

# Metagenomic and metaproteomic analyses of *Accumulibacter phosphatis* enriched floccular and granular biofilm

Jeremy J. Barr<sup>1,2,3\*</sup>, Bas E. Dutilh<sup>4,5,6</sup>, Connor T. Skennerton<sup>2,7,8</sup>, Toshikazu Fukushima<sup>2,9</sup>,  
Marcus L. Hastie<sup>10</sup>, Jeffrey J. Gorman<sup>10</sup>, Gene W. Tyson<sup>2,7</sup>, Philip L. Bond<sup>2,3</sup>

<sup>1</sup> *San Diego State University, Department of Biology, San Diego, CA 92182, USA*

<sup>2</sup> *The University of Queensland, Advanced Water Management Centre (AWMC), QLD 4072, Australia*

<sup>3</sup> *Environmental Biotechnology Cooperative Research Centre (EBCRC), Sydney, Australia*

<sup>4</sup> *Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands*

<sup>5</sup> *Centre for Molecular and Biomedical Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Centre, Greet Grooteplein 28, 6525 GA, Nijmegen, The Netherlands*

<sup>6</sup> *Department of Marine Biology, Institute of Biology, Federal University of Rio de Janeiro, Brazil*

<sup>7</sup> *The University of Queensland, Australian Centre for Ecogenomics, School of Chemistry and Molecular Bioscience, QLD 4072, Australia*

<sup>8</sup> *Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>9</sup> *Division of Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan*

<sup>10</sup> *Protein Discovery Centre, Queensland Institute of Medical Research (QIMR) Berghofer Medical Research Institute, Herston, QLD 4006, Australia*

\*Corresponding author.

Jeremy J. Barr

Department of Biology,

San Diego State University,

San Diego, CA 92182, USA

jeremybarr85@gmail.com

Running Title: Meta-omics of floccular and granular biofilms

---

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.13019

This article is protected by copyright. All rights reserved.

## Abstract

Biofilms are ubiquitous in nature, forming diverse adherent microbial communities that perform a plethora of functions. Here we operated two laboratory-scale sequencing batch reactors enriched with *Candidatus Accumulibacter phosphatis* (*Accumulibacter*) performing enhanced biological phosphorus removal (EBPR). Reactors formed two distinct biofilms, one floccular biofilm, consisting of small, loose, microbial aggregates, and one granular biofilm, forming larger, dense, spherical aggregates. Using metagenomic and metaproteomic methods we investigated the proteomic differences between these two biofilm communities, identifying a total of 2,022 unique proteins. To understand biofilm differences, we compared protein abundances that were statistically enriched in both biofilm states. Floccular biofilms were enriched with pathogenic secretion systems suggesting a highly competitive microbial community. Comparatively, granular biofilms revealed a high stress environment with evidence of nutrient starvation, phage predation pressure, and increased extracellular polymeric substance (EPS) and cell lysis. Granular biofilms were enriched in outer membrane transport proteins to scavenge the extracellular milieu for amino acids and other metabolites, likely released through cell lysis, to supplement metabolic pathways. This study provides the first detailed proteomic comparison between *Accumulibacter*-enriched floccular and granular biofilm communities, proposes a conceptual model for the granule biofilm, and offers novel insights into granule biofilm formation and stability.

## Introduction

Biofilms are aggregates of microorganisms that may attach to a surface or exist as suspended biomass. In many environments, bacteria preferentially exist within these sessile biofilm communities rather than as planktonic and free-swimming counterparts (Costerton *et al.*, 1987; Davies *et al.*, 1998). Biofilms are characterized by the environmental conditions, surfaces, and species required for their formation, and provide numerous benefits to their microbial inhabitants (O'Toole *et al.*, 2000; Watnick & Kolter, 2000). Biofilm structure can be highly varied and complex, with the microbial inhabitants embedded within a matrix of extracellular polymeric substances (EPS). EPS are defined as organic polymers of microbial origin, consisting of but not limited to polysaccharides, proteins, and lipids. Together, EPS are responsible for binding cells, water, and other particulate matter together (Wingender *et al.*, 1999). Biofilms can be of serious clinical and commercial importance. An estimated 60% of nosocomial infections manifest as biofilms, and significant commercial efforts are underway to control and manipulate biofilms in situations that range from biofouling on ships to the treatment of wastewater (Fridkin *et al.*, 1997; Schultz, 2007; Nicolella *et al.*, 2000).

Biotechnological applications of biofilms are critical in the treatment of wastewater, whereby microbial aggregates remove and degrade waste nutrients from the wastewater, allowing the safe release of the treated water back into the environment (Stephenson *et al.*, 2000; Cornel & Krause, 2006). Activated sludge wastewater treatment processes can enrich various bacterial types to achieve the biological removal of carbon, nitrogen and phosphorus nutrients (Fuhs & Chen, 1975; Blackall *et al.*, 2002). Enhanced biological phosphorus removal (EBPR) is a desired activated sludge treatment process whereby microorganisms remove inorganic phosphate (Pi) and carbon from wastewater. By use of laboratory-scale sequencing batch reactors (SBRs) many well-controlled experiments have been performed to reveal ecosystem insight of these biofilms performing EBPR (He *et al.*, 2006; García Martín *et al.*, 2006). EBPR is mediated by groups of bacteria known as polyphosphate accumulating organisms (PAOs) (Hesselmann *et al.*, 1999; Kong *et al.*, 2005). *Candidatus Accumulibacter phosphatis* (Accumulibacter) is identified as a principal PAO within laboratory-scale studies of EBPR reactors (Bond *et al.*, 1995; Hesselmann

*et al.*, 1999). Previous metagenomic studies have elucidated metabolisms from a number of *Accumulibacter* species, providing a systems level understanding of the microbiology behind EBPR processes (García Martín *et al.*, 2006; Flowers *et al.*, 2013; Skennerton *et al.*, 2014; Mao *et al.*, 2014). Moreover, these metagenomic studies provide the genetic blueprints for subsequent transcriptomic and proteomic analyses of *Accumulibacter* enriched microbial communities, identifying core metabolisms and pathways utilized in EBPR (Wilmes, Wexler, *et al.*, 2008; Wilmes, Andersson, *et al.*, 2008; He & McMahon, 2011a). This has ultimately led to an unprecedented level of understanding of the microbial community function within EBPR wastewater treatment and the optimization of full-scale treatment processes (Nielsen *et al.*, 2012).

Within conventional EBPR the activated sludge is typically small, suspended, biofilm aggregates, with diameters of the range 30-200  $\mu\text{m}$  that are called floccular sludge or 'Flocs'. Following the passage through anaerobic and aerobic reaction phases, the floccular sludge is separated from the treated wastewater by gravitational settling. The concentrated floccular sludge is then returned to the reaction stages for the treatment of new incoming wastewater. This biomass separation and return is characteristic of activated sludge wastewater treatment, and the repeated passage through the wastewater treatment stages provides alternating conditions essential for the EBPR process. Floccular activated sludge is a tried and true process used for wastewater treatment over the past 100 years (Mino *et al.*, 1998).

