will tend to be 0 as this vector is an outcome of the simulation to optimize growth. To enforce the synthesis of these compounds, Rügen, et al. [32] introduced the concept of quota compounds following the relation:

$$B_{quota_t} M_t \geq C_{quota_t} (w^T . M_t)$$

In this relation, B_{quota_t} is a matrix containing the index positions in the *M* vector of the metabolites for which the quota is being defined at time point *t*. C_{quota_t} is a vector containing the values for the said quotas. The indicated relation was used to define minimal quota levels at time 0 for glycogen, polyphosphate and PHAs based on data published by Acevedo, et al. [7]. It was also used to define an all-time minimal level of biomass precursors and non-catalytic proteins.

For this research, the concept of quota compounds was slightly modified to include specific and maximum quota compounds at indicated time points.

$$B_{quota\_eq_t} M_t = C_{quota\_eq_t} (w^T . M_t)$$

$$B_{quota\_max_t} M_t \leq C_{quota\_max_t} (w^T . M_t)$$

These additional concepts were used to define a set initial amount of substrate (acetate) at time 0 hours and to enforce its uptake by setting a maximum quota of 0 at 0.5 hours of the cycle.

Catalytic and non-catalytic constraints on fluxes

Comparable to conventional FBA, each reaction can be limited according to pre-assigned values in a lower (*lb*) and upper bound (*ub*) vector at each time point.

$$lb_t \leq v_t \leq ub_t$$

The lower and upper bounds were used to specify environmental limits on reactions (*e.g.* during the anaerobic period, oxygen consuming had an *ub* of 0

mmol/g_{DW}/h). Otherwise, the upper bounds were defined by the amount of a specific catalyst and the turnover rate of this catalyst ($k_{cat,E}$ values) by:

$$v_t \leq M_{t,E}.k_{cat,E}$$

where $M_{t,E}$ denotes a subsection of M_t containing the molar amount of enzymes at time t and $k_{cat,E}$ a vector with the catalytic turnover number ($k_{cat,E}$) of each enzyme. All reactions from the system except for CO_2 diffusion are catalysed by enzymes or ribosomes. The $k_{cat,E}$ values for each reaction were adapted from the work from Rügen, et al. [32] and if not considered in their research it was obtained from the BRENDA database [51]. The molar cost of all the enzymes was the assumed to be the same based on the work from Rügen, et al. [32].

Maintenance requirements:

Metabolic models commonly consider the expenditure of energy (ATP) for basic maintenance purposes. Here, a constant flux of 0.3980 mmol ATP/g_{DW}/h was used, based on literature values for PAOs [16]. This value is assumed for the anaerobic as well as aerobic phase and is independent of the amount of available substrate.

Simulation of dynamic conditions:

Simulations were performed considering a typical EBPR cycle: an anaerobic phase of 1.5 hours followed by an aerobic phase of 3.5 hours. The anaerobic conditions were simulated by setting the upper bound for the electron transport chain (ETC) reaction to 0 mmol O_2/g_{DW} /h. Conversely, during the aerobic phase the upper bound of this reaction was only limited by the capacity of the respective enzyme constraint. The substrate (acetate) was available at the beginning of the anaerobic phase (t=0h) by setting an initial metabolite quota for the starting amount. To enforce rapid consumption, a maximum quota of 0 mmol/g_{DW}/h was set at 30 minutes.

Optimization target and algorithm used:

The previously defined constraints (equalities and inequalities of the model) are discretized in defined time intervals throughout the cycle. Linear programming was then used to identify possible flux distributions at each time interval without an optimization objective. Thus, the obtained flux distribution for each time point is not an optimal unique solution, but rather a possible solution that fulfils the equality and inequality constraints mentioned so far (steady state metabolites, initial metabolite composition, quotas and enzyme capacities).

The possible solutions can be further constrained for any given α value (fold change of the system). The overall optimal target (maximum possible α value) was

set following the approach introduced by Rügen, et al. [32] and used by Reimers, et al. [52] and Guedes da Silva, et al. [33]. Briefly, a search for α values that fulfil the linear constraints is performed using an algorithm of binary search until the highest achievable α value is found with a defined accuracy (minimal step-size) of 10^{-7}. Due to the presence of internal cycles in the metabolic model, the solution obtained is not unique. To further limit the solution space, we identified and limited internal cycles from the available solutions with no effect on the outcome of the simulation.

Temperature dependencies:

Temperature was implemented as a parameter affecting the turnover rates (k_{cat}) of individual reactions following the simplified Arrhenius equation from 4 to 20 °C. This relation has been used previously to describe temperature effect on the kinetics of PAOs using a black-box kinetic model [17,38] and is expressed as:

$$k_{cat}\,(T) = k_{cat\,(20)} * \theta^{\,(T-20\,^{\circ}\mathrm{C})},$$

where $k_{cat}\,(T)$ represents the k_{cat} value at temperature T, $k_{cat\,(20)}$ the k_{cat} value at 20 °C (defined as optimal temperature for PAOs) and theta (θ) the temperature coefficient for the specific reaction in question. The values for θ are reaction specific, and were adapted from the previously determined θ values found in Lopez-Vazquez, et al. [17] and summarized in *Table S2*. The values for θ of reactions that have not been determined were set within the range of estimated θ values for comparable bioconversions. The effect of the set θ values on the model prediction was evaluated with a general parameter sweep.

Simulation with and without biomass synthesis constrains:

To evaluate the effect that biomass growth during the entire cycle or only during the aerobic phase, two simulations were performed. These simulations are referred to as the *unconstrained growth* (no constrains on biosynthesis) and the *aerobic growth* (biosynthesis can only occur during the aerobic period) models. The only difference between both models lays on the upper bounds allowed for the reactions involved in protein and ribosomes production (including enzymes) and synthesis of biomass precursors (BMP) synthesis (See *Table S1* for information on the stoichiometry of these reactions). In the *unconstrained growth* model, these bounds were constrained like all other reactions (set by the catalytic limitations), whereas in the *aerobic growth* model, these upper bounds were set to 0 mmol/g_{dw}/h during the anaerobic phase.

Software and model availability

All simulations were performed in MATLAB version 9.4 (R2019b) using LINPROG as the linear optimization solver. The original cFBA model was retrieved from Rügen, et al. [32] and the PAOs specific model retrieved from Guedes da Silva, et al. [33]. The adapted model used in this study is available at GitLab project ID 39202430 (https://gitlab.com/delft_paos/cFBA_temperature).

References

1 Konopka, A. What is microbial community ecology? *The ISME journal* **3**, 1223-1230 (2009).

2 Widder, S. *et al.* Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal* **10**, 2557-2568 (2016).

3 Van Loosdrecht, M., Lopez-Vazquez, C., Meijer, S., Hooijmans, C. & Brdjanovic, D. Twenty-five years of ASM1: past, present and future of wastewater treatment modelling. *Journal of Hydroinformatics* **17**, 697-718 (2015).

4 Beun, J. *et al.* Aerobic granulation in a sequencing batch reactor. *Water research* **33**, 2283-2290 (1999).

5 Strous, M., Heijnen, J., Kuenen, J. G. & Jetten, M. The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-oxidizing microorganisms. *Applied microbiology biotechnology* **50**, 589-596 (1998).

6 Lawson, C. E. *et al.* Autotrophic and mixotrophic metabolism of an anammox bacterium revealed by in vivo 13C and 2H metabolic network mapping. *The ISME journal* **15**, 673-687 (2021).

7 Acevedo, B. *et al.* Metabolic shift of polyphosphate-accumulating organisms with different levels of polyphosphate storage. *Water Research* **46**, 1889-1900 (2012).

8 Singleton, C. M. *et al.* Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nature communications* **12**, 1-13 (2021).

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

10 Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. *The ISME journal* **10**, 810-822 (2016).

11 Kleikamp, H. B. *et al.* Database-independent de novo metaproteomics of complex microbial communities. *Cell Systems* **12**, 375-383. e375 (2021).

12 Zomorrodi, A. R. & Segrè, D. Synthetic ecology of microbes: mathematical models and applications. *Journal of molecular biology* **428**, 837-861 (2016).

13 McDaniel, E. A. *et al.* Prospects for multi-omics in the microbial ecology of water engineering. *Water Research* **205**, 117608 (2021).

14 Mino, T., Van Loosdrecht, M. & Heijnen, J. Microbiology and biochemistry of the enhanced biological phosphate removal process. *Water Research* **32**, 3193-3207 (1998).

15 Oehmen, A. *et al*. Advances in enhanced biological phosphorus removal: from micro to macro scale. *Water Research* **41**, 2271-2300 (2007).

16 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. A structured metabolic model for anaerobic and aerobic stoichiometry and kinetics of the biological phosphorus removal process. *Biotechnology and Bioengineering* **47**, 277-287 (1995).

17 Lopez-Vazquez, C. M. *et al*. Modeling the PAO–GAO competition: effects of carbon source, pH and temperature. *Water Research* **43**, 450-462 (2009).

18 Oehmen, A., Carvalho, G., Lopez-Vazquez, C., Van Loosdrecht, M. & Reis, M. Incorporating microbial ecology into the metabolic modelling of polyphosphate accumulating organisms and glycogen accumulating organisms. *Water research* **44**, 4992-5004 (2010).

19 Varga, E. *et al*. Recent advances in bio-P modelling–a new approach verified by full-scale observations. *Water Science Technology* **78**, 2119-2130 (2018).

20 Santos, J. M. *et al*. A novel metabolic-ASM model for full-scale biological nutrient removal systems. *Water Research* **171**, 115373 (2020).

21 Comeau, Y., Hall, K. J., Hancock, R. E. & Oldham, W. K. Biochemical model for enhanced biological phosphorus removal. *Water Research* **20**, 1511-1521 (1986).

22 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnology and bioengineering* **43**, 461-470 (1994).

23 Sawers, G. The aerobic/anaerobic interface. *Current opinion in microbiology* **2**, 181-187 (1999).

24 Ribes, J., Keesman, K. & Spanjers, H. Modelling anaerobic biomass growth kinetics with a substrate threshold concentration. *Water research* **38**, 4502-4510 (2004).

25 Kleiner, M. Metaproteomics: much more than measuring gene expression in microbial communities. *Msystems* **4**, e00115-00119 (2019).

26 Sievers, S. in *Microbial Proteomics* 21-33 (Springer, 2018).

27 Helbig, A. O., Heck, A. J. & Slijper, M. Exploring the membrane proteome—challenges and analytical strategies. *Journal of proteomics* **73**, 868-878 (2010).

28 Wilmes, P. *et al*. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *The ISME journal* **2**, 853-864 (2008).

29 Wexler, M., Richardson, D. J. & Bond, P. L. Radiolabelled proteomics to determine differential functioning of Accumulibacter during the anaerobic and aerobic phases of a bioreactor operating for enhanced biological phosphorus removal. *Environmental microbiology* **11**, 3029-3044 (2009).

30 Hesselmann, R., Von Rummell, R., Resnick, S. M., Hany, R. & Zehnder, A. Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Research* **34**, 3487-3494 (2000).

31 Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nature biotechnology* **28**, 245-248 (2010).

32 Rügen, M., Bockmayr, A. & Steuer, R. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Scientific reports* **5**, 15247 (2015).

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

34 Russell, N. Cold adaptation of microorganisms. *Philosophical Transactions of the Royal Society of London. B,* **326**, 595-611 (1990).

35 Bosdriesz, E., Molenaar, D., Teusink, B. & Bruggeman, F. J. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *The FEBS journal* **282**, 2029-2044 (2015).

36 Metzl-Raz, E. *et al*. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *Elife* **6**, e28034 (2017).

37 Mairet, F., Gouzé, J.-L., de Jong, H. & Applications. Optimal proteome allocation and the temperature dependence of microbial growth laws. *npj Systems Biology* **7**, 1-11 (2021).

38 Brdjanovic, D. *et al*. Influence of temperature on biological phosphorus removal: process and molecular ecological studies. *Water Research* **32**, 1035-1048 (1998).

39 Brdjanovic, D., Loosdrecht, M. C. v., Hooijmans, C. M., Alaerts, G. J. & Heijnen, J. J. Temperature effects on physiology of biological phosphorus removal. *Journal of environmental engineering* **123**, 144-153 (1997).

40 Brdjanovic, D., Van Loosdrecht, M. C., Hooijmans, C. M., Alaerts, G. J. & Heijnen, J. J. Minimal aerobic sludge retention time in biological phosphorus removal systems. *J Biotechnology bioengineering* **60**, 326-332 (1998).

41 Garcia-Usach, F., Ferrer, J., Bouzas, A. & Seco, A. Calibration and simulation of ASM2d at different temperatures in a phosphorus removal pilot plant. *Water Science Technology* **53**, 199-206 (2006).

42 Lopez-Vazquez, C. M. *et al*. Temperature effects on the aerobic metabolism of glycogen-accumulating organisms. *Biotechnology* **101**, 295-306 (2008).

43 Whang, L. M. & Park, J. K. Competition between polyphosphate-and glycogen-accumulating organisms in enhanced-biological-phosphorus-removal systems: Effect of temperature and sludge age. *Water Environment Research* **78**, 4-11 (2006).

44 Welles, L. *et al*. Accumulibacter clades Type I and II performing kinetically different glycogen-accumulating organisms metabolisms for anaerobic substrate uptake. *Water research* **83**, 354-366 (2015).

45 Chen, Y. & Nielsen, J. Energy metabolism controls phenotypes by protein efficiency and allocation. *Proceedings of the National Academy of Sciences* **116**, 17592-17597 (2019).

46 Sun, F., Dai, C., Xie, J. & Hu, X. Biochemical issues in estimation of cytosolic free NAD/NADH ratio. *PloS one* **7**, e34525 (2012).

47 Zhao, J., Xie, S., Luo, Y. & Li, X. NADH accumulation during DPAO denitrification in a simultaneous nitrification, denitrification and phosphorus removal system. *Environmental Science: Water Research Technology* **7**, 1819-1827 (2021).

48 Urry, D. W. in Application of thermodynamics to biological and materials science (IntechOpen, 2011).

49 da Silva, L. G. *et al*. Revealing the Metabolic Flexibility of "Candidatus Accumulibacter phosphatis" through Redox Cofactor Analysis and Metabolic Network Modeling. *Applied Environmental Microbiology* **86** (2020).

50 Yücesoy, E., Lüdemann, N., Lucas, H., Tan, J. & Denecke, M. Protein analysis as a measure of active biomass in activated sludge. *Water science technology* **65**, 1483-1489 (2012).

51 Schomburg, I. *et al*. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic acids research* **41**, D764-D772 (2012).

52 Reimers, A.-M., Knoop, H., Bockmayr, A. & Steuer, R. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences* **114**, E6457-E6465 (2017).

6

METABOLIC IMPLICATIONS FOR DUAL SUBSTRATE GROWTH IN "CANDIDATUS ACCUMULIBACTER"

Published as: Páez-Watson, T., Jansens, C., van Loosdrecht, M. C. M., & Roy, S. Metabolic implications for dual substrate growth in "Candidatus Accumulibacter". bioRxiv, (2025).

Supplementary figures and tables are available in the online publication

Abstract

This study explores the metabolic implications of dual substrate uptake in *"Candidatus Accumulibacter"*, focusing on the co-consumption of volatile fatty acids and amino acids under conditions typical of enhanced biological phosphorus removal (EBPR) systems. Combining batch tests from highly enriched "*Ca.* Accumulibacter" cultures with conditional flux balance analysis (cFBA) predictions, we demonstrated that co-consumption of acetate and aspartate leads to synergistic metabolic interactions, lowering ATP loss compared to individual substrate consumption. The metabolic synergy arises from the complementary roles of acetate and aspartate uptake: acetate uptake provides acetyl-CoA to support aspartate metabolism, while aspartate conversion generates NADH, reducing the need for glycogen degradation during acetate uptake. We termed this type of metabolic interaction as reciprocal synergy. We further expanded our predictions to uncover three types of interactions between catabolic pathways when substrates are co-consumed by "*Ca.* Accumulibacter": (i) neutral, (ii) one-way synergistic and (iii) reciprocal synergistic interactions. Our results highlight the importance of network topology in determining metabolic interactions and optimizing resource use. These findings provide new insights into the metabolism "*Ca.* Accumulibacter" and suggest strategies for improving EBPR performance in wastewater treatment plants, where the influent typically contains a mixture of organic carbon compounds.

Introduction

Enhanced biological phosphorous removal (EBPR) systems are complex, dynamic environments that foster the growth of mixed microbial communities. Amongst these, members of the genus "*Ca.* Accumulibacter" are dominant, both in terms of biovolume[1, 2] and protein content[3, 4], highlighting their key role in the biological conversions that drive this process. Their unique metabolic capabilities allow them to accumulate large amounts of polyphosphate (PolyP)[5, 6] under cyclic, dynamic environments. No pure cultures of "*Ca.* Accumulibacter" are available, making research reliant on highly enriched cultures (> 90 % biovolume[1, 2]) for characterization.

The metabolism of "*Ca.* Accumulibacter" depends on the cycling of polyhydroxyalkanoates (PHAs), PolyP and glycogen during an EBPR cycle[7-9]. This enables the rapid anaerobic uptake and accumulation of volatile fatty acids (VFAs) inside the microbial cells. The stored polymers can be subsequently oxidized aerobically, generating enough energy for growth and the accumulation of PolyP. Research on this metabolic strategy has predominantly focused on feeding single substrates, such as VFAs like acetate [10-12] and propionate [13], as well as non-VFA substrates like glucose [14, 15], and amino acids like aspartate and glutamate[16].

It is noteworthy that environments containing only a single substrate are rare in nature, and wastewater treatment plants are no exception. Wastewater typically contains a mixture of organic substrates, including fatty acids, amino acids, sugars and lipids[17-19]. Understanding the effects of multiple substrates on the metabolism of microorganisms involved in phosphate removal is therefore crucial for optimizing EBPR processes.

Despite the importance of mixed substrates, studies on the metabolic mechanisms of "*Ca.* Accumulibacter" during the simultaneous uptake of two or more substrates are very limited. Studies have reported co-consumption of acetate with glucose [15, 20], acetate with propionate and lactate [21] and acetate with glycerol[22] in enrichment cultures, but the interactions between the metabolic strategies for consumption of each substrate remains poorly understood.

A notable study by Qiu et al. [16] stands out for its detailed examination of the concurrent uptake of acetate with either aspartate or glutamate. This research identified a potential synergy between acetate and aspartate, where aspartate uptake led to a net energy gain (~9 % ATP gain), enhancing acetate uptake. The proposed mechanism was the operation of fumarate reductase (reducing fumarate to succinate) which contributes to the *proton motive force* (pmf) required for acetate transport. More recently, a similar energetic gain was

evaluated for acetate and succinate co-consumption[23], though the entry point of succinate bypasses fumarate reductase, raising the question of whether this synergy is indeed driven by *pmf* or an unidentified metabolic interaction.

Understanding the interactions within metabolic networks is challenging due to the complexity and interconnectedness of metabolites, especially energy carriers like ATP, NADH, $FADH_2$[24]. In this regard, metabolic modelling provides valuable tools for studying these networks and how interactions emerge from stoichiometric rules. Techniques such as Flux Balance Analysis (FBA) can model steady-state metabolic operations[25], but more advanced methods are needed for dynamic systems like EBPR[26-28]. One such method is conditional FBA (cFBA)[29], which has been successfully used to model "*Ca.* Accumulibacter" metabolism, where intracellular storage polymers cycling emerged as a property of model stoichiometry and environmental conditions[30]. However, how these emergent properties change with multiple substrates remains an open question.

To address this, we aimed to understand the metabolic implications of dual substrate uptake in "*Ca.* Accumulibacter" under the dynamic conditions characteristic of EBPR. We developed and tested a metabolic model for the uptake of varying acetate and aspartate ratios to uncover synergistic interactions that improve growth yields. These findings were validated with lab-scale enrichments and the mechanisms behind the synergy are described. Finally, we extended this modelling approach to explore interactions with additional substrates, identifying a basic biological principle of metabolic interactions. This work lays the groundwork for further exploration of multiple substrate consumption, which could potentially lead to increased biomass yields compared to individual substrate consumption.

Results

Different anaerobic stoichiometries are employed for acetate, aspartate and combined substrate uptake by "*Ca.* Accumulibacter"

We expanded a previous metabolic model of "*Ca.* Accumulibacter"[32] to incorporate uptake mechanisms for acetate and aspartate (Figure 1.A). Next, we simulated the concurrent uptake of both substrates at varying ratios. To normalize the simulations, we ensured that the combined substrates provided the same amount of electron equivalents (analogous to the chemical oxygen demand, most

widely used in engineering), meaning the ratios were adjusted to achieve 'electron equivalence' rather than molar equivalence (see methods for details). The predictions indicated that the optimal anaerobic strategies employed for the individual uptake of acetate or aspartate were different, and that a mixed uptake also resulted in different individual anaerobic strategies rather than linear combinations of the individual strategies (Figure 1.B).

With acetate as the sole substrate, polyP and glycogen were degraded to supply resources for PHA accumulation primarily as acetyl-CoA precursors with a smaller fraction of propionyl-CoA. As the fraction of aspartate uptake increased, less glycogen degradation was required and larger fraction of PHAs as propionyl-CoA precursors was synthesized. At an equal electron equivalence ratio of acetate to aspartate, glycogen degradation halted. In scenarios with higher aspartate fractions, glycogen degradation was not observed, and the PHA pool showed a higher dominance of propionyl-CoA precursors with a higher demand for polyphosphate degradation (Figure 1.B).

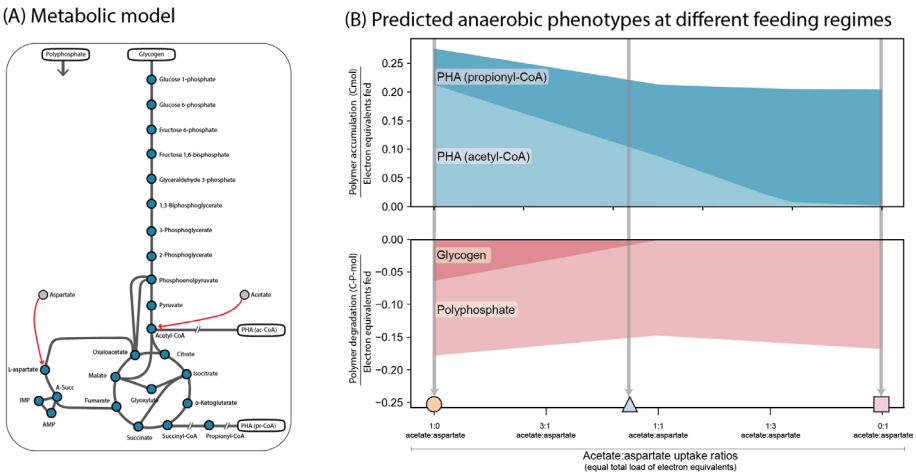


Figure 1. *Different anaerobic stoichiometries employed for the uptake of acetate, aspartate and combined substrates.* ***(A)*** *Schematic representation of the metabolic model of "Ca. Accumulibacter" used for simulations. Metabolites are represented as filled circles connected via reactions with grey lines. Red arrows indicate the entry point for substrate uptake in central carbon metabolism. Intracellular storage polymers are represented as white boxes.* ***(B)*** *Anaerobic conversions during the uptake of acetate/aspartate at varying electron equivalent ratios simulated with cFBA. (In blue) PHA accumulation subdivided into PHAs from acetyl-CoA (C2) and from propionyl-CoA (C3). (In red) Glycogen and Polyphosphate consumption. Three specific simulations marked with symbols (circle, triangle and square) were confirmed experimentally in Figure 2.*

To validate the modelling results, batch tests were performed on a lab reactor enrichment culture. qFISH analysis estimated the biovolume abundance of "*Ca.* Accumulibacter" at 89 ± 3 %., while metagenomics analysis revealed the enrichment of a clade I strain closely related to "*Ca.* Accumulibacter regalis". The enriched genome harboured the complete genetic potential required for aspartate metabolism (Figure 2.A).

The batch confirmed the optimal strategies predicted by the cFBA simulations. Substrate compositions were evaluated under three regimes: acetate, aspartate and a 45:55 electron equivalence ratio of acetate to aspartate consumption (regimes marked in Figure 1.B). The experimental results closely matched the predicted stoichiometries (Figure 2.B). Specifically, in the acetate-fed regime, PHAs accumulated anaerobically mainly as acetyl-CoA precursors, accompanied by the degradation of polyphosphate and glycogen. In the mixed substrate regime, PHAs accumulated as a balanced mixture of both acetyl-CoA and propionyl-CoA precursors, with lower glycogen degradation per electron equivalent consumed compared to the acetate-only regime, consistent with the predictions. Finally, in the aspartate-fed regime, PHAs accumulated with a substantial decrease in acetyl-CoA precursors. This regime required the highest

polyphosphate degradation and no glycogen degradation, aligning with the predictions (See Figure 1.B and Figure 2.B for comparison).