Aerobic granular sludge has recently emerged as an alternative biofilm technology for wastewater treatment (Liu & Tay, 2004; de Kreuk *et al.*, 2007). Like Flocs, aerobic granules are a suspended biofilm of mixed-microbial communities. However granular biofilms are significantly larger (200-2000  $\mu\text{m}$ ), with the microbial cells embedded in large amounts of EPS. As such they exhibit superior settling capacities than Flocs, enabling both longer wastewater treatment times and a reduced ecological footprint (Morgenroth *et al.*, 1997; Beun *et al.*, 1999). There is significant commercial interest in the application of granular sludge for wastewater treatment processes (Winkler *et al.*, 2011; Weissbrodt *et al.*, 2013), but questions regarding granular biofilm formation, stability, and reproducibility remain.

Here we operated two laboratory-scale EBPR reactors, one as a floccular biofilm, and one that transitioned from a floccular biofilm through to a completely granular biofilm. Both reactors were enriched with *Accumulibacter*, performing comparable levels of EBPR performance. Reactors were sampled sequentially throughout their operation for both total cellular and extracellular proteins to investigate the microbial community-level differences in protein expression between the two biofilms. Proteins were identified by liquid chromatography-mass spectrometry (LC-MS) analysis of the entire microbial community. Moreover, to enable accurate protein identification, shotgun metagenomes from the floccular and granular reactors were attained together with the final protein extraction. The sequential sampling of a stable floccular biofilm and a developing granular biofilm, both enriched with *Accumulibacter*, provided proteomes to identify proteins that were significantly enriched in each biofilm state. Together, we provide the first direct proteomic comparison between a floccular and granular, *Accumulibacter*-enriched biofilm community and provide biological insights into granule biofilm formation, stability, and metabolism. The results presented here will lead to a better understanding of granule biofilms and may impact their application to full-scale wastewater treatment processes.

## Materials and Methods

### ***Reactor operations***

Two laboratory-scale sequencing batch reactors (Floc and Gran) were seeded with floccular activated sludge from a domestic WWTP operating for EBPR. This first reactor, called “Floc” had a working volume of 4 L, a 6 hr cycle time (10 min decant, 6 min feed, 120 min anaerobic, 180 min aerobic, 4 min waste and 40 min settle), 24 hr hydraulic retention time (HRT) and a 10-day solids retention time (SRT). The second reactor, called “Gran” had a working volume of 8 L, a 6 hr cycle time (as above except 218 min aerobic and 2 min settle), a 24 hr HRT and 18-day SRT. Both reactors were fed with synthetic wastewater containing volatile fatty acids (VFA) and orthophosphate ( $\text{PO}_4^{3-}$ ) for the enrichment of PAO. All other operating conditions are as previously described (Barr, Cook, *et al.*, 2010).

### ***Fluorescence in situ hybridization (FISH) quantification***

FISH was performed as previously described (Amann, 1995). Samples were hybridized with PAOMix (Crocetti *et al.*, 2000) and EUBMix (Amann *et al.*, 1990; Daims *et al.*, 1999) probes for *Accumulibacter* and total bacterial abundance respectively. Bacterial populations binding fluorescent DNA probes were visualized as previously described (Barr, Slater, *et al.*, 2010). Quantification of *Accumulibacter* abundance was carried out using DAIME version 1.2 (Daims *et al.*, 2006).

### ***Metagenome sequencing***

Microbial DNA was extracted from both reactors at one time point by centrifuging sludge to remove the supernatant and using ~500 mg of biomass for the Fast DNA SPIN Kit for soil (MP-Bio, USA). Approximately 2  $\mu\text{g}$  of metagenomic DNA was sequenced at the Beijing Genome Institute (Beijing, China) using the Illumina HiSeq2000 with paired end 100 $\times$ 100 library with a 300 bp insert size producing ~12 Gbp of raw data for each of the Floc and Gran communities (see Supplemental Information for full methods details).

### ***Data availability***

Raw sequencing data is available in the sequence read archive under the NCBI project ID PRJNA231881. Annotated and assembled scaffolds from the metagenome can be found in IMG/M under accession numbers 3300003282 and 3300003281 for Floc and Gran respectively.

Draft *Accumulibacter* genome sequences have been deposited in the NCBI whole genome shotgun database under the accession numbers: JDVG000000000 (BA-91), JEMX000000000 (BA-92), JEMY000000000 (BA-93), and JEMZ000000000 (BA-94) (Skenneron *et al.*, 2014).

### ***Protein extraction and purification***

Samples were collected 15 min from the end of the aerobic phase and biomass pelleted by centrifugation at 1200 g for 5 min. To ensure comparable extractions, all samples were concentrated to 10 g volatile suspended solids (VSS)/L prior to extraction. Samples were manually homogenized using a 15 mL glass homogenizer in an icebox. Both total cellular and extracellular proteins were extracted, using the B-PER extraction method and cation exchange resin (CER), respectively. Protein was then precipitated in 13% w/v trichloroacetic acid (TCA) overnight at 4 °C, followed by centrifugation at 15,000 g for 15 min at 4 °C. The protein pellet was then washed twice in 80% (v/v) ice-cold acetone, air-dried, and stored at -80 °C until LC-MS analysis (see Supplemental Information for full methods details).

### ***Protein database, mapping, and analysis***

For mapping the metaproteomic peptides, we created a custom database consisting of all the proteins from the Floc and Gran metagenome assemblies. These proteins were annotated to the orthologous groups (COGs) in the STRING 8.0 database (Jensen *et al.*, 2009) using the Signature web server that identifies and separates fusion proteins and assigns proteins to a COG if at least three of the five top hits belong to that COG (Dutilh *et al.*, 2008). Human keratin 9 and keratin 77 (Genbank identifiers NP\_000217.2 and NP\_778253.2, respectively) were included as controls to account for possible contamination in the proteomic sample, and metaproteomic peptides were mapped to the database with Mascot version 2.2 (Matrix Science Inc., USA) (Sinzinger *et al.*, 2013). Mapping parameters were: enzyme type, trypsin; maximum missed cleavages, 1; fixed modification, Carbamidomethylation (C); variable modification, Oxidation (M), Deamidation (N,Q); peptide tolerance, 10 ppm; MS/MS fragment ion tolerance, 0.8 Da. Criteria for protein identification for single peptide hits included: Mascot peptide score  $\geq 50$ ; modified delta score  $\geq 10$ ; E-value  $< 0.02$ ; maximum variable modification = 1. Criteria for protein identification with  $> 1$  peptide hits included: Mascot peptide score  $\geq 30$ ; modified delta score  $\geq 10$ ; E-value  $< 0.05$ ; maximum variable modification = 5. We performed a reverse database search to calibrate the results to a false discovery rate of 0.05 and assign significance to the protein hosts, as described previously (Sinzinger *et al.*, 2013). Protein quantitation was based

on the number of spectra assigned to all predicted peptides from a given protein, and normalized for the total number of spectra in a sample (Floc or Gran). The abundance of a protein in a sample was calculated from the normalized counts as the log ratio of the average in the four Floc replicates divided by the average in the three Gran replicates. We included a pseudocount of 0.0001 to avoid division by zero (the minimum measured value was 0.00014). Differentially expressed proteins were identified at an alpha cutoff value of 0.05, corresponding to a Z score of 1.77 and 1.9 for the Floc and Gran proteomics, respectively. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno *et al.*, 2014) via the PRIDE partner repository with the dataset identifier PXD002404.