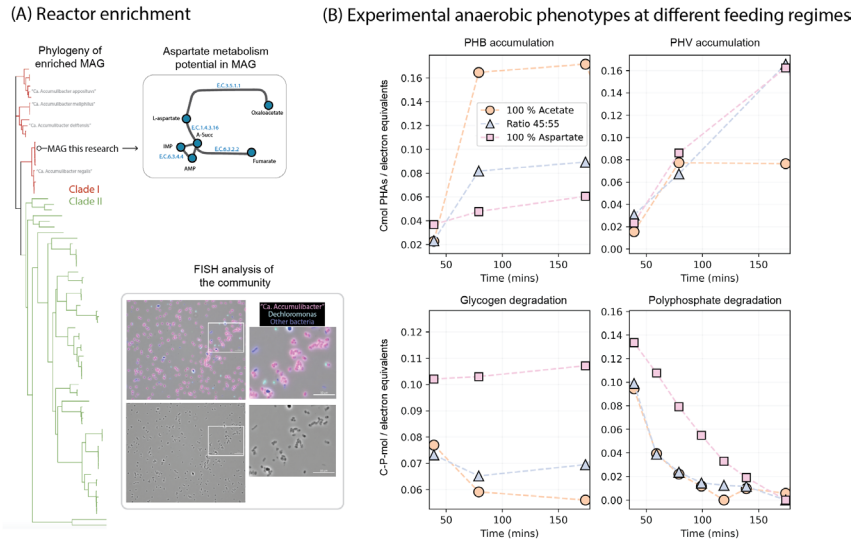


Figure 2. Lab culture enrichment and experimental validation of the model. (A) *Microbial community analysis with metagenomics and FISH. Metagenomics revealed the enrichment of an MAG closely associated to Ca. Accumulibacter Clade I "Ca. Accumulibacter regalis" and harbouring all the genes necessary for aspartate metabolism. FISH image of the enrichment in which magenta colour represents the overlap of "Ca. Accumulibacter" (red) and eubacteria (blue). Bottom image of phase contrast highlighting typical morphology observed from PAOs enrichments.* ***(B)*** *Experimental validation of the anaerobic phase metabolic strategies observed during batch tests for acetate, aspartate, and mixed acetate:aspartate (45:55 electron equivalence) regimes. Distinct markers (circles, triangles, and squares) facilitate direct comparison with the modeled predictions in Figure 1.B.*

Metabolic and energetic balances reveal complementary strategies for the uptake of acetate and aspartate leading to enhanced growth yields

Simulations with varying ratios of acetate and aspartate in the feed revealed not only changes in internal storage polymer utilization under anaerobic conditions, but also predicted maximum growth yields per electron equivalents for each cycle (Figure 3.A). Growth on aspartate was more efficient than growth on acetate. Notably, the highest growth yield was achieved with a combination of both substrates, specifically at a 1:4 electron equivalent ratio of acetate to aspartate.

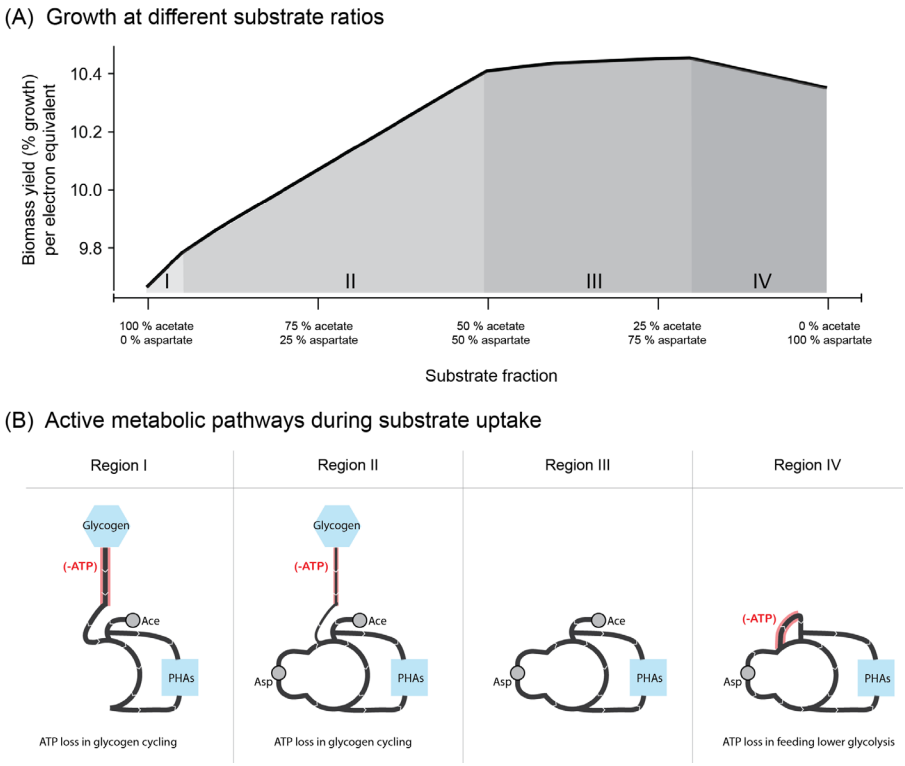


Figure 3. (A) *Biomass growth yield as a duplication (% growth) in one cycle per electron equivalents fed at varying fractions of acetate aspartate. Each region represents a unique metabolic strategy employed by the model to optimize growth.* ***(B)*** *Active metabolic operation during anaerobic substrate uptake predicted via the cFBA model at each region indicated in (A). Metabolic reactions contributing to a net ATP loss of the system considering the whole EBPR cycle have been highlighted in red. Lower thickness in Region II illustrates the lower demand on glycogen degradation than in Region I.*

During acetate uptake, glycogen was degraded to supply NADH necessary for PHA accumulation. The most efficient strategy for PHA accumulation reflected in the model involved glycogen degradation via glycolysis to phosphoenolpyruvate (PEP), then converting PEP to oxaloacetate (OAA) to fuel the TCA cycle. This allowed partial acetate oxidation in the right branch of the TCA cycle, producing NADH and Propionyl-CoA-type PHAs (Figure 3.B – region I). However, during the aerobic phase, this strategy required glycogen replenishment and the resulting glycolysis/gluconeogenesis operation over the cycle results in a net ATP loss, thus reducing the overall growth yield.

As the aspartate-to-acetate ratio increased, aspartate metabolism generated additional NADH via aspartate oxidase that countered the NADH requirement from glycogen degradation (Figure 3.B – region II), thus lowering the net ATP loss in the glycolysis/gluconeogenesis cycle, which improved growth yields. When the

ratio reached 1:1, glycogen degradation was no longer necessary. Beyond this point, up to a 1:4 acetate-to-aspartate ratio, there were no net ATP losses, resulting in the highest growth yields (Figure 3.B – region III). In this range reciprocal benefits were observed when aspartate consumption provided NADH that benefited acetate metabolism, while acetate consumption supplied acetyl-CoA equivalents that supported aspartate metabolism (as described below).

At higher aspartate fractions, the metabolic strategy necessitated the operation of the right branch of the TCA cycle, which required acetyl-CoA equivalents. These equivalents could be supplied through acetate uptake. At insufficient acetate fractions, part of the consumed aspartate was channelled towards acetyl-CoA generation via PEP carboxykinase (PEPCK), raising the demand for ATP (Figure 3.B – region IV). This was met with an increased polyphosphate degradation, necessitating increased ATP requirements in the aerobic phase to replenish the polyphosphate pools resulting in lower growth yields.

Synergistic effects of substrate co-consumption vary by metabolic entry point and network topology

Several substrates were incorporated into the existing metabolic model of "*Ca*. Accumulibacter", and their co-consumption with acetate at varying ratios was simulated using cFBA, as described in the previous section. The predictions indicated that multiple substrates could support PHA accumulation without relying on reducing equivalents from glycogen degradation, which is typically required during anaerobic acetate uptake (Figure 4).

Interestingly, the co-consumption of certain substrates with acetate mirrored the reciprocal synergistic effect observed with aspartate. Specifically, these combinations led to an enhanced biomass yield per electron equivalent compared to the yield of individual substrates. We referred to these as reciprocal synergistic interactions. Substrates exhibiting this behaviour included succinate, fumarate, malate, oxaloacetate, and aspartate, all of which enter the reducing branch (left-hand side) of the TCA cycle. The metabolism of these substrates resulted in sufficient NADH production to alleviate the dependence from glycogen degradation, the main limitation during acetate uptake. Complementarily, metabolising these substrates benefited from the uptake of acetate to feed acetyl-CoA equivalents into the TCA cycle, as was the case with aspartate.

In contrast, another class of substrates co-consumed with acetate produced biomass yields greater than the sum of the individual parts but did not surpass

the yield of the more favourable substrate on its own. These substrates were able to generate sufficient reducing power (NADH) to alleviate the reliance on glycogen degradation, releasing the limitation for acetate metabolism. However, they did not benefit from the additional acetate uptake, since their metabolism did not require acetyl-CoA to be fed into the TCA cycle using PEPCK or similar reactions. We classified these as one-way synergistic interactions, observed for butyrate, lactate, pyruvate and citrate.

Finally, certain substrates such as propionate and alpha-ketoglutarate resulted in biomass yields that closely matched the sum of the individual yields, with no additional gain from co-consumption. We classified these as neutral interactions, wherein the metabolic demands of these substrates closely resembled that of acetate. These substrates required similar resources (NADH and ATP) as acetate, leading to overlapping metabolic strategies that did not enhance overall growth yield.

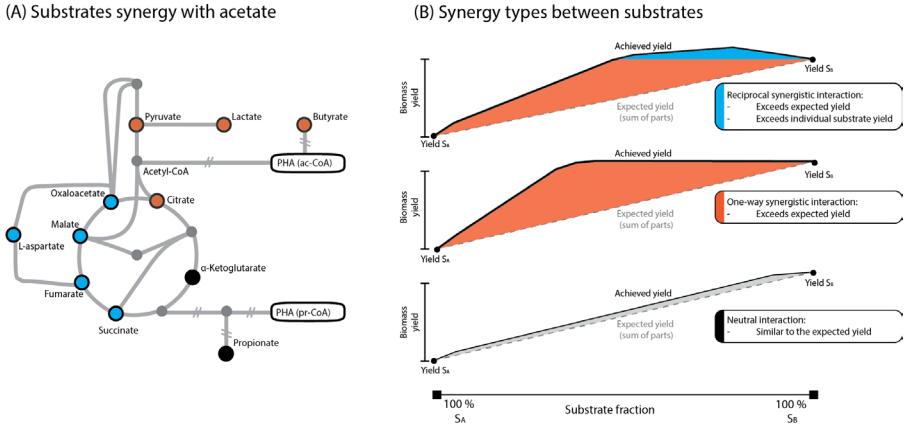


Figure 4. *Substrates interactions when co-consumed.* ***(A)*** *Metabolic model of "Ca. Accumulibacter" highlighting the external substrates that were tested and their type of interaction when co-fed with acetate. Blue: extra-ordinary interaction. Orange: positive interaction. Black: neutral interaction.* ***(B)*** *Types of interactions (substrate synergy) existing between two substrates. Top shows reciprocal synergy between acetate and aspartate in which the biomass yield can exceed that of the sum of parts and the individual maximum biomass yields. Middle panel shows a positive interaction between acetate and lactate in which the biomass yield can exceed that of the sum of parts but not the individual maximum yield. Bottom panel shows a neutral interaction between acetate and propionate in which the biomass yield is similar to the expected from the sum of parts.*

Importantly, when thermodynamic constraints were removed from the model, allowing all reactions to operate reversibly, no significant differences were observed between the substrates. In this scenario, all co-consumed substrates resulted in similar biomass yields, producing a horizontal line in the interaction

plot, demonstrating the absence of any metabolic synergy under reversible reaction conditions.

Discussion

This study demonstrates that "*Ca.* Accumulibacter" exhibits synergistic metabolic interactions when acetate is co-consumed with aspartate, leading to enhanced growth yields and reduced net ATP loss. These findings offer new insights into the metabolic mechanisms of substrate co-consumption in EBPR systems and highlight the importance of network topology in determining interaction outcomes. Below, we contextualize these results within existing research on EBPR, biological adaptation, and energy cycling while also acknowledging the limitations of the predictive models used.

Wastewater typically consists of a mix of organic substrates[17-19], and many studies have documented "Ca. Accumulibacter" as capable of consuming multiple substrates, both through genomic analysis[16, 51] and *in situ* studies[52, 53]. Our findings align with these studies, showing that co-consumption of acetate and aspartate not only is possible but also results in a metabolic interaction that optimizes cellular resource. The reciprocal synergy observed arises from a two-way release of metabolic limitations for each substrate, improving overall biomass yield (Figure 2). Specifically, acetate uptake is typically limited by anaerobic glycogen degradation, resulting in net ATP loss[9, 54], while aspartate uptake requires ATP-consuming conversion of acetyl-CoA via PEPCK[16]. The combination of both substrates alleviates these limitations, leading to improved growth. It is noteworthy that the identified release in metabolic limitations were not dependent on potential *pmf* generation as hypothesized by Qiu et al.[16] and thus can also explain the results obtained by Chen et al.[23].

The energy losses associated with glycogen cycling during acetate consumption in "*Ca.* Accumulibacter" has been well documented[5, 9, 10, 55-58], but their direct connection to physiological effects, such as growth yields, has not been previously established. Research manipulating growth rates by adjusting biomass retention time suggest that higher glycogen cycling corresponds to lower biomass yields[59-61], though the connection to energetics and metabolism has not been discussed extensively. This behaviour can be compared to the broader context of "ATP demanding yet useful" cycles[54]. Since the ATP loss in cycling glycogen is temporally separated, this is not a futile cycle in strict sense. Yet the temporal separation of glycogen degradation and replenishment serves an adaptive purpose, allowing organisms to maintain metabolic flexibility in response to

dynamic environments. This flexibility supports survival and growth under fluctuating conditions, similarly as that of apparent ATP-demanding pathways[62-65]. Further consideration of the metabolic and energetic cost/benefit of these temporally separated cycles needs to be considered.