## Results

### *Reactor characterization*

Two laboratory-scale sequencing batch reactors were operated for enhanced biological phosphorus removal (EBPR) for a 161-day period. Both reactors were seeded with floccular EBPR sludge from a domestic wastewater treatment plant (WWTP) and initially enriched with *Candidatus* Accumulibacter phosphatis (Accumulibacter). EBPR activity achieved comparably high levels in both reactors as evident from the high levels of phosphate release in the anaerobic period and low levels of phosphate detected in the effluent following the aerobic period (**Figure 1A and 1C**). The first SBR (called Floc) was maintained as a conventional floccular biofilm (defined by an average particle size of less than 200  $\mu\text{m}$ ) throughout the operation of the reactor (**Figure 1B**). This was achieved by using a typical settling period of 40 min (de Kreuk *et al.*, 2007). The second reactor (called Gran) was manipulated in order to transition from the initial floccular biofilm through to a completely granular biofilm (**Figure 1D**). During this period, the reactor went through a transition stage (Trans), from day 60 to day 105, where larger granular aggregates ( $>200 \mu\text{m}$ ) formed simultaneously with the presence of smaller flocs. The slower settling flocs were selectively removed from the granular reactor by decreasing the settling time to 2 min, generating a completely granular biofilm (Gran) from day 105 onwards (Barr *et al.*, 2010). Biofilms from both reactors were examined at day 142 by microscopy, particle size, mixed liquor suspended solids (MLSS), and fluorescence *in-situ* hybridization (FISH) to confirm completely the floccular and granular biofilm states that were sufficiently enriched with Accumulibacter (**Figure 2A–F**). Both MLSS and particle size were higher in the granular reactor compared to the floccular, indicative of denser and larger biofilm aggregates (SI Figure 1). The bacterial populations of the floccular and granular biofilms, as determined by quantitative FISH, were enriched with 75% and 80% Accumulibacter respectively.

### *Metagenomic analysis*

Biomass was collected from floccular and granular reactors on day 142 of reactor operation for metagenomic sequencing of the bacterial communities (**Figure 1**). Metagenomic sequencing and assembly resulted in 63 Mbp and 100 Mbp of assembled contigs with 99,223 and 135,943 open reading frames for the Floc and Gran samples, respectively. Community analysis showed that

Accumulibacter were the dominant organisms in both samples with both reactors containing similar abundances of flanking community members (**Figure 3A–B**). Contig binning resulted in a single high coverage draft genome of a clade IA Accumulibacter from the Floc reactor and three draft genomes of clade IIC, IIF & and IC Accumulibacter from the Gran reactor (SI Figure 2). There were many large contigs that were not part of Accumulibacter bins suggesting that other members of the community were also present and well assembled in the metagenomes (SI Figure 2). Although we identified numerous different Accumulibacter clades between the reactors, major EBPR pathways involving central carbon and phosphorus metabolisms were identical. For a full comparison between floccular and granular metagenomes please see Skennerton et. al., 2014.

### ***Metaproteomic analysis***

Proteins were extracted from biomass collected at the end of the aerobic phase from both the floccular and granular reactors once comparable levels of EBPR activity were obtained (**Figure 1**). Sequential protein samples were collected from the floccular reactor (Floc,  $n = 4$ ) and from the granular reactor, during its transition period from a floccular to granular sludge (Trans,  $n = 3$ ) and again once a completely granular sludge was obtained (Gran,  $n = 3$ ). For each time point, both cellular and extracellular proteins were extracted from the whole microbial community and analyzed by LC-MS. Peptide tandem mass spectra from the 20 metaproteomes were matched to predicted peptides from the Floc and Gran metagenomes. This resulted in a total of 2,022 unique proteins identified (SI Table 1). The majority of identified proteins originated from Accumulibacter (83% and 80% in Floc and Gran, respectively), with proteins from Gammaproteobacteria and Actinobacteria comprising most of the remaining matches from the flanking community (**Figure 3C–D**). There were numerous discrepancies when comparing reactor community compositions identified by metagenomes versus proteomes. Metagenomes identified both cellular and extracellular DNA present in the activated sludge biofilm (Cheng *et al.*, 2011), while proteomes provided a closer indication of the functional community, closely matching quantitative FISH measurements.

We compared the total Floc and Gran proteomes to identify those proteins that were statistically enriched in either biofilm state (**Figure 4A**, SI Table 2 and 3). Proteins significantly enriched in both the Floc and Gran samples were compared based on their COG functional categories (**Figure 4B**). Floc samples were enriched with proteins associated with translation and energy production, indicating increased cellular growth. Conversely, Gran samples were enriched with proteins associated with cell wall biogenesis, cell motility, intracellular trafficking and secretion, post-translational modification, protein turnover, chaperone functions, and the transport and metabolism of carbohydrates, amino acids, coenzymes, lipids and inorganic ions, suggesting increased extracellular components and diverse transport and metabolisms in granular sludge. The Gran proteome was surprisingly stable over time, while the Trans and Floc proteomes showed evidence of changes (SI Figure 3&4). Finally, in an attempt to identify molecular triggers for granulation we investigated proteins that were consistently increasing in abundance over time within the granular reactor during its transition from a floccular to a granular sludge (Trans) (**Figure 1**). Proteins were identified using a Spearman's rank correlation coefficient analysis of normalized spectra counts and the proteome time points collected during the transition period (SI Table 4).

## Discussion

### *Proteins involved in Accumulibacter core metabolism*

Since the metabolic reconstruction of the first Accumulibacter genome (García Martín *et al.*, 2006), there has been conjecture regarding its metabolic potential. Recent studies have utilized the genomic information for proteomic and transcriptomic approaches to determine active metabolic pathways important for Accumulibacter growth (Wilmes, Wexler, *et al.*, 2008; Wilmes, Andersson, *et al.*, 2008; Burow *et al.*, 2008; He & McMahon, 2011a). The phylogeny of Accumulibacter has been resolved to the strain level through analysis of the polyphosphate kinase gene (*ppk1*), grouping Accumulibacter into two major divisions (type I and II) each with multiple sub-divisions (type IA-IE, IIA-IIF) (He *et al.*, 2007; Skennerton *et al.*, 2014; Flowers *et al.*, 2013). We found the Floc reactor to be dominated by Accumulibacter type IA, whereas the Gran reactor contained a mix of Accumulibacter types IIC, IC, and IIF (Skennerton *et al.*, 2014). In this section we discuss Accumulibacter-identified proteins within our data set, and address questions regarding Accumulibacter metabolism and physiology.

Proteins for acetate uptake and subsequent activation into acetyl-CoA were abundant, and are consequently suggested as the major carbon uptake mechanism utilized by Accumulibacter here. Glycogen degradation was indicated to be active only through the Embden-Meyerhof-Parnas (EMP) pathway (García Martín *et al.*, 2006; Oehmen *et al.*, 2007; Wilmes, Wexler, *et al.*, 2008). We did not detect any peptide hits to the novel cytochrome *b/b6* that is proposed to provide the required anaerobic reoxidation of reduced quinones, indicating this was not a mechanism to release reducing power generated by anaerobic operation of the TCA cycle (García Martín *et al.*, 2006; Flowers *et al.*, 2013). All components of the tricarboxylic acid (TCA) cycle were detected, along with key enzymes involved in both the glyoxylate shunt and split TCA cycle (Hesselmann *et al.*, 2000; Oehmen *et al.*, 2007). Polyhydroxybutyrate (PHB) was the PHA storage polymer utilized, with supplemental polyhydroxyvalerate (PHV) storage also active.

We detected the low-affinity phosphate transporter (Pit) only in Floc biofilm, and the high-affinity phosphate transporter (Pst) in both Floc and Gran biofilms, although not significantly

different. This supports the notion that *Accumulibacter* varies its phosphate transport capabilities under changing EBPR conditions, and suggests that floccular and granular biofilms may have further phosphate transport differences (Wilmes, Andersson, *et al.*, 2008; Burow *et al.*, 2008; He & McMahon, 2011b). There were also differences between *Accumulibacter* proteins involved in denitrification pathways detected in Floc and Gran biofilms. We detected the partial denitrification pathway (periplasmic nitrate reductase, and nitrite reductase) in the Floc biofilm, and the near-complete denitrification pathway (periplasmic nitrate reductase, nitrite reductase, and nitrous oxide reductase) in the Gran biofilm, with only the nitric oxide reductase protein being absent. Interestingly, the complete nitrous oxide reductase (*nosDFLZ*) gene was found to be present in *Accumulibacter* clade IA identified in the floccular reactor, but not in any of the *Accumulibacter* clades detected in the granular reactor. Yet the detection of nitrous oxide reductase precursor protein in the Gran biofilm only suggests that the complete *nos* gene was present, and further analysis revealed that *Accumulibacter* clade IC from the granular reactor contained an incomplete nitrous oxide reductase gene (*nosDZ*). Thus the presence of a near-complete denitrification pathway suggests that the denitrification processes through to completion only in granular sludge (Skenneron *et al.*, 2014). This is important as nitrous oxide is a potent greenhouse gas and suggests that aerobic granular sludge nutrient removal systems may be suitable for its removal (Beaulieu *et al.*, 2011; Quan *et al.*, 2012).

#### ***Proteins enriched in floccular biofilm community***

Proteins enriched in Floc biofilm community as a whole were typical of EBPR systems, but revealed enrichment of pathogenic secretion systems suggesting a highly competitive microbial community (**Figure 5**). We detected the siderophore enterobactin – one of the strongest known siderophores found in microbial systems – in the Floc biofilm that likely plays a role in the parasitic scavenging of extracellular iron. Also enriched were critical iron-containing enzymes such as Ferredoxins involved in electron transfer, respiration, and cellular metabolism, suggesting an increased uptake and utilization of ferric iron (Loiseau *et al.*, 2007). The most highly enriched protein in the Floc biofilm was the Por secretion system (PorSS). This recently described pathogenic protein secretion system is thought to be involved with gliding motility, the secretion of major extracellular proteinases, and chitinase secretion (McBride & Zhu, 2013;

Shoji *et al.*, 2011; Sato *et al.*, 2010). We also detected putative chitinase, which in conjunction with the PorSS, suggests that the Floc community is able to sequester extracellular polymers such as peptidoglycan and chitin that may be present in activated sludge (Eckert *et al.*, 2013). Colicin outer membrane receptors were enriched, and are primarily produced by gram-negative bacteria in competitive mixed microbial communities. These colicin toxins are possibly released into the floccular environment to mediate cytotoxic effects (Jakes & Cramer, 2012). Microbial defense mechanisms were enriched, including the SapC transport system and glutathione S-transferase (GSTs) that provide resistance to antimicrobial peptides and detoxify xenobiotics respectively, providing further support for the highly competitive Floc community (Arca *et al.*, 1988; Shelton *et al.*, 2011).

### ***Proteins involved in the transition from a floccular to granular biofilm***

To elucidate potential trigger proteins involved in the transition from a floccular to a granular biofilm, we analyzed the metaproteomes from the Gran reactor during its transition period (Trans, **Figure 1**). Proteins that consistently increased over time during the Trans period revealed the importance of glycolysis, TCA cycle operation, increased EPS synthesis, and fatty acid anabolism for granule formation (**Figure 6**, SI Table 4).

The bifunctional PGK/TIM was the most detected protein within the Trans samples, and forms a fusion protein of phosphoglycerate kinase (PGK) and triosephosphate isomerase (TIM), two ubiquitous enzymes central to glycolysis (Beaucamp *et al.*, 1997). The bifunctional PGK/TIM fusion protein operates at greater efficiency than the individual enzyme counterparts (Katebi, 2013), allowing for rapid glycolysis for cellular growth and EPS synthesis. Capsular glucan formed the major polysaccharide in the outermost cellular layer (Sambou *et al.*, 2008) and may be important for the establishment of the granular EPS matrix.

Following glycolysis, the oxidative decarboxylation of pyruvate to acetyl-CoA and CO<sub>2</sub> is a critical step in carbon metabolism, typically performed by pyruvate dehydrogenase. Yet here we detect pyruvate synthase (PorABC) that can also catalyze this critical step, although using ferredoxins rather than NAD<sup>+</sup> to transfer electrons (Vita *et al.*, 2008). PorABC can also function in the bio-energetically unfavorable reverse direction under anaerobic conditions (Lin *et al.*,

2003), that may enable the generation of pyruvate and oxidized ferredoxin to subsequently feed into the TCA cycle, generating energy and substrates for cellular growth (Furdui & Ragsdale, 2000). We further detected enzymes indicative of the operation of the full TCA cycle and glyoxylate shunt. The NADP-dependent malic enzyme is an important anaplerotic enzyme that redirects TCA cycle intermediates to pyruvate and provides the required NADPH for fatty acid biosynthesis (Rodriguez *et al.*, 2012).

Numerous enzymes involved in fatty acid biosynthesis were detected in Trans samples, including the synthesis of very-long chain fatty acids (C<sub>60</sub>-C<sub>90</sub>), typically called mycolic acids (Sacco *et al.*, 2007). These mycolic acids require chaperones for their correct synthesis, forming esters anchored to the cell envelope that are crucial for biofilm architecture (Ojha *et al.*, 2005). We identified two chaperones in transition samples that may assist in mycolic acid folding: heat shock protein 70 (HSP70) and heat shock protein G (HSP-G) (Maier *et al.*, 2005). The biosynthesis of mycolic acids, and their correct folding, may explain the previously observed ‘white-ish’ and ‘waxy’ appearance of *Accumulibacter*-dominated granular biofilms (Barr *et al.*, 2010). We also found evidence for the increased production of polyhydroxyalkanoate (PHA) polymers and synthesis of acyl-homoserine lactones (AHLs), which are critical molecules for quorum sensing and biofilm formation (Schaefer *et al.*, 1996; Hoang *et al.*, 2002; Nomura *et al.*, 2005).

### ***Proteins enriched in the granular biofilm community***

Proteins that were statistically enriched in Gran biofilm community as a whole revealed diverse metabolism, increased EPS synthesis, a high stress environment, and bacteriophage predation pressure among others (**Figure 7**). Numerous proteins involved in the production and attachment to EPS were enriched, highlighting the importance of this extracellular matrix in granular sludges. We detected the synthesis of trehalose, an alpha-linked disaccharide that has high water retention and carbohydrate storage capabilities. Trehalose is a gel-forming exo-polysaccharide that helps to maintain protein integrity under conditions of stress (Jain & Roy, 2009). The secretion of gel-forming exo-polysaccharides has recently been implicated as important in the formation of aerobic granular sludge (Seviour *et al.*, 2009). Trehalose may be particularly

important for granular biofilms as it functions as both a gel-like storage polymer and protects against high-stress cellular conditions.