The interaction between the metabolic operations of acetate and aspartate is complex, due to both the complexity inherent of metabolic networks[66] and the dynamic nature of the EBPR cycle. The metabolic consequences of anaerobic uptake strategies can only be fully understood when considering the entire cycle. Most research in EBPR focuses on anaerobic processes, where the cell's redox state is tightly constrained, limiting reaction feasibility[32]. However, without a holistic perspective on the aerobic phase, it is difficult to assess how anaerobic pathways impact overall metabolic fitness. The cFBA model employed here provides a tool to explore these interconnected processes, revealing metabolic strategies that minimize ATP losses as emergent properties of the system rather than being predefined *a priori*. Our experimental results aligned well with the model's predictions, indicating that cFBA model successfully captured the key features of substrate interactions and energy flows. This agreement between predictions and experimental data reinforces the utility of cFBA in understanding complex metabolic behaviors.

While our model successfully identified synergistic interactions, it was not able to detect any substrate combinations that would lower biomass yields (in essence, negative interactions). This limitation stems from the model's focus on global biomass yield optimization. Potential negative interactions might arise when dealing with substrates that activate stress responses or an overproduction of reductive potential and warrant further investigation.

Understanding the complete EBPR cycle is essential for uncovering the metabolic costs associated with PAO strategies. Studies examining different carbon substrates, such as butyrate and lactate, have shown shifts in glycogen use that may impact EBPR performance, but these effects have not been integrated into a broader understanding of biomass growth. Our results suggest that the energy wasteful use of glycogen in "*Ca.* Accumulibacter" when consuming acetate can be greatly released when co-consumed with many other substrates (amongst them, butyrate, lactate, pyruvate, citrate, oxaloacetate, malate, fumarate and succinate. See Figure 4). Further experiments are needed to uncover the metabolic effects of multiple organic substrates and how it can be employed to improve phosphorus removal.

Conclusions

- Co-consumption of acetate and aspartate by “Ca. Accumulibacter” results in synergistic metabolic interactions that improve biomass yield and reduce ATP losses.
- Acetate conversion to PHA benefits from NADH generated during aspartate metabolism, while aspartate uptake is supported by acetyl-CoA produced from acetate uptake.
- Glycogen cycling related to growth on acetate is energy demanding, co-consumption with other substrates (e.g., aspartate, succinate, fumarate) reduces this energy demand.
- A holistic consideration of the entire EBPR cycle is essential to fully understand the metabolic strategies and optimize the performance of PAOs.
- Synergistic interactions arising from metabolic optimization present an opportunity for co-utilization of carbon substrates that can be exploited to enhance the yield of bio-based processes.

Materials and Methods

Metabolic model and cFBA simulations

The metabolism of "*Ca.* Accumulibacter" was simulated with cFBA using the py_cFBA toolkit implementation[31]. A basic metabolic model was constructed as a stoichiometric matrix (**S**), representing the relationships between metabolites and reactions. Stoichiometries for reactions involved in glycogen degradation, glycolysis, the TCA cycle, anaplerotic routes, and PHA synthesis were adapted from an earlier study on "*Ca.* Accumulibacter"[32], excluding the reaction *MalE* which was present in only a few genomes within this genus. Stoichiometries for aspartate metabolism were obtained from[16] and the presence of this pathway in our enrichment culture was confirmed with metagenomics (see later in this section). A reaction representing synthesis of 1 c-mole of biomass was implemented in **S** following the stoichiometry from [30], which combined the energy (ATP) requirements for bacterial growth from acetyl-CoA from [33] and the overall stoichiometry of PAOs growth from [8, 34].

In the model, selected metabolites were defined as imbalanced, allowing their accumulation or depletion over time during simulations. These included acetate, aspartate, glycogen, PHB, PH2MV, CO2, polyP, and biomass. All other metabolites were balanced, adhering to the steady-state assumption of FBA. Biomass was defined as the sole contributor to the weights vector (*w*) in the cFBA formulation.

The model was implemented in Python using the py_cFBA toolkit, which generated SBML files for each configuration. Simulations of an EBPR cycle consisted of five time points (Δt = 1 hour), with no enzyme capacity constraints. Anaerobic and aerobic phases were simulated by allowing the reactions ETC_NADH, ETC_FADH (electron transport chain oxygen consumption), and biomass synthesis to occur exclusively in the final two time points of each cycle. Reaction reversibility was defined using upper and lower bounds based on a prior thermodynamic evaluation[32].

Substrate uptake during the anaerobic phase was enforced using quota definitions. An equality-quota at the initial time point specified the concentration of substrate fed, followed by a max-quota of zero in subsequent time points. This enforced the anaerobic uptake of substrate. Substrate concentrations were normalized to provide equivalent electron equivalents, even for substrate mixtures, based on their degree of reduction (e.g., 8 electrons for acetate, 12 electrons for aspartate). All simulations optimized biomass synthesis as the global target across the entire cycle, rather than at each time step, consistent with cFBA

methodology. All SBML models, simulation files and results are available at https://github.com/TP-Watson/PAOs_co-substrates_cFBA.

"*Ca* Accumulibacter" Enrichment

A "Ca. Accumulibacter" enrichment was obtained in a 1.5 L (1 L working volume) sequencing batch reactor (SBR), following conditions described earlier[2] with some adaptations. The reactor was inoculated using enriched sludge from the work of Páez-Watson et al.[2], which was previously inoculated with activated sludge from a municipal wastewater treatment plant (Harnaschpolder, The Netherlands). Each SBR cycle lasted 6 hours, consisting of 30 minutes of settling, 50 minutes of effluent removal, 10 minutes of N_2 sparging, 30 minutes of anaerobic feeding, 105 minutes of anaerobic phase and 135 minutes of aerobic phase. N_2 gas and compressed air were sparged at 500 ml/min into the reactor broth to maintain anaerobic and aerobic conditions respectively. The hydraulic retention time (HRT) was 12 hours (removal of 500 mL of broth per cycle, each cycle of 6 hours). The average solids retention time (SRT) was controlled to 9 days by the removal of 27,7 ml of mixed broth at the end of the mixed aerobic phase in each cycle. The pH was controlled at 7.3 ± 0.1 by dosing 0.5 M HCl or 0.5 M NaOH. The temperature was maintained at 20 ± 1 °C.

The reactor was fed with three separate media components diluted in demineralized water: a concentrated COD medium (400 mg COD/L) of acetate (13 g/L sodium acetate ×$3H_2O$); a concentrated mineral medium (0.69 g/L NH_4Cl, 2.16 g/L $MgSO_4$×$7H_2O$, 0.54 g/L $CaCl_2$×$2H_2O$, 0.64 KCl, 0.06 g/L N-allylthiourea (ATU), 0.06 g/L yeast extract and 6 mL/L of trace element solution prepared following Smolders et al.[7]; and a phosphate solution containing 0.76 g/L NaH_2PO_4×H_2O and 0.8 g/L Na_2HPO_4×$2H_2O$. In each cycle, 75 mL of COD medium, 75 ml of mineral medium and 360 mL of phosphate solution were added to the reactor during the 30 minutes of feeding. The final feed contained 400 mg COD/L of acetate.

Batch tests

Batch tests were conducted in the bioreactor on the enriched biomass once a *pseudo* steady state was reached (determined by a constant phosphate release and removal over multiple days). For the batch tests, 400 ml of H_2O and 50 ml of mineral media were fed as usual during the anaerobic phase. Later, 50 ml of organic substrate (containing either acetate, aspartate or a mix) was pulse fed and considered the beginning of the anaerobic phase of the cycle. The anaerobic phase on these batch tests was extended by 30 minutes to compensate for the delay in feed. The organic media was prepared such that the final feed contained 400 mg COD/L of acetate, aspartate or a mix (calculated by using the degree of

reduction of 8 and 12 e^-/mol for acetate and aspartate, respectively). Thus, the organic substrate solution for the tests contained only acetate, only aspartate or mix of acetate and aspartate as follows: (i) 13.1 g/L sodium acetate trihydrate ($C_2H_3NaO_2 \cdot 3H_2O$), (ii) 19.35 g/L sodium aspartate ($C_4H_6NNaO_4$), or (iii) 5,8 g/L sodium acetate trihydrate with 9,6 g/L sodium aspartate. For mixed substrates, the net consumption of acetate and aspartate was used to determine the acetate:aspartate uptake ratio.

Reactor and biomass analyses

Extracellular concentrations of phosphate and ammonium were measured with a Gallery Discrete Analyzer (Thermo Fisher Scientific, Waltham, MA). Acetate was measured by high performance liquid chromatography (HPLC) with an Aminex HPX-87H column (Bio-Rad, Hercules, CA), coupled to RI and UV detectors (Waters, Milford, MA), using 0.0015 M phosphoric acid as eluent supplied at a flowrate of 1 mL/min.

The biomass concentration (total and volatile suspended solids – TSS and VSS) was measured in accordance with Standard Methods as described in Smolders et al.[7] with some modifications: 10 ml of mixed broth were obtained at the end of the aerobic phase, centrifuged at 3600*g during 3 minutes and washed twice with demineralized water to remove salts. The sludge was then dried at 100 °C for 24 hours and weighed on a microbalance to determine the dry content – TSS. The ash content was determined by incinerating the dry material in an oven at 550 °C, and the difference used to calculate the VSS.

For glycogen and PHA determination, biomass samples (10 ml mixed broth) were collected throughout the batch test and stored in 15 ml conical tubes containing 0.3 ml of 37 % formaldehyde to stop biological activity. After each batch test, the biomass tubes were pottered to break the granular structure of the biomass, centrifuged at 3700 *g for 5 minutes and washed twice. The pellet was then frozen at-80 °C for at least 3 hours and freeze dried. For glycogen analysis the method described by Smolders et al.[7] was used: 5 mg of dry biomass was digested in 0.9 M HCl solutions in glass tubes at 100 °C for 5 hours. After this time, tubes were cooled at room temperature, filtered with 0.45 µm Whatman disk filters and neutralized with equal volumes of 0.9 M NaOH. The glucose resulting from digestion was quantified using the D-Glucose Assay Kit (GOPOD Format) from Megazyme (Bray, Ireland).

Microbial community characterization

The microbial community of the reactor was characterized at pseudo steady state as defined earlier. Two orthogonal approaches were used for the community characterization: metagenomics and Fluorescence in-situ hybridization (FISH).

For FISH, samples underwent handling, fixation, and staining procedures outlined by Winkler et al[35]. Bacteria were selectively identified using a blend of EUB338, EUB338-II, and EUB338-III probes[36, 37]. "*Ca*. Accumulibacter" was visualized employing mixtures of the probes Acc1011, Acc471, Acc471_2, Acc635, Acc470 designed and tested previously for different "*Ca.* Accumulibacter" lineages[38]. The images were captured with an epifluorescence microscope equipped with filter set Cy3 (ET545/25x ET605/70 m T565LPXR), Cy5 (ET640/30x ET690/50 m T660LPXR), and FITC (ET470/40x ET525/50 m T495LPXR) (Axio Imager M2, Zeiss, Germany). Quantitative FISH (qFISH) was done as a percentage of total biovolume over 12 representative pictures using the Daime software (DOME, Vienna, Austria)[39].

For metagenomics, DNA from the biomass samples was extracted using the DNeasy PowerSoil Pro-Kit (Qiagen, Germany) following the manufacturer's protocol. Shotgun sequencing was performed by Hologenomix (Delft, Netherlands). Paired-end sequencing with a read length of 150 bp was conducted using the Illumina NovaSeq X sequencing system. Library preparation was carried out using the Nextera XT DNA Library Preparation Kit. Approximately 10 Gbp of sequencing data were generated per sample.

The quality of raw sequenced reads was assessed using FastQC (version 0.11.7) with default parameters[40], and results were visualized with MultiQC (version 1.19). Low-quality paired-end reads were trimmed and filtered using Fastp (version 0.23.4) in paired-end mode[41]. Taxonomic classification of raw reads was performed to profile the microbiome in each sample using Kraken2 (version 2) with the standard database, which includes all complete bacterial, archaeal, and viral genomes in the NCBI RefSeq database, complemented by a curated wastewater database (sludgeDB)[42].

Clean reads were assembled into contigs using MetaSPAdes (version 3.15.5) with default parameters[43]. The resulting contigs were binned using MetaBAT (version 2.2.15) to reconstruct metagenome-assembled genomes (MAGs) with default parameters[44]. Bin completeness and contamination were assessed using CheckM (version 1.2.2) with the "lineage_wf" workflow [45]. Relative abundance of bins with contamination below 5% was determined in each sample using CoverM (version 0.7.0, https://github.com/wwood/CoverM) with default parameters.

For phylogenetic analysis, bins were classified using GTDB-Tk (version 2.4.0) and GTDB release 220[46]. The *ppk1* gene was utilised as a marker in bins identified as Accumulibacter. hmmsearch [47] was used with the ppk1.hmm profile, taking the best hit as the *ppk1* gene. Identified *ppk1* genes were combined with those in an existing database and aligned with MUSCLE (version 5.1)[48]. A phylogenetic tree of these *ppk1* sequences was generated with RaxML-NG (version 1.2.2)[49] and visualised using iTol (v6)[50].

References

1. Lu, H.; Oehmen, A.; Virdis, B.; Keller, J.; Yuan, Z., Obtaining highly enriched cultures of Candidatus Accumulibacter phosphates through alternating carbon sources. *Water Research* **2006,** *40* (20), 3838-3848.

2. Páez-Watson, T.; Tomás-Martínez, S.; de Wit, R.; Keisham, S.; Tateno, H.; van Loosdrecht, M. C.; Lin, Y., Sweet Secrets: Exploring Novel Glycans and Glycoconjugates in the Extracellular Polymeric Substances of "Candidatus Accumulibacter". *ACS Es&t Water* **2024,** *4* (8), 3391-3399.

3. Kleikamp, H. B.; Grouzdev, D.; Schaasberg, P.; van Valderen, R.; van der Zwaan, R.; van de Wijgaart, R.; Lin, Y.; Abbas, B.; Pronk, M.; van Loosdrecht, M. C., Comparative metaproteomics demonstrates different views on the complex granular sludge microbiome. *bioRxiv* **2022**, 2022.03. 07.483319.

4. Kleikamp, H. B.; Grouzdev, D.; Schaasberg, P.; van Valderen, R.; van der Zwaan, R.; van de Wijgaart, R.; Lin, Y.; Abbas, B.; Pronk, M.; van Loosdrecht, M. C., Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. *Water research* **2023,** *246*, 120700.

5. Mino, T.; Arun, V.; Tsuzuki, Y.; Matsuo, T., Effect of phosphorus accumulation on acetate metabolism in the biological phosphorus removal process. In *Biological phosphate removal from wastewaters*, Elsevier: 1987; pp 27-38.

6. Kortstee, G. J.; Appeldoorn, K. J.; Bonting, C. F.; van Niel, E. W.; van Veen, H. W., Biology of polyphosphate-accumulating bacteria involved in enhanced biological phosphorus removal. *FEMS microbiology reviews* **1994,** *15* (2-3), 137-153.

7. Smolders, G.; Van der Meij, J.; Van Loosdrecht, M.; Heijnen, J., Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnology and bioengineering* **1994,** *43* (6), 461-470.

8. Smolders, G.; Van der Meij, J.; Van Loosdrecht, M.; Heijnen, J., Stoichiometric model of the aerobic metabolism of the biological phosphorus removal process. *Biotechnology and bioengineering* **1994,** *44* (7), 837-848.

9. Smolders, G.; Van der Meij, J.; Van Loosdrecht, M.; Heijnen, J., A structured metabolic model for anaerobic and aerobic stoichiometry and kinetics of the biological phosphorus removal process. *Biotechnology and Bioengineering* **1995,** *47* (3), 277-287.

10. Van Loosdrecht, M.; Smolders, G.; Kuba, T.; Heijnen, J., Metabolism of micro-organisms responsible for enhanced biological phosphorus removal from wastewater, Use of dynamic enrichment cultures. *Antonie van Leeuwenhoek* **1997,** *71* (1-2), 109-116.

11. Mino, T.; Van Loosdrecht, M.; Heijnen, J., Microbiology and biochemistry of the enhanced biological phosphate removal process. *Water Research* **1998,** *32* (11), 3193-3207.

12. Welles, L.; Tian, W.; Saad, S.; Abbas, B.; Lopez-Vazquez, C.; Hooijmans, C.; Van Loosdrecht, M.; Brdjanovic, D., Accumulibacter clades Type I and II performing kinetically different glycogen-accumulating organisms metabolisms for anaerobic substrate uptake. *Water research* **2015,** *83*, 354-366.

13. Pijuan, M.; Saunders, A. M.; Guisasola, A.; Baeza, J. A.; Casas, C.; Blackall, L., Enhanced biological phosphorus removal in a sequencing batch reactor using propionate as the sole carbon source. *Biotechnology and bioengineering* **2004,** *85* (1), 56-67.

14. Elahinik, A.; Li, L.; Pabst, M.; Abbas, B.; Xevgenos, D.; van Loosdrecht, M. C.; Pronk, M., Aerobic granular sludge phosphate removal using glucose. *Water Research* **2023,** *247*, 120776.

15. Ziliani, A.; Bovio-Winkler, P.; Cabezas, A.; Etchebehere, C.; Garcia, H. A.; López-Vázquez, C. M.; Brdjanovic, D.; van Loosdrecht, M. C.; Rubio-Rincón, F. J., Putative metabolism of Ca. Accumulibacter via the utilization of glucose. *Water Research* **2023,** *229*, 119446.

16. Qiu, G.; Liu, X.; Saw, N. M. M. T.; Law, Y.; Zuniga-Montanez, R.; Thi, S. S.; Ngoc Nguyen, T. Q.; Nielsen, P. H.; Williams, R. B.; Wuertz, S., Metabolic traits of Candidatus Accumulibacter clade IIF strain SCELSE-1 using amino acids as carbon sources for enhanced biological phosphorus removal. *Environmental Science & Technology* **2019,** *54* (4), 2448-2458.

17. Huang, M.-h.; Li, Y.-m.; Gu, G.-w., Chemical composition of organic matters in domestic wastewater. *Desalination* **2010,** *262* (1-3), 36-42.

18. Bengtsson, S.; Hallquist, J.; Werker, A.; Welander, T., Acidogenic fermentation of industrial wastewaters: Effects of chemostat retention time and pH on volatile fatty acids production. *Biochemical Engineering Journal* **2008,** *40* (3), 492-499.

19. Qiu, G.; Zuniga-Montanez, R.; Law, Y.; Thi, S. S.; Nguyen, T. Q. N.; Eganathan, K.; Liu, X.; Nielsen, P. H.; Williams, R. B.; Wuertz, S., Polyphosphate-accumulating organisms in full-scale tropical wastewater treatment plants use diverse carbon sources. *Water research* **2019,** *149*, 496-510.

20. Gebremariam, S. Y.; Beutel, M. W.; Christian, D.; Hess, T. F., Effects of glucose on the performance of enhanced biological phosphorus removal activated sludge enriched with acetate. *Bioresource Technology* **2012,** *121*, 19-24.

21. Rubio-Rincón, F. J.; Welles, L.; Lopez-Vazquez, C. M.; Abbas, B.; van Loosdrecht, M. C.; Brdjanovic, D., Effect of lactate on the microbial community and process performance of an EBPR system. *Frontiers in microbiology* **2019,** *10*, 125.

22. Yuan, Q.; Sparling, R.; Lagasse, P.; Lee, Y.; Taniguchi, D.; Oleszkiewicz, J., Enhancing biological phosphorus removal with glycerol. *Water Science and Technology* **2010,** *61* (7), 1837-1843.

23. Chen, L.; Wei, G.; Zhang, Y.; Wang, K.; Wang, C.; Deng, X.; Li, Y.; Xie, X.; Chen, J.; Huang, F., Candidatus Accumulibacter use fermentation products for enhanced biological phosphorus removal. *Water Research* **2023,** *246*, 120713.

24. Nielsen, J., Systems biology of metabolism. *Annual review of biochemistry* **2017,** *86* (1), 245-275.

25. Orth, J. D.; Thiele, I.; Palsson, B. Ø., What is flux balance analysis? *Nature biotechnology* **2010,** *28* (3), 245-248.

26. Mahadevan, R.; Edwards, J. S.; Doyle III, F. J., Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical journal* **2002,** *83* (3), 1331-1340.

27. Sarkar, D.; Mueller, T. J.; Liu, D.; Pakrasi, H. B.; Maranas, C. D., A diurnal flux balance model of Synechocystis sp. PCC 6803 metabolism. *PLoS computational biology* **2019,** *15* (1), e1006692.

28. Liu, L.; Bockmayr, A., Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling. *Journal of Theoretical Biology* **2020,** *501*, 110317.

29. Rügen, M.; Bockmayr, A.; Steuer, R., Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using

conditional FBA. *Scientific reports* **2015,** *5*, 15247.

30. Páez-Watson, T.; van Loosdrecht, M. C.; Wahl, S. A., Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: Application to polyphosphate accumulating organisms. *Water Research* **2023,** *228*, 119365.

31. Páez-Watson, T.; Hernández Medina, R.; Vellekoop, L.; van Loosdrecht, M. C.; Wahl, S. A., Conditional Flux Balance Analysis (cFBA) Toolbox for python: application to research metabolism in cyclic environments. *Bioinformatics Advances* **2024**, vbae174.

32. Páez-Watson, T.; van Loosdrecht, M. C.; Wahl, S. A., From metagenomes to metabolism: Systematically assessing the metabolic flux feasibilities for "Candidatus Accumulibacter" species during anaerobic substrate uptake. *Water Research* **2024,** *250*, 121028.

33. Gommers, P.; Van Schie, B.; Van Dijken, J.; Kuenen, J., Biochemical limits to microbial growth yields: an analysis of mixed substrate utilization. *Biotechnology and bioengineering* **1988,** *32* (1), 86-94.

34. Oehmen, A.; Zeng, R. J.; Keller, J.; Yuan, Z., Modeling the Aerobic Metabolism of Polyphosphate-Accumulating Organisms Enriched with Propionate as a Carbon Source. *Water Environment Research* **2007,** *79* (13), 2477-2486.

35. Winkler, M.-K.; Bassin, J.; Kleerebezem, R.; De Bruin, L.; Van den Brand, T.; Van Loosdrecht, M., Selective sludge removal in a segregated aerobic granular biomass system as a strategy to control PAO–GAO competition at high temperatures. *Water research* **2011,** *45* (11), 3291-3299.

36. Amann, R. I.; Binder, B. J.; Olson, R. J.; Chisholm, S. W.; Devereux, R.; Stahl, D., Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and environmental microbiology* **1990,** *56* (6), 1919-1925.

37. Daims, H.; Brühl, A.; Amann, R.; Schleifer, K.-H.; Wagner, M., The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Systematic and applied microbiology* **1999,** *22* (3), 434-444.

38. Petriglieri, F.; Singleton, C. M.; Kondrotaite, Z.; Dueholm, M. K.; McDaniel, E. A.; McMahon, K. D.; Nielsen, P. H., Reevaluation of the Phylogenetic Diversity and Global Distribution of the Genus "Candidatus Accumulibacter". *mSystems* **2022,** *7* (3), e00016-22.

39. Daims, H.; Lücker, S.; Wagner, M., Daime, a novel image analysis program for microbial ecology and biofilm research. *Environmental microbiology* **2006,** *8* (2), 200-213.

40. Andrews, S., FastQC: a quality control tool for high throughput sequence data. Cambridge, United Kingdom: 2010.

41. Chen, S., Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2023,** *2* (2), e107.

42. Wood, D. E.; Lu, J.; Langmead, B., Improved metagenomic analysis with Kraken 2. *Genome biology* **2019,** *20*, 1-13.

43. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P. A., metaSPAdes: a new versatile metagenomic assembler. *Genome research* **2017,** *27* (5), 824-834.

44. Kang, D. D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z., MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from

metagenome assemblies. *PeerJ* **2019,** *7*, e7359.

45. Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **2015,** *25* (7), 1043-1055.

46. Chaumeil, P.-A.; Mussig, A. J.; Hugenholtz, P.; Parks, D. H., GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **2022,** *38* (23), 5315-5316.

47. Johnson, L. S.; Eddy, S. R.; Portugaly, E., Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC bioinformatics* **2010,** *11*, 1-8.

48. Edgar, R. C., MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* **2004,** *5*, 1-19.

49. Kozlov, A. M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A., RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019,** *35* (21), 4453-4455.

50. Letunic, I.; Bork, P., Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **2024**, gkae268.

51. Oyserman, B. O.; Noguera, D. R.; del Rio, T. G.; Tringe, S. G.; McMahon, K. D., Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. *The ISME journal* **2016,** *10* (4), 810-822.

52. Kong, Y.; Nielsen, J. L.; Nielsen, P. H., Microautoradiographic study of Rhodocyclus-related polyphosphate-accumulating bacteria in full-scale enhanced biological phosphorus removal plants. *Applied and environmental microbiology* **2004,** *70* (9), 5383-5390.

53. Nguyen, H. T. T.; Kristiansen, R.; Vestergaard, M.; Wimmer, R.; Nielsen, P. H., Intracellular accumulation of glycine in polyphosphate-accumulating organisms in activated sludge, a novel storage mechanism under dynamic anaerobic-aerobic conditions. *Applied and Environmental Microbiology* **2015,** *81* (14), 4809-4818.

54. Sharma, A. K.; Khandelwal, R.; Wolfrum, C., Futile cycles: Emerging utility from apparent futility. *Cell Metabolism* **2024**.

55. da Silva, L. G.; Gamez, K. O.; Gomes, J. C.; Akkermans, K.; Welles, L.; Abbas, B.; van Loosdrecht, M. C.; Wahl, S. A., Revealing the Metabolic Flexibility of "Candidatus Accumulibacter phosphatis" through Redox Cofactor Analysis and Metabolic Network Modeling. *Applied Environmental Microbiology* **2020,** *86* (24).

56. Brdjanovic, D.; Van Loosdrecht, M.; Hooijmans, C.; Mino, T.; Alaerts, G.; Heijnen, J., Effect of polyphosphate limitation on the anaerobic metabolism of phosphorus-accumulating microorganisms. *Applied microbiology biotechnology*

1998, *50* (2), 273-276.

57. Brdjanovic, D.; van Loosdrecht, M. C.; Hooijmans, C. M.; Mino, T.; Alaerts, G. J.; Heijnen, J. J., Bioassay for glycogen determination in biological phosphorus removal systems. *Water Science and Technology* **1998,** *37* (4-5), 541-547.