Catalases, proteases, stress response proteins, and associated regulatory proteins were enriched in Gran, indicative of oxidative stress, degradation of damaged proteins, and controlled cell death (Rice & Bayles, 2008). Catalase HPII was the most highly enriched protein in Gran biofilms and is synthesized in response to metabolic growth on TCA cycle intermediates, rather than in response to peroxide (Loewen *et al.*, 1985). The enrichment of stress and starvation proteins within Gran proteomes suggests that the microenvironment within the granular biofilm is nutrient and substrate limited, leading to controlled cell death in the inner most regions of granules (Barr, Cook, *et al.*, 2010). This is likely a result of mass transfer limitations and substrate availability patterns that occur within the larger biofilm (Liu *et al.*, 2005; Toh *et al.*, 2003).

Gran biofilms were enriched with a large diversity of specific and non-specific outer membrane proteins (Omp) and transporters, including ABC transporters, C4-dicarboxylate transporters, and a NAD(P) transhydrogenase. The increased cell death observed in the Gran biofilm would lead to turnover of intracellular components and nutrients, which are in turn taken up by these outer membrane transporters and utilized for cell growth and metabolism. Omp allow for the scavenging of diverse extracellular nutrients and operation of supplemental metabolisms, but they also serve as cellular receptors for phage infection (Lindberg, 1973). Bacteriophage tail, tail sheath proteins, and protein activators of lytic and lysogenic phage lifecycles were enriched in Gran, suggesting phage predation was prevalent within granular biofilms (Hansen *et al.*, 2003; Craig & Nash, 1984). Thus we hypothesize that this increase in Omp seen in granular sludge acts as a double-edged sword: enabling diverse supplemental metabolisms while simultaneously increasing phage predation. The compact structure and higher cellular density within granular biofilms would further increase phage predation pressure compared to floccular biofilms (Barr, Slater, *et al.*, 2010).

### ***A conceptual model of the granular biofilm***

Through metagenomic sequencing and metaproteomic analysis of EBPR laboratory-scale reactors enriched with *Accumulibacter*, we identified a large number of proteins enriched in floccular and granular biofilms, and hinted at trigger proteins that may be involved in the switch from a floccular to a granular biofilm. Future analysis of our proteomic dataset using the recently described two-step database search methods and quantitative statistical methods may reveal further protein identifications and insights into the granule biofilm (Cox & Mann 2012; Jagtap *et al.*, 2013). Based on all of our comparative proteomic analysis, we propose a conceptual model for the formation, structure, and metabolism of the granular biofilm.

It was apparent that there was a substantial increase in the diversity of substrates being utilized by the granular biofilm. In this instance, the synthetic feed was identical for both the Floc and Gran reactors. Consequently, differences in metabolically available substrates had to emerge from differing processes intrinsic to the respective biofilms. A key difference seen within granular biofilms was the extent of cellular stress and lysis, evident through increased detection of stress response proteins, cell death, and phage predation. The inner regions of granules are thought to experience limited nutrient availability due to exterior biofilm substrate use and mass transfer limitations, leading to regions of higher cell death, and increased turnover of cellular components. Indeed, the inner parts of granules are known to be void of active cells (Barr, Cook, *et al.*, 2010). Consequently, the granule biofilm utilizes diverse outer membrane transporters to scavenge for amino acids and other cellular substrates, released into the surrounding extracellular environment through cell lysis, to supplement metabolism. This increase in outer membrane transporters perpetuates phage predation of the inner granule biofilm (Barr, Slater, *et al.*, 2010). Due to the larger size and stratified structure of aerobic granules, a life cycle of granule size is proposed. Here, due to metabolic and physical limitations, the granule biofilm grows to a critical size, above this size the inner core of the granule is significantly weakened and these fracture into smaller granule particles (Verawaty *et al.*, 2013). The findings from this study provide a greater understanding of the differences between these biofilm states, and this improved insight may contribute to better application and management of aerobic granular wastewater treatment processes.

## **Acknowledgements**

This work was funded by the Environmental Biotechnology Cooperative Research Centre (EBCRC), which was established and funded by the Australian Government, together with industry and university partners. JJB acknowledges EBCRC for funding of a Ph.D. scholarship. PLB acknowledges EBCRC, Waste Technologies of Australia, The University of Queensland, and the Queensland Government Smart State Fellowship Program for funding of a senior research fellowship. Access to proteomic infrastructure in the QIMR Berghofer Protein Discovery Centre was made possible by funding from Bioplatforms Australia and the Queensland State Government provided through the Australian Government National Collaborative Infrastructure Scheme (NCRIS) and Education Investment Fund (EIF). The authors thank Jolein Gloerich and Hans Wessels of the Radboud Proteomics Centre for their help in analyzing the proteomics data. The authors declare no conflict of interest.

## Reference

- Amann RI. (1995). Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology. *Mol Ecol* **4**:543–554.
- Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Envir Microbiol* **56**:1919–1925.
- Arca P, Rico M, Brana AF, Villar CJ, Hardisson C, Suarez JE. (1988). Formation of an adduct between fosfomycin and glutathione: a new mechanism of antibiotic resistance in bacteria. *Antimicrob Agents Chemother* **32**:1552–1556.
- Barr JJ, Cook AE, Bond PL. (2010). Granule formation mechanisms within an aerobic wastewater system for phosphorus removal. *Appl Environ Microbiol* **76**:7588–97.
- Barr JJ, Slater FR, Fukushima T, Bond PL. (2010). Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. *FEMS Microbiol Ecol* **74**:631–42.
- Beaucamp N, Hofmann A, Kellerer B, Jaenicke R. (1997). Dissection of the gene of the bifunctional PGK-TIM fusion protein from the hyperthermophilic bacterium *Thermotoga maritima*: design and characterization of the separate triosephosphate isomerase. *Protein Sci* **6**:2159–65.
- Beaulieu JJ, Tank JL, Hamilton SK, Wollheim WM, Hall RO, Mulholland PJ, *et al.* (2011). Nitrous oxide emission from denitrification in stream and river networks. *Proc Natl Acad Sci U S A* **108**:214–9.
- Beun J., Hendriks A, van Loosdrecht MC., Morgenroth E, Wilderer P., Heijnen J. (1999). Aerobic granulation in a sequencing batch reactor. *Water Res* **33**:2283–2290.
- Blackall LL, Crocetti GR, Saunders AM, Bond PL. (2002). A review and update of the microbiology of enhanced biological phosphorus removal in wastewater treatment plants. *Antonie Van Leeuwenhoek* **81**:681–691.
- Bond P, Hugenholtz P, Keller J, Blackall L. (1995). Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl Envir Microbiol* **61**:1910–1916.
- Burow LC, Mabbett AN, Blackall LL. (2008). Anaerobic glyoxylate cycle activity during simultaneous utilization of glycogen and acetate in uncultured *Accumulibacter* enriched in enhanced biological phosphorus removal communities. *ISME J* **2**:1040–51.
- Cheng M, Cook AE, Fukushima T, Bond PL. (2011). Evidence of compositional differences between the extracellular and intracellular DNA of a granular sludge biofilm. *Lett Appl Microbiol* **53**:1–7.
- Cornel P, Krause S. (2006). Membrane bioreactors in industrial wastewater treatment--European experiences, examples and trends. *Water Sci Technol* **53**:37–44.
- Costerton JW, Cheng KJ, Geesey GG, Ladd TI, Nickel JC, Dasgupta M, *et al.* (1987). Bacterial biofilms in nature and disease. *Annu Rev Microbiol* **41**:435–64.
- Cox J, Mann M. (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13**.