58. Oehmen, A.; Carvalho, G.; Lopez-Vazquez, C.; Van Loosdrecht, M.; Reis, M., Incorporating microbial ecology into the metabolic modelling of polyphosphate accumulating organisms and glycogen accumulating organisms. *Water research* **2010,** *44* (17), 4992-5004.

59. Rodrigo, M.; Seco, A.; Ferrer, J.; Penya-Roja, J., The effect of sludge age on the deterioration of the enhanced biological phosphorus removal process. *Environmental technology* **1999,** *20* (10), 1055-1063.

60. Whang, L. M.; Park, J. K., Competition between polyphosphate- and glycogen-accumulating organisms in enhanced-biological-phosphorus-removal systems: Effect of temperature and sludge age. *Water Environment Research* **2006,** *78* (1), 4-11.

61. Onnis-Hayden, A.; Majed, N.; Li, Y.; Rahman, S. M.; Drury, D.; Risso, L.; Gu, A. Z., Impact of solid residence time (SRT) on functionally relevant microbial populations and performance in full-scale enhanced biological phosphorus removal (EBPR) systems. *Water Environment Research* **2020,** *92* (3), 389-402.

62. Chitraju, C.; Mejhert, N.; Haas, J. T.; Diaz-Ramirez, L. G.; Grueter, C. A.; Imbriglio, J. E.; Pinto, S.; Koliwad, S. K.; Walther, T. C.; Farese, R. V., Triglyceride synthesis by DGAT1 protects adipocytes from lipid-induced ER stress during lipolysis. *Cell metabolism* **2017,** *26* (2), 407-418. e3.

63. Sharma, A. K.; Wolfrum, C., Lipid cycling isn't all futile. *Nature metabolism* **2023,** *5* (4), 540-541.

64. Zhang, Y.; Li, C.; Li, H.; Song, Y.; Zhao, Y.; Zhai, L.; Wang, H.; Zhong, R.; Tang, H.; Zhu, D., miR-378 activates the pyruvate-PEP futile cycle and enhances lipolysis to ameliorate obesity in mice. *EBioMedicine* **2016,** *5*, 93-104.

65. Adolfsen, K. J.; Brynildsen, M. P., Futile cycling increases sensitivity toward oxidative stress in Escherichia coli. *Metabolic engineering* **2015,** *29*, 26-35.

66. Liu, N.; Santala, S.; Stephanopoulos, G., Mixed carbon substrates: a necessary nuisance or a missed opportunity? *Current opinion in biotechnology* **2020,** *62*, 15-21.

7

OUTLOOK

Introducing this thesis, I explained how assumptions about what is fundamental in biology shapes not only our reasoning but also the direction of research itself. These assumptions influence how labs operate, what tools are developed, and ultimately, they set boundaries of what a scientist can explain. Personally, I find theories that view energy flows as fundamental to life more compelling than those that hold information molecules (such as DNA or RNA) as central to biology. This perspective, whether or not it holds as the sole truth, enables a more comprehensive exploration of cellular responses to their environments, without an overly causal focus on DNA.

In studying "*Ca.* Accumulibacter", I approached the subject from this less conventional perspective, aiming to understand the energetic limitations shaping these bacteria's lives. This thesis contains the development and application of both experimental and computational methods to deepen our understanding of "*Ca.* Accumulibacter" and to probe *what* they do and *why* they do it. More concretely, the previous chapters revealed that:

- Describing EPS from genetic potential alone is misleading for both individual species and microbial communities. Complex glycans and glycoconjugates are currently impossible to infer from genetic information alone and require metabolic analysis (**Chapter 2**).
- Metabolic models based on genetics provide a map of metabolites connections, while thermodynamics constrains the direction within this map, defining the possibilities within a given network. By implementing these constrains, the immense solution space of a PAOs metabolic network was considerably reduced to a handful of thermodynamically feasible operations (**Chapter 3**).
- Accounting for the cyclic nature of cells within their environment allows for the first time a consistent, assumption-free, prediction and quantitative understanding of metabolic phenotypes of "*Ca.* Accumulibacter" in an EBPR cycle (**Chapters 4**, **5**, and **6**).
- The dynamic reoccurrence of oxygen in EBPR systems leads to temporal separation of biosynthesis in "*Ca.* Accumulibacter" cells (**Chapter 5**).
- Complex interactions in metabolic networks are challenging to infer without mathematical tools, which reveal synergistic effects when "*Ca.* Accumulibacter" consumes multiple substrates simultaneously (**Chapter 6**).

Each chapter inevitably opens new questions and unknowns. While it's easy to become absorbed in specific details, stepping back to consider the broader

challenges in researching PAOs – and microorganisms in dynamic environments more generally – can help to guide the direction research should take and the pitfalls to avoid. Our understanding of these microbes depends on measurements taken at multiple biological levels, each with intrinsic limitations. In the following sections, we will explore these levels – DNA, RNA, proteins and metabolites – following the central dogma of biology, and examine the challenges unique to each. I then address the complexities in integrating these data into a comprehensive, quantitative understanding of microbial ecology and reflect on how rarely we fully consider dynamic environments in microbiology.

DNA: A Metagenomics Race at Full Speed

While writing this section, I attended the Long Read Sequencing in Upsala, Sweden. There I fully grasped the momentum of the "genomic-era" in research and witnessed firsthand the striking commercialization of DNA sequencing. The race is unmistakably in full swing. Yet, challenges remain. Even high-quality genomes are not entirely complete, and many species – including "*Ca.* Accumulibacter" – lack circular genome representation in databases. Efforts are underway, combining long- and short-read sequencing, to uncover the 'genomic dark matter' in these and other species. However, what remains hidden in these regions is yet to be known.

Genomic analysis offers valuable insights into the evolutionary history of a species, genera and families, but it is best suited to eukaryotes, where DNA recombination is primarily driven by sexual reproduction. Bacterial DNA, in contrast, is highly dynamic, with horizontal gene transfer often dominating as a source of recombination. This genetic fluidity reflects bacteria's energetic constraints; populations tend to favour smaller, adaptable genomes, enabling them to acquire beneficial genes from a communal gene pool. Yet, how horizontal gene transfer shapes bacterial populations under dynamic conditions is still largely unexplored.

Xie et al. [1] inferred gene horizontal transfer events in genomes of "*Ca.* Accumulibacter" and other PAOs, seeking to pinpoint the genetic basis for PAO traits. Similarly, Alder et al. [2] found high frequencies of mobile genetic elements and non-genomic vectors in "*Ca.* Accumulibacter". Plasmid DNA, which is often overlooked in metagenomic studies, may add significant genetic diversity to these environments. This could challenge the relevance of traditional species definitions, but that might be an exaggeration. Or maybe not? We simply don't know enough.

Finally, as discussed in Chapter 2, gene presence doesn't imply function, nor does absence confirm lack of function. This issue has been discussed in detail earlier and won't be elaborated further here. Nonetheless this is a crucial consideration when

relying solely on DNA data, which leads us to the next stage in the central dogma: RNA.

RNA: Molecules of Activity, right?

Cells produce various types of RNA, including messenger RNA (mRNA), which is often studied to infer microbial activity. The assumption is that mRNA levels directly reflect gene transcription, and thus meta-transcriptomics studies are growing in microbial ecology [3-6]. However, interpreting mRNA levels in slow-growing and dynamic systems presents challenges that have not been widely recognized. In fast-growing organisms, mRNA levels lead quickly to new proteins. But in slow-growing cells – especially in fluctuating conditions with temporal allocation of biosynthesis – how reliably does mRNA at a given time correlate with activity?

Several factors obscure the relation between mRNA levels and activity. For example, mRNA can undergo post-transcriptional modifications (PTcrpM) which can vary among species [7,8]. How these changes affect mRNA efficiency and stability, particularly in ecological microbes, remains unknown. This is one of the first barriers to linking mRNA levels with activity.

Additionally, mRNA must be translated into a protein, a process influenced by cellular resource availability (as an example, see Iwańska, et al. [9]). Chapter 5 suggested that *"Ca.* Accumulibacter" may favour biosynthesis to occur aerobically. How does this temporal separation affect the correlation between mRNA levels and proteins presents an open question, adding more uncertainty to transcriptomic data. In other words, are mRNAs present in the anaerobic phase translated into proteins as efficiently in the aerobic phase? Furthermore, slow-growing organisms exhibit gradual and slow proteomic changes which are orders of magnitude slower than that of mRNA lifetime. In this sense, mRNA shifts would affect only a fraction of the proteome.

Once synthesized, a protein may not be active or catalysing reactions as expected. This brings us to the next stage in the central dogma: proteins.

Proteins: Molecules that *Actually* Perform Functions

To understand a community member's activity, proteomic analysis is often more reliable than mRNA. Proteins are the molecules that "walk the walk", executing cellular functions. Consequently, proteo-genomics is an extremely powerful tool to

understand microbial abundance in terms of potential activity a given group could carry, and to understand reactions to environmental changes [10-14].

Like mRNA, proteins can undergo post-translational modifications (PTMs) that can significantly alter their structure and function [15]. While PTMs in bacteria have received less attention than in eukaryotes, they play a crucial role in environmental bacteria [16-19], especially for EPS as discussed in Chapter 2. The impact of PTMs on metabolic enzymes – and whether they can be reversibly modified in response to dynamic conditions – remains largely unexplored but could provide critical insights into bacterial physiology. Further research must focus on developing methods to detect changes in proteins within microbial communities [20] and to identify proteins that elude detection due to hydrophobicity [21].

Metabolites and Energy Carriers: The Ultimate Challenge

As proteins interact with and respond to metabolic fluxes, we move into the realm of metabolites. They are at the core of what makes a cell truly *alive*. DNA, RNA and proteins are studied largely because they eventually lead to metabolic activity. Yet, direct metabolic analysis remains underused, in part because of the complexity compared to other biomolecules[22,23]. In this sense, efforts should be put into advancing analytical methods to accurately measure metabolites and energy states within cells in ecology. Experiments using isotopically labelled substrates represent a great starting point in further developing these techniques and metabolic understanding [24-26].

In Chapter 2, measurements of glycans in the EPS of PAOs demonstrated that EPS composition is more complex than genetic inference alone suggests. However, methods to measure central metabolites, such as those in the TCA cycle or glycolysis, remain limited. Such metabolites are highly reactive and turnover quickly, requiring methods for immediate quenching, effective extraction and precise quantification techniques. This represents an opportunity for researchers to record the time dependent metabolic state of cells within an environment, which ultimately could explain the *why* of phenotypic behaviours.

In Chapter 3 mathematical tools led to identifying how TCA cycle and glycolysis metabolite concentrations would affect metabolic direction and driving forces. In this sense, accurately quantifying these metabolites – including energy careers such as ATP and NADH – during an EBPR cycle could conclusively indicate the directionality of metabolic reactions over time. A similar study in yeast cells found that indeed metabolic changes – independent of gene expression – predominantly influenced glycolytic direction and flux [27]. Chapter 5 predicted that PAOs biosynthesis might be

temporally separated throughout the cycle, and measuring energy carriers like ATP and NAD(P)H could yield more concrete answers. Nevertheless, first the analytical methods for this quantification need to be advanced.

Integrating omics- data for a quantitative understanding of metabolism

Achieving a quantitative understanding of microbial metabolism within complex environments requires integrating multi-omics data – (meta)genomics, (meta)transcriptomics, (meta)proteomics and (meta)metabolomics[28]. Each layer provides unique insights: genomics reveals ancestry and metabolic potential, transcriptomics shows gene activation as a consequence of the environment, proteomics reveals to what extent that gene activation was successful and what a microbe can actually perform, and metabolomics captures the real-time intricacies of metabolism. However, using these data types in isolation is insufficient to identify fundamental and testable hypotheses to further our understanding of complex systems like PAOs in dynamic environments.

Efforts are already being made with research integrating genomics-transcriptomics [1,4,5] or genomics-proteomics [29] data. Nevertheless, very little has been done in the forefront of transcriptomics-proteomics and even less with proteomics-metabolomics. We need robust frameworks that are able to merge different levels of datasets allowing us to view causes and consequences of microbial activity more comprehensively. Researching the dynamic interplay between different levels could lead to novel insights into how a given phenotype is observed and, with views into the future, how we can precisely manipulate a microbial community to achieve a desired outcome.

Recurrent dynamic conditions: much more than PAOs

While "*Ca.* Accumulibacter" has been the focus of this thesis, the insights can extend towards general microbial survival strategies in dynamic environments. Natural microbial communities rarely face static conditions, instead encountering fluctuating resources, oxygen levels, and environmental shifts. These dynamics many times require adaptations that go beyond genetic control, due to the incompatibility between the change itself and the time for reaction from genomic control. It is thus interesting to point out that the strategy "*Ca.* Accumulibacter" employs to drastically switch metabolism – which is metabolic control – may be employed by many other

microorganisms. The only requirement is that the change in the environment is faster than that of protein turnover. And this can be found in nearly every environment I can think of – from oceans that experience sudden increase in temperature from a hydrothermal vent, to soils where a massive input of fertilizers are applied by a farmer. From wastewater treatment plants that suddenly receive massive flows of water during the mid-term of the world cup final, to a cell within a controlled industrial bioreactor that suddenly gets pushed by a massive impeller towards a substrate rich region within the broth. We must always consider that cells react *metabolically* to a given change before they can even attempt to react *genetically*. Genetic control, in this sense, is an effect rooted on metabolic change.

References

1 Xie, X. *et al*. Integrated genomics provides insights into the evolution of the polyphosphate accumulation trait of Ca. Accumulibacter. *Environmental Science and Ecotechnology* **20**, 100353 (2024).

2 Adler, A., Poirier, S., Pagni, M., Maillard, J. & Holliger, C. Disentangle genus microdiversity within a complex microbial community by using a multi-distance long-read binning method: example of Candidatus Accumulibacter. *Environmental microbiology* **24**, 2136-2156 (2022).

3 Cardona, L., Saini, J. S., Rodilla Ramirez, P. N., Adler, A. & Holliger, C. Multiple extracellular polymeric substances pathways expressed by Accumulibacter and the flanking community during aerobic granule formation and after influent modification. *bioRxiv*, 2024.2009. 2004.611157 (2024).

4 Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. *The ISME journal* **10**, 810-822 (2016).

5 McDaniel, E. A. *et al*. Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics. *Msystems* **6**, e00474-00421 (2021).

6 He, S. *et al*. Metatranscriptomic array analysis of 'Candidatus Accumulibacter phosphatis'-enriched enhanced biological phosphorus removal sludge. *Environmental microbiology* **12**, 1205-1217 (2010).

7 Menendez-Gil, P. & Toledo-Arana, A. Bacterial 3' UTRs: a useful resource in post-transcriptional regulation. *Frontiers in molecular biosciences* **7**, 617633 (2021).

8 Mohanty, B. K. & Kushner, S. R. The majority of Escherichia coli mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic acids research* **34**, 5695-5704 (2006).

9 Iwańska, O. *et al*. Translation in Bacillus subtilis is spatially and temporally coordinated during sporulation. *Nature Communications* **15**, 7188 (2024).

10 Pabst, M. *et al*. A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. *The ISME Journal* **16**, 346-357 (2022).

11 Kleikamp, H. B. *et al*. Comparative metaproteomics demonstrates different views on the complex granular sludge microbiome. *bioRxiv*, 2022.2003. 2007.483319 (2022).

12 Kleikamp, H. B. *et al*. Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. *Water research* **246**, 120700 (2023).

13 Wexler, M., Richardson, D. J. & Bond, P. L. Radiolabelled proteomics to determine differential functioning of Accumulibacter during the anaerobic and aerobic phases of a bioreactor operating for enhanced

biological phosphorus removal. *Environmental microbiology* **11**, 3029-3044 (2009).

14 Kleiner, M. Metaproteomics: much more than measuring gene expression in microbial communities. *Msystems* **4**, e00115-00119 (2019).

15 Duan, G. & Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS computational biology* **11**, e1004049 (2015).

16 Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology* **8**, 765-778 (2010).

17 Schuster, B. & Sleytr, U. B. Relevance of glycosylation of S-layer proteins for cell surface properties. *Acta biomaterialia* **19**, 149-157 (2015).

18 Van Teeseling, M. C. *et al*. The S-layer protein of the anammox bacterium Kuenenia stuttgartiensis is heavily O-glycosylated. *Frontiers in microbiology* **7**, 1721 (2016).

19 Li, H. *et al*. Understanding protein glycosylation pathways in bacteria. *Future Microbiology* **12**, 59-72 (2017).

20 Virág, D. *et al*. Current trends in the analysis of post-translational modifications. *Chromatographia* **83**, 1-10 (2020).

21 de Jong, S. I., Sorokin, D. Y., van Loosdrecht, M. C., Pabst, M. & McMillan, D. G. Membrane proteome of the thermoalkaliphile Caldalkalibacillus thermarum TA2. A1. *Frontiers in Microbiology* **14**, 1228266 (2023).

22 Pinu, F. R., Villas-Boas, S. G. & Aggio, R. Analysis of intracellular metabolites from microorganisms: quenching and extraction protocols. *Metabolites* **7**, 53 (2017).

23 Van Gulik, W. M. *et al*. Fast sampling of the cellular metabolome. *Microbial Systems Biology: Methods and Protocols*, 279-306 (2012).

24 Chokkathukalam, A., Kim, D.-H., Barrett, M. P., Breitling, R. & Creek, D. J. Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis* **6**, 511-524 (2014).

25 Maurer, M., Gujer, W., Hany, R. & Bachmann, S. Intracellular carbon flow in phosphorus accumulating organisms from activated sludge systems. *Water Research* **31**, 907-917 (1997).

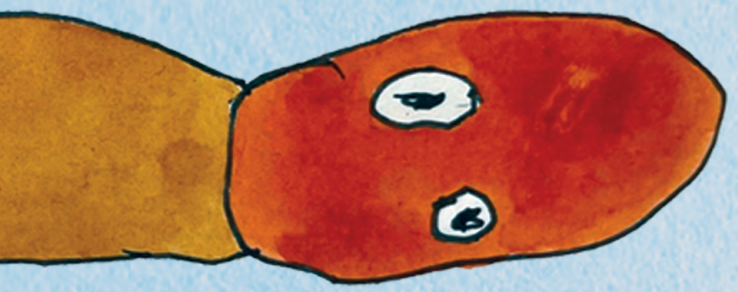
26 Buckstein, M. H., He, J. & Rubin, H. Characterization of nucleotide pools as a function of physiological state in Escherichia coli. *Journal of bacteriology* **190**, 718-726 (2008).

27 Daran-Lapujade, P. *et al*. The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. *Proceedings of the National Academy of Sciences* **104**, 15753-15758 (2007).

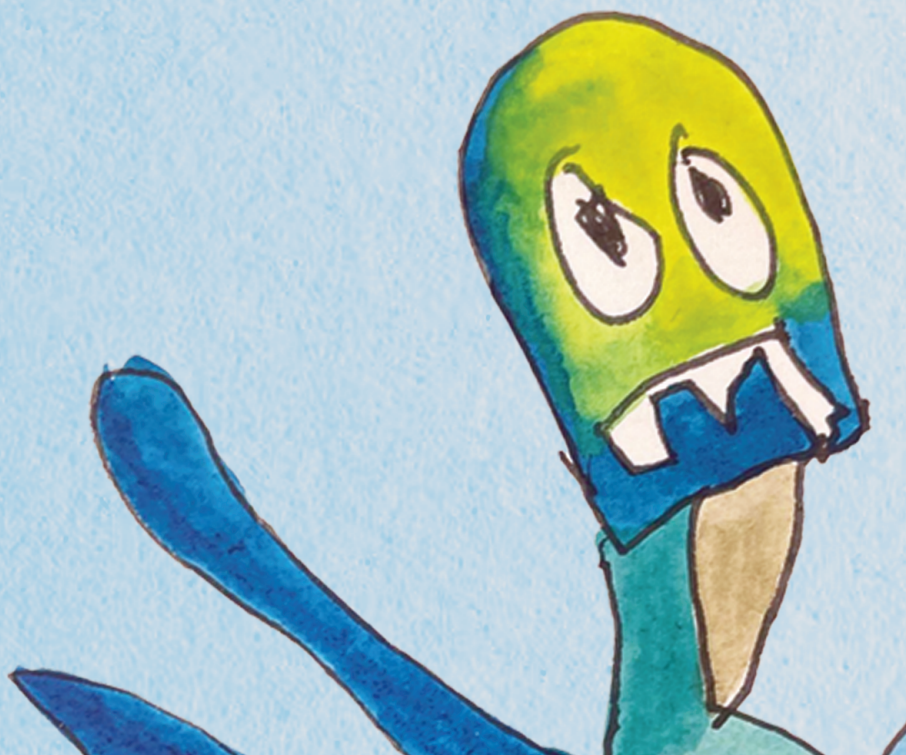
28 McDaniel, E. A. *et al*. Prospects for multi-omics in the microbial ecology of water engineering. *Water Research* **205**, 117608 (2021).

29 Wilmes, P. *et al*. Community proteogenomics highlights microbial strain-variant protein expression

within activated sludge performing enhanced biological phosphorus removal. *The ISME journal* **2**, 853-864 (2008).



ACKNOWLEDGEMENTS



Acknowledgements

Luck

The odds of our existence are low – so low that there's even a theory suggesting it's more probable that a self-aware brain spontaneously forms, complete with your memories, before instantly disintegrating. Yet here I am, incredibly lucky to live this journey. Luck has contributed enormously to my life, and this section is dedicated to acknowledging it.

I started off this journey with so much luck to be in Environmental Biotechnology (EBT). I had no idea what I was getting into when I accepted the PhD. I just loved the idea of being in university and having free coffee with my employee card. But it turned out to be the coolest place I've ever worked in – and a lot of that has to do with Mark, thanks for caring about us as people, not just researchers. Your feedback—especially on my social efforts—pushed me to organize more pub quizzes than I ever imagined. I feel lucky to have had you as a supervisor, and I hope I helped keep EBT a fun place.

On the other corner of my supervisory table, we have Aljoscha. At first, it seemed quite strange – worrying, even – that you were leaving Delft. How was I supposed to do a PhD without my main supervisor? But it was a blessing in disguise. Your enthusiasm only grew stronger from Germany. I guess being in a place where the Jumbo-bridge isn't the biggest mountain helps – especially for a fellow climber. Thanks for believing in my work even when I didn't, for pushing me to make the cFBA toolbox, and for even teaching it and sharing it with other people. And, of course, thank you so much for proposing my name to teach AMN when you left. You helped me discover my true passion in life.

During my PhD, I was fortunate to have the freedom to collaborate with anyone I wanted. There were many, and I am thankful to all of you for sharing ideas, comments and patience with me. Ricardo, tal vez no lo sepas, pero mi PhD se soporta en tus hombros. Ñaño, gracias por tu ayuda con Python y con cFBA. Yuemei and Sergio, thanks for letting me jump on a project on our sweet secrets (which does not mean gossip time!). Martin and Jitske, thanks for your incredible work with the metabolomics project. Although not in this thesis, it will be a fantastic piece of work! Karel, gracias por ser una persona tan curiosa e interesante. Conversar contigo en una tarde es equivalente a leer diez artículos

científicos, dos libros de historia y una novela. Estoy muy agradecido de haber podido contribuir a tu proyecto de RubisCO. Samarpita, for introducing a new substrate into my PAOs work. Siem, although our collaboration was short, you always amazed me with the deep level of understanding you have on, well, everything. I was lucky enough to learn some of your wisdom. Finally, Yubo, I am so grateful to be able to work with such a caring and lovely person. People like you make me trust this world.

Now, since I'm a bit of an educator myself, let's get into the world of student-supervisor relationships. Being a supervisor was very dear to me. I feel very lucky because I think very few get to experience the type of connection I felt with my students – of course, sometimes a bit of drama is involved, but yet… Firstly, Daan with your good vibes and honesty. Together, we got through very stressful moments in our lives and I thank you for that. Robin, the student that always disappeared! Thanks for teaching me that pure cultures are not the way to go in EBT! Bea (short version of a long complex Portuguese name), I was lucky enough to have you work on cFBA of PAOs. You were the first to teach me, well, organization in the broadest sense. I'm thrilled that we are now colleagues.

After this bunch, I worked with Jim, 'the lonely wolf'. Thanks for being so ambitious that you got me started into learning thermodynamics for your BEP – without any biological background! I ended up doing quite a lot of thermodynamics later in my PhD. Next up, I worked with two very opposite personalities: Maaike and Rene. This probably was the first time I realized I needed to adapt my style – relaxed yet strict, fun yet serious. Thanks for pushing me to both sides. Maaike, I still enjoy walking in the forest to calm down, and I learnt it from you. Rene… you are, by far, the most genuinely unconventional person I have ever met. Thanks for the time we worked together. And Timo, Rene's calmer hippy brother. I remember thinking both of you belonged in a music band when I met you in your first year. Thanks for doing your BEP with me; I learned more from you than you from me. And to both of you, thanks for bringing me a Christmas tie to class! Tobias, you came to show me that BEP students can be even more mature and responsible than myself. Thanks for your hard work in the lab and for helping me with my teacher training.

Soon, I was lucky to work with students that could bridge collaborations with different partners. Casper, you fit perfectly into the role of managing two sometimes frustratingly different views. Thanks for integrating them so well, just like you did with the lab and models. Lilo, although not directly supervised by others, you experienced the stress of receiving strong opinions from many different people. Thanks for having the strength and patience to handle it.

Roeland, Sergio and I were lucky to have you on board. It was a lot of fun to see you become super responsible and independent in your work – and I still need to beat you at tennis. Dafni, thanks for your enthusiasm and strong presence. Any group working with you will be lucky to have you. I still look at your cat drawing almost every day at work. And how could I forget the youngest, coolest Gen Z ever, Scott? Thanks for sharing some of your style and youth with EBT. Yubo and I were very lucky to work with you.

Finally, my personal PhD projects with students who connected with me through my teaching. Isa, I am so lucky to have worked with you. Although we are very opposite in many ways, things just worked so smoothly. Maybe that's how the life of organized people is. Thanks for letting me experience it for 9 months! The reactors never recovered from your absence, even after cleaning them with a 'flessenlikker'. Next, a 360 degrees change – Loek. Uniquely artistic, perfectly chaotic. Thanks for all the great times we had together. Dinner and drinks with your family in Wageningen was a unique experience – reinforcing the idea that these strong connections are very rare.

To all my students, I was honored to work with you all. I will always be available for a karaoke night. You can count on me for the first round of drinks and songs.

When it comes to teaching, I was lucky enough to have Jack and Ulf supporting my journey. Thank you both for trusting me and helping me apply for the Van Rijn teaching position. While teaching, I had the great fortune of working with the incredibly capable hands of teaching assistants: Effie & Nick, Nina & Esther, Nuria & Mehrab, and Marco & Uros. Thanks to each powerful duo for making AMN fun and, especially, for helping me spot my mistakes and endless typos*. Djordje, my AMN-apprentice! Gracias por tu ayuda, por enseñarme nuevas técnicas sobre modelos y, mas que nada, por ser tan chévere. AMN queda en grandes manos.

C1.380. If anything is improbable, that is! I started my PhD in an office with Stefan, my half-Latin, karaoke-addicted brother. Thanks for being there for almost six years! I still don't fully comprehend how fast you forget stuff, but I love it. In my mind, you are the oldest flatmate I´ve lived with. Then came Francesc, el Kiko-Corbera, from Delft to the world. What better way to complete our pseudo-Latin trio than with this blunt, hilariously inappropriate and injury-prone personality? Together, we made the office our own – filled with clothes, towels, lab material, a deflated volleyball, broken badmington rackets, shooting unicorns, and more. We thought we had it all. Until the revolution happened – Rodoula, our divine

* Here's a quick game. For every new typo that is found, I owe a drink to the discoverer.

intervention. You changed absolutely everything for us. The office finally felt like a small yet warm family. I loved every day I went to work, and I was lucky I could share 90 % of my time with you guys.

Outside of the office, I also was fortunate enough to make great EBT-friends. We were privileged enough to do part of the PhD during lockdowns. Who needed the botanical when we had each other! I cherish those memories and thank you all. Maxim, your unnecessarily responsible finishing of the PhD left many stories untold. Thanks for all the long conversations sitting on the office floor, all the gossip, and your unique ability to interrogate Sam and Ali. Chris, your responses always seem delayed—probably because your neurons have such a long way to travel! Samuel, your unique charisma that allowed us to interrogate your intimate life. Sergio, we made the best trio living with Leo. Thanks for making my PhD fun. I can still hear the echoes of your voice in the corridors, especially when someone's name is mentioned. Who was it? Lemin, thanks for being so kind and awkward when hugging – I know you try your best. Ali, thanks for your help when discussing against all the U.S. haters. And on that note, David, thanks for being so annoying and winning every competition. I don't know how you do it, but you sure well know how to brag about it. Marçelo, always conflicted with your daily apple consumption. Rootless, always trying unsuccessfully to escape long lunches. That won't work! Life is short.

Although not entirely EBT, my EBT-related friends extended to very lovely people. Mariana, thanks for being so incredibly plastic, fitting into any group. Every time our eyes met, I knew I was in for a wave of long-needed-gossip. Thanks, Άγγελος, for our first handshake and hug – you are the warmest person I've ever met. Thanks Jan – you might not know this, but you radically changed my work-life balance with the Bullet Journal. Matteo, thanks for always being so encouraging in everything, from Dutch to climbing.

To many other old and new BT members, thanks for sharing these years with me and for taking 10 extra minutes at lunch to have a nice chat. Alexandra, Bea, Chiara (the chemister), Claudia, Georg, Heike, Hugo, Ismail, Jelle, Jelmer, Ji, Jisk, Jitske, João, Jules, Jure, Linghang, Luz, Mario, Marit, Marta, Martijn (fellow 14 L beer drinker), Marte, Natalia (the almost palindrome), Puck, Ramon, Rebeca, Robbert, Sebastian, Sirous and Sofia. And of course, a gigantic THANK YOU to Ben-Dita-Zita (in Spanish: blessed Zita), Dirk, Kevin, Miranda, Apilena, Jannie, Gea, Kawieta and Nayyar for all your help in and out of the lab. None of the work done in any PhD would be possible without you.

As a foreigner in the Netherlands, I only get to see my family once a year. But I had the pleasure and great fortune to find my very own non-blood related family here in Europe. Starting with Álvaro –you would never forgive me if I didn't. Gracias por ser un canelo, un menudo canelo. Tuve la enorme suerte de contar con tu amistad desde el inicio del master, escuchando chichi peralta e inventándonos juegos con los nombres de LST. Juntos hacemos los mejores tres latinos. Ricardo, *odio* que no estés aquí en Holanda, pero *amo* cada momento cuando nos vemos. Aunque nos engañes para no jugar *Overcooked* o llevarnos a museos, eres mi ñaño del alma. Melina, nuestra diosa griega. I never had a sister, but our same sense of humor and constant small fights makes me feel like I do. Thanks for everything – I miss you living nearby. Jorge, gracias por: *"que pedos, esa maje si que es pija de pajera maje lampiña lempira!"* Gracias por tu amistad incondicional y por hacerme sentir que mi carbon footprint no es tan grande. Sara, gracias por tu honestidad y por aprenderte todos los nombres de LST para poder reírnos juntos.

My paranymphs, I purposely left you down here to give you a little scare. Nina, I am so lucky to be your friend. Thanks for that. You have been with me through so many good moments, but also through tough, sad, and difficult ones. Thank you for being that voice of support, reason, and guidance when I needed it, and for always believing in me. And you're welcome for introducing you to Leiden – I'm so glad we are neighbors! Rodoula, how fortunate that you, out of anyone, would start a PhD in Biotechnology? You managed to finish your PhD without knowing what biomass *is*. You earned the most paranymph points ever—counting both positive and negative. But more importantly, you filled my journey with light. Thanks for your everyday smile. In case you didn't know, it makes everyone around you happy. So do your never-ending changing hair styles, your jokes and your personality. To both my lovely paranymphs, thanks for accepting this difficult long and frustrating job! And equally for trusting me to be yours. I couldn't be happier (or busier :p).

To my parents, to who (or whomst? I'm not using ChatGPT here!) I owe my strong personality. Jools, thanks for making me the *livianito*, easygoing, spontaneous and fun person I am – and extraordinarily humble too it seems – *weheey!* I try every day to be a bit more like you. Your kids are great teachers because we learnt from, nothing short from the best storyteller in history! And thanks for your amazing drawings in this book. Patuleco, *Dios te pague, hermano*. Tú me has dado el mayor ejemplo de ser una persona trabajadora, fuerte y luchadora. Al pasar los años, me doy cuenta de que me parezco más y más a ti. Me enseñaste que hacer deporte es súper importante, al igual que tratar de siempre dar lo mejor de ti. No

seré siempre el mejor, pero intento dar lo mejor de mí, y eso me lo enseñaste tú. A mis hermanos: "*Alex, deja de fumar. Timmy, Charlie, nunca fumen*". ¡Si tan solo supiera el Patu! Gracias por todo. Siempre intento aprender canciones en el piano o en la guitarra pensando que algún día les haré escuchar. Pasar tiempo con ustedes y sus nuevos núcleos familiares me recuerda lo que es importante en esta vida. Gracias Igorete por tus diseños para este libro. Y a ti y a la Pandorita por escuchar mis historias. Graicas Ari y Lau, por todos los buenos momentos que hemos pasado juntos. Mis últimos días en Ecuador siempre son los mejores gracias a ustedes, a nuestros partidos de fútbol, bailes y parrilladas. Grandsil and Dawn, thanks for taking me in for some weeks every year to spend some time being truly spoilt in England. A mi casi-familia Felipollito, Gutto, Merino y Sebas gracias por siempre hacer el esfuerzo de verme cuando estoy de visita. ¡Oh como me encantan mis ventanas!

And last, the one person who supported me the most on this journey – my lovely little Vends. This, for sure, must be the self-assembled brain I mentioned at the beginning of the section, because having a life so perfect with you seems truly impossible. I've always said that the PhD is a fun and not-so-stressful journey. This is hugely biassed, of course, because I speak from the privilege of having my best friend by my side at every step of the way.
My best friend who rewatches with me every joke I think I made during AMN.
My best friend who spins weird theories with me about everything.
My best friend who plays along with all my stupid little guessing games.
My best friend who gets me, who makes me laugh, who makes me happy.
My best friend who – though it's hard to admit – is the creative mastermind behind all the games and videos I've made.
I'm so lucky that I live with my best friend.

Thank you, universe of endless probabilities, for giving me this self-aware brain filled with all these lovely memories.

Publications

Paez-Watson, T., Jansens, C., van Loosdrecht, M. C. M., & Roy, S. (2025). Metabolic implications for dual substrate growth in “Candidatus Accumulibacter”. bioRxiv.

Páez-Watson, T.,Tomás-Martínez, S., de Wit, R., Keisham, S., Tateno, H., & Wahl, S. A. (2024). Sweet secrets: Exploring novel glycans and glycoconjugates in the extracellular polymeric substances of “Candidatus Accumulibacter”. ACS ES&T Water, 4(8), 3391-3399.

Páez-Watson, T.,Hernández Medina, R., Vellekoop, L., van Loosdrecht, M. C. M., & Wahl, S. A. (2024). Conditional flux balance analysis toolbox for Python: Application to research metabolism in cyclic environments. Bioinformatics Advances, 4(1), vbae174.

Páez-Watson, T., van Loosdrecht, M. C. M., & Wahl, S. A. (2024). From metagenomes to metabolism: Systematically assessing the metabolic flux feasibilities for “Candidatus Accumulibacter” species during anaerobic substrate uptake. Water Research, 250, 121028.

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 Research, 228, 119365.

Páez-Watson, T., Álvarez-Suárez, J. M., Rivas-Romero, F., Estrada, L., López, D., & Cabrera, J. C. (2020). Increased salinity stress tolerance of Nicotiana tabacum L. in vitro plants with the addition of xyloglucan oligosaccharides to the culture medium. In Vitro Cellular & Developmental Biology-Plant, 56, 325-334.

González-Pérez, L., Páez-Watson, T., Álvarez-Suárez, J. M., Rivas-Romero, F., Estrada, L., López, D., & Cabrera, J. C. (2018). Application of exogenous xyloglucan oligosaccharides affects molecular responses to salt stress in Arabidopsis thaliana seedlings. Journal of Soil Science and Plant Nutrition, 18(4), 1187-1205.

Alvarez-Suarez, J. M., Giampieri, F., Gasparrini, M., Mazzoni, L., Santos-Buelga, C., González-Paramás, A. M., Forbes-Hernández, T. Y., Afrin, S., Páez-Watson, T.,Quiles, J. L., & Battino, M. (2017). The protective effect of acerola (Malpighia emarginata) against oxidative damage in human dermal fibroblasts through the improvement of antioxidant enzyme activity and other mechanisms. Food & Function, 8(9), 3250-3258.

Manuscripts in Preparation

van Ede, J.*, Páez-Watson, T.*, van Loosdrecht, M. C. M., & Pabst, M. Temporal variations of energy metabolites in granular biomass of phosphate-accumulating microbial communities. *Shared co-authorship

Olavarria, K., Roy, S., van Ede, J., Páez-Watson, T., Pabst, M., & van Loosdrecht, M. A new glycolytic route involving Rubisco in heterotrophic phosphate-accumulating bacteria [Manuscript in preparation].

About the Author

Timothy is a dedicated researcher and educator with broad expertise in microbiology, environmental biotechnology and systems biology. Born in Quito in 1994, he pursued a BSc in Biotechnology Engineering at Universidad de las Américas (UDLA) in Ecuador, where he gained experience in plant biotechnology for his bachelor's thesis. Following this, he worked as a technician in UDLA's research laboratories under Dr. Lien Gonzalez and Dr. José Miguel Álvarez-Suarez, focusing on the antioxidative properties of natural products in mammalian cell cultures.

In 2017, Timothy moved to Delft, Netherlands, to pursue an MSc in Life Science and Technology with a specialization in Biocatalysis and Cell Factory. During this time, he participated in the iGEM synthetic biology competition and gained hands-on experience in molecular techniques related to producing, purifying, and testing novel fusion proteins for targeted next-generation sequencing. He later shifted his focus to studying *Escherichia coli* adaptation to microaerobic conditions in pure culture chemostats for his MSc thesis under the supervision of Dr. Duncan McMillan.

In 2020, he began his PhD at TU Delft in the Environmental Biotechnology group under the supervision of Prof. Dr. Aljoscha Wahl and Prof. Dr. Ir. Mark van Loosdrecht. His research focuses on dynamic metabolic modeling, resource allocation models, and open culture bioreactors for microbial communities involved in phosphorus removal from wastewater.

Alongside his research, Timothy has a strong passion for teaching. He began teaching at a young age as a high school tutor in math and chemistry, a role that continued at UDLA in Ecuador, where he lectured in Biochemistry and Molecular Techniques. Currently, he is a Junior Teacher at TU Delft, where he designs and teaches the Analysis of Metabolic Networks course.