Craig NL, Nash HA. (1984). E. coli integration host factor binds to specific sites in DNA. *Cell* **39**:707–716.

Crocetti GR, Hugenholtz P, Bond PL, Schuler A, Keller J, Jenkins D, *et al.* (2000). Identification of Polyphosphate-Accumulating Organisms and Design of 16S rRNA-Directed Probes for Their Detection and Quantitation. *Appl Environ Microbiol* **66**:1175–1182.

Daims H, Brühl A, Amann R, Schleifer KH, Wagner M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**:434–44.

Daims H, Lückner S, Wagner M. (2006). daime, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8**:200–13.

Davies DG, Parsek MR, Pearson JP, Iglewski BH, Costerton JW, Greenberg EP. (1998). The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm. *Science (80- )* **280**:295–298.

Dutilh BE, He Y, Hekkelman ML, Huynen MA. (2008). Signature, a web server for taxonomic characterization of sequence samples using signature genes. *Nucleic Acids Res* **36**:W470–4.

Eckert EM, Baumgartner M, Huber IM, Pernthaler J. (2013). Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. *Environ Microbiol* **15**:2019–30.

Flowers JJ, He S, Malfatti S, del Rio TG, Tringe SG, Hugenholtz P, *et al.* (2013). Comparative genomics of two “Candidatus Accumulibacter” clades performing biological phosphorus removal. *ISME J* **7**:2301–14.

Fridkin SK, Welbel SF, Weinstein RA. (1997). Magnitude and prevention of nosocomial infections in the intensive care unit. *Infect Dis Clin North Am* **11**:479–496.

Fuhs GW, Chen M. (1975). Microbiological basis of phosphate removal in the activated sludge process for the treatment of wastewater. *Microb Ecol* **2**:119–38.

Furdui C, Ragsdale SW. (2000). The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. *J Biol Chem* **275**:28494–9.

García Martín H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, *et al.* (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* **24**:1263–9.

Hansen A-M, Lehnerr H, Wang X, Mobley V, Jin DJ. (2003). Escherichia coli SspA is a transcription activator for bacteriophage P1 late genes. *Mol Microbiol* **48**:1621–1631.

He S, Gall DL, McMahon KD. (2007). “Candidatus Accumulibacter” population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Appl Environ Microbiol* **73**:5865–74.

He S, Gu AZ, McMahon KD. (2006). Fine-scale differences between Accumulibacter-like bacteria in enhanced biological phosphorus removal activated sludge. *Water Sci Technol* **54**:111–7.

He S, McMahon KD. (2011a). “Candidatus Accumulibacter” gene expression in response to dynamic EBPR conditions. *ISME J* **5**:329–40.

He S, McMahon KD. (2011b). Microbiology of “Candidatus Accumulibacter” in activated sludge. *Microb Biotechnol* **4**:603–19.

Hesselmann R, Rummell R, Resnick S, Hany R, Zehnder A. (2000). Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Res* **34**:3487–3494.

Hesselmann RP, Werlen C, Hahn D, van der Meer JR, Zehnder AJ. (1999). Enrichment, phylogenetic analysis and detection of a bacterium that performs enhanced biological phosphate removal in activated sludge. *Syst Appl Microbiol* **22**:454–65.

Hoang TT, Sullivan SA, Cusick JK, Schweizer HP. (2002).  $\beta$ -Ketoacyl acyl carrier protein reductase (FabG) activity of the fatty acid biosynthetic pathway is a determining factor of 3-oxo-homoserine lactone acyl chain lengths. *Microbiology* **148**:3849–3856.

Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, Griffin TJ. (2013). A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* **13**:1352-7.

Jain NK, Roy I. (2009). Effect of trehalose on protein structure. *Protein Sci* **18**:24–36.

Jakes KS, Cramer WA. (2012). Border crossings: colicins and transporters. *Annu Rev Genet* **46**:209–31.

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, *et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**:412–6.

Katebi AR. (2013). Building and simulating protein machines. Iowa State University

Kong Y, Nielsen JL, Nielsen PH. (2005). Identity and ecophysiology of uncultured actinobacterial polyphosphate-accumulating organisms in full-scale enhanced biological phosphorus removal plants. *Appl Environ Microbiol* **71**:4076–85.

De Kreuk MK, Kishida N, van Loosdrecht MCM. (2007). Aerobic granular sludge--state of the art. *Water Sci Technol* **55**:75–81.

Lin WC, Yang Y-L, Whitman WB. (2003). The anabolic pyruvate oxidoreductase from *Methanococcus maripaludis*. *Arch Microbiol* **179**:444–56.

Lindberg AA. (1973). Bacteriophage receptors. *Annu Rev Microbiol* **27**:205–41.

Liu Y, Tay J-H. (2004). State of the art of biogranulation technology for wastewater treatment. *Biotechnol Adv* **22**:533–63.

Liu Y-Q, Liu Y, Tay J-H. (2005). Relationship between size and mass transfer resistance in aerobic granules. *Lett Appl Microbiol* **40**:312–5.

Loewen PC, Switala J, Triggs-Raine BL. (1985). Catalases HPI and HPII in *Escherichia coli* are induced independently. *Arch Biochem Biophys* **243**:144–149.

Loiseau L, Gerez C, Bekker M, Ollagnier-de Choudens S, Py B, Sanakis Y, *et al.* (2007). ErpA, an iron sulfur (Fe S) protein of the A-type essential for respiratory metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **104**:13626–31.

Maier T, Ferbitz L, Deuerling E, Ban N. (2005). A cradle for new proteins: trigger factor at the ribosome. *Curr Opin Struct Biol* **15**:204–12.

Mao Y, Yu K, Xia Y, Chao Y, Zhang T. (2014). Genome Reconstruction and Gene Expression of “Candidatus Accumulibacter phosphatis” Clade IB Performing Biological Phosphorus Removal. *Environ Sci Technol* **48**:10363–71.

McBride MJ, Zhu Y. (2013). Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J Bacteriol* **195**:270–8.

Mino T, van Loosdrecht MCM, Heijnen JJ. (1998). Microbiology and biochemistry of the enhanced biological phosphate removal process. *Water Res* **32**:3193–3207.

Morgenroth E, Sherden T, Van Loosdrecht MCM, Heijnen JJ, Wilderer PA. (1997). Aerobic granular sludge in a sequencing batch reactor. *Water Res* **31**:3191–3194.

Nicolella C, van Loosdrecht MCM, Heijnen JJ. (2000). Wastewater treatment with particulate biofilm reactors. *J Biotechnol* **80**:1–33.

Nielsen PH, Saunders AM, Hansen AA, Larsen P, Nielsen JL. (2012). Microbial communities involved in enhanced biological phosphorus removal from wastewater—a model system in environmental biotechnology. *Curr Opin Biotechnol* **23**:452–9.

Nomura CT, Taguchi K, Gan Z, Kuwabara K, Tanaka T, Takase K, *et al.* (2005). Expression of 3-ketoacyl-acyl carrier protein reductase (fabG) genes enhances production of polyhydroxyalkanoate copolymer from glucose in recombinant *Escherichia coli* JM109. *Appl Environ Microbiol* **71**:4297–306.

O’Toole G, Kaplan HB, Kolter R. (2000). Biofilm formation as microbial development. *Annu Rev Microbiol* **54**:49–79.

Oehmen A, Lemos PC, Carvalho G, Yuan Z, Keller J, Blackall LL, *et al.* (2007). Advances in enhanced biological phosphorus removal: from micro to macro scale. *Water Res* **41**:2271–300.

Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF. (2005). GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. *Cell* **123**:861–73.

Quan X, Zhang M, Lawlor PG, Yang Z, Zhan X. (2012). Nitrous oxide emission and nutrient removal in aerobic granular sludge sequencing batch reactors. *Water Res* **46**:4981–90.

Rice KC, Bayles KW. (2008). Molecular control of bacterial death and lysis. *Microbiol Mol Biol Rev* **72**:85–109, table of contents.

Rodriguez E, Navone L, Casati P, Gramajo H. (2012). Impact of malic enzymes on antibiotic and triacylglycerol production in *Streptomyces coelicolor*. *Appl Environ Microbiol* **78**:4571–9.

Sacco E, Covarrubias AS, O’Hare HM, Carroll P, Eynard N, Jones TA, *et al.* (2007). The missing piece of the type II fatty acid synthase system from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **104**:14628–33.

Sambou T, Dinadayala P, Stadthagen G, Barilone N, Bordat Y, Constant P, *et al.* (2008). Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Mol Microbiol* **70**:762–74.

Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, *et al.* (2010). A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A* **107**:276–81.

Schaefer AL, Val DL, Hanzelka BL, Cronan JE, Greenberg EP. (1996). Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein. *Proc Natl Acad Sci* **93**:9505–9509.

Schultz MP. (2007). Effects of coating roughness and biofouling on ship resistance and powering. *Biofouling* **23**:331–41.

Seviour T, Pijuan M, Nicholson T, Keller J, Yuan Z. (2009). Gel-forming exopolysaccharides explain basic differences between structures of aerobic sludge granules and floccular sludges. *Water Res* **43**:4469–4478.

Shelton CL, Raffel FK, Beatty WL, Johnson SM, Mason KM. (2011). Sap transporter mediated import and subsequent degradation of antimicrobial peptides in *Haemophilus*. Seifert, HS (ed). *PLoS Pathog* **7**:e1002360.

Shoji M, Sato K, Yukitake H, Kondo Y, Narita Y, Kadowaki T, *et al.* (2011). Por secretion system-dependent secretion and glycosylation of *Porphyromonas gingivalis* hemin-binding protein 35. Adler, B (ed). *PLoS One* **6**:e21372.

Sinzinger MD, Ruttekkolk IR, Gloerich J, Wessels H, Chung Y-D, Adjobo-Hermans MJW, *et al.* (2013). Peptide microarrays to probe for competition for binding sites in a protein interaction network. *J Proteomics* **89**:71–80.

Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. (2014). Expanding our view of genomic diversity in *Candidatus Accumulibacter* clades. *Environ Microbiol Rep*.

Stephenson T, Judd S, Jefferson B, Brindle K. (2000). Membrane bioreactors for wastewater treatment.

Toh SK, Tay JH, Moy BYP, Ivanov V, Tay STL. (2003). Size-effect on the physical characteristics of the aerobic granule in a SBR. *Appl Microbiol Biotechnol* **60**:687–95.

Verawaty M, Tait S, Pijuan M, Yuan Z, Bond PL. (2013). Breakage and growth towards a stable aerobic granule size during the treatment of wastewater. *Water Res* **47**:5338–49.

Vita N, Hatchikian EC, Nouailler M, Dolla A, Pieulle L. (2008). Disulfide bond-dependent mechanism of protection against oxidative stress in pyruvate-ferredoxin oxidoreductase of anaerobic *Desulfovibrio* bacteria. *Biochemistry* **47**:957–64.

Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, *et al.* (2014). ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination. *Nature Biotechnol.* **30**(3):223-226.

Watnick P, Kolter R. (2000). Biofilm, City of Microbes. *J Bacteriol* **182**:2675–2679.

Weissbrodt DG, Schneiter GS, Fürbringer J-M, Holliger C. (2013). Identification of trigger factors selecting for polyphosphate- and glycogen-accumulating organisms in aerobic granular sludge sequencing batch reactors. *Water Res* **47**:7006–18.

Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B, *et al.* (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J* **2**:853–64.

Wilmes P, Wexler M, Bond PL. (2008). Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS One* **3**:e1778.

Wingender J, Neu TR, Flemming H. (1999). *Microbial Extracellular Polymeric Substances*. Springer Berlin Heidelberg: Berlin, Heidelberg

Winkler M-KH, Bassin JP, Kleerebezem R, de Bruin LMM, van den Brand TPH, van Loosdrecht MCM. (2011). Selective sludge removal in a segregated aerobic granular biomass system as a strategy to control PAO-GAO competition at high temperatures. *Water Res* **45**:3291–9.

Accepted Article

## Figure Legends

**Figure 1.** Two laboratory-scale sequence batch reactors (SBR) were operated over a 161-day period as floccular and granular biofilms. Enhanced biological phosphorus removal (EBPR) performance (A and C) and particle size (B and D) for the Floc and Gran reactors were monitored throughout this period. Samples indicate time points when metagenome (M) and metaproteome (P) samples were collected. The 50<sup>th</sup> percentile is the median particle size. Inorganic phosphate (Pi) release was measured at the end of anaerobic phase and Pi effluent was measured at the end of aerobic phase.

**Figure 2.** Microscopic analysis by stereomicroscopy (A and D), light microscopy (B and E), and confocal microscopy using fluorescence *in-situ* hybridization (FISH) (C and F) of Floc and Gran biofilms collected on day 142 of operation. Mixed liquor suspended solids (MLSS) indicate the concentration of suspended solids and biomass. Average particle size is representative of median. FISH images show *Accumulibacter* in green and all other bacteria in blue. All scale bars = 200  $\mu\text{m}$ .

**Figure 3.** Relative abundance of the bacterial community composition from Floc and Gran biofilms as assessed by raw metagenomic reads (A and B) and protein identifications (C and D).

**Figure 4.** Community-level changes in protein abundance. (A) Fold change in normalized protein abundance between Floc and Gran reactors at the community level. Dashed lines indicate the statistical cutoff (alpha value = 0.05), resulting in the selection of the top 5% of significantly enriched proteins for each biofilm. This resulted in 101 proteins significantly enriched in each biofilm state, corresponding to a fold change in normalized protein abundance of 1.22 and -1.15, and a Z score cutoff of 1.77 and 1.9, for Floc and Gran biofilms respectively. (B) Number of proteins assigned to Clusters of Orthologous Groups (COG) that are significantly enriched in Floc and Gran reactors at the community level. COG categories; J: Translation, ribosomal structure and biogenesis; A: RNA processing and modification; K: Transcription; L: Replication, recombination and repair; B: Chromatin structure and dynamics; D: Cell cycle control, cell division, chromosome partitioning; Y: Nuclear structure; T: Signal transduction mechanisms; M:

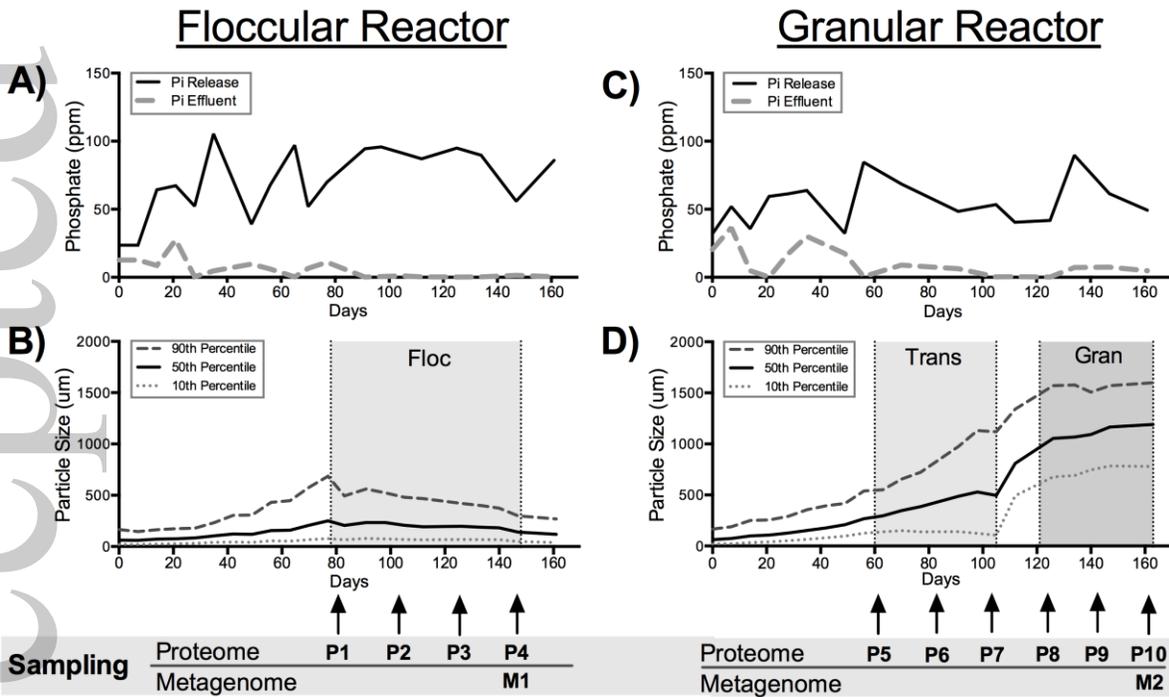
Cell wall/membrane/envelope biogenesis; N: Cell motility; Z: Cytoskeleton; U: Intracellular trafficking, secretion and vesicular transport; O: Posttranslational modification, protein turnover, chaperones; C: Energy production and conversion; G: Carbohydrate transport and metabolism; E: Amino-acid transport and metabolism; F: Nucleotide transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General functional prediction only; and S: Function unknown.

**Figure 5.** Floccular biofilm schematic of proteins that were significantly enriched within the Floc metaproteomes (black arrows). Grey arrows represent core metabolic proteins that were present, but not sufficiently enriched in Floc. Abbreviations: 50S, 50S ribosomal protein; acn, aconitate hydratase; BDH1, D-beta-hydroxybutyrate dehydrogenase; erpA, iron-sulfur cluster insertion protein; FRD, fumarate reductase; G-3-P, glycerol-3-phosphate; Gde, glycogen debranching enzyme; GS, glutamate synthase; GSTs, glutathione S-transferase; nudix, NUDIX domain; OGOR, 2-oxoglutarate ferredoxin oxidoreductase; pap, polyphosphate:AMP phosphotransferase; PhaF, polyhydroxyalkanoate granule-associated protein; SCS, succinyl-CoA ligase; ywqF, UDP-glucose 6-dehydrogenase.

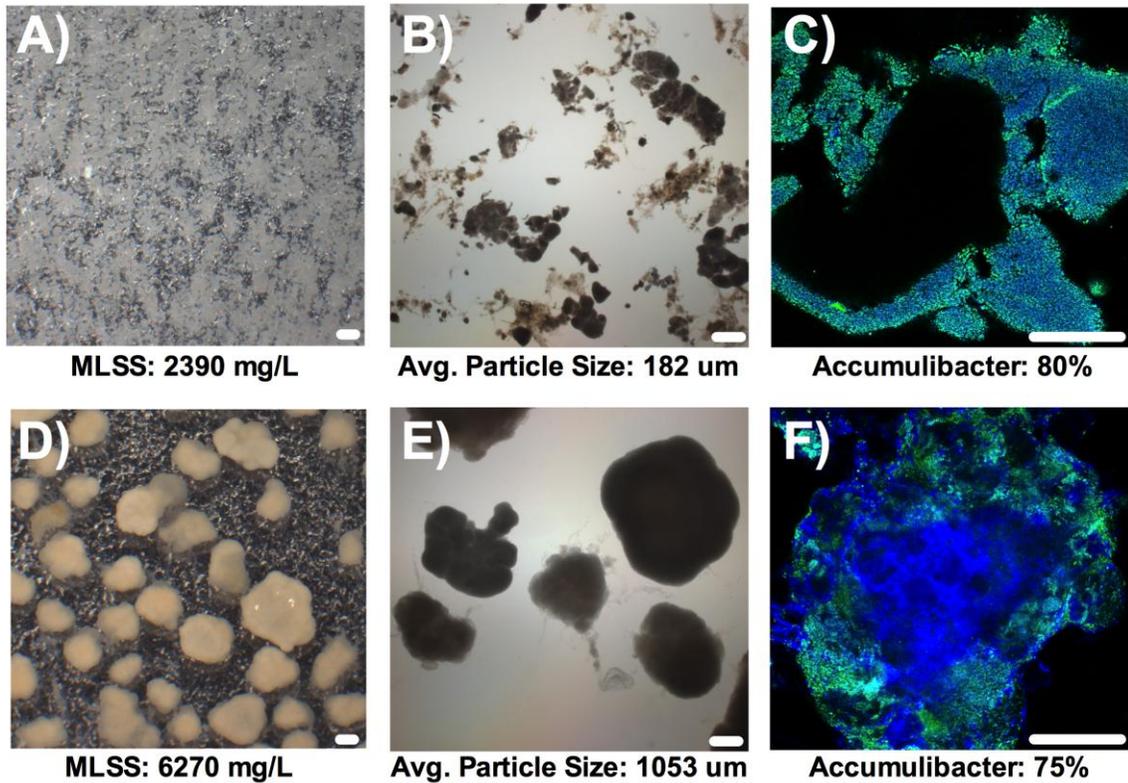
**Figure 6.** Reconstruction of proteins that were continuously increasing in abundance throughout the Trans metaproteomes. Red arrows indicate proteins that were continuously increasing throughout Trans, black arrows indicate proteins that were detected but were not continuously increasing in abundance. Abbreviations: acct, acetyl-CoA acetyltransferase; acn, aconitate hydratase; Bifunctional PGK/TIM, Bifunctional phosphoglycerate kinase/triosephosphate isomerase; CGS, capsular glucan synthase; CS, citrate synthase; FabG, 3-oxoacyl reductase; GIPT, glucose-1-phosphate adenylyltransferase; GlaS, glutamate synthase; GliS, glutamine synthase; GmpS, GMP-synthase; HadB, hydroxyacyl-ACP dehydratase; HSP70, heat shock protein; HSP-G, high temperature protein; ICL, isocitrate lyase; IDH, isocitrate dehydrogenase; LCFAS, long-chain-fatty-acid-CoA ligase; ME, malic enzyme; MMC, methylmalonyl-CoA carboxyltransferase; OGDC, 2-oxoglutarate dehydrogenase; PFP, diphosphate-fructose-6-

phosphate 1-phosphotransferase; PorABC, pyruvate synthase; SamS, S-adenosylmethionine synthase; SCS, succinyl-CoA ligase.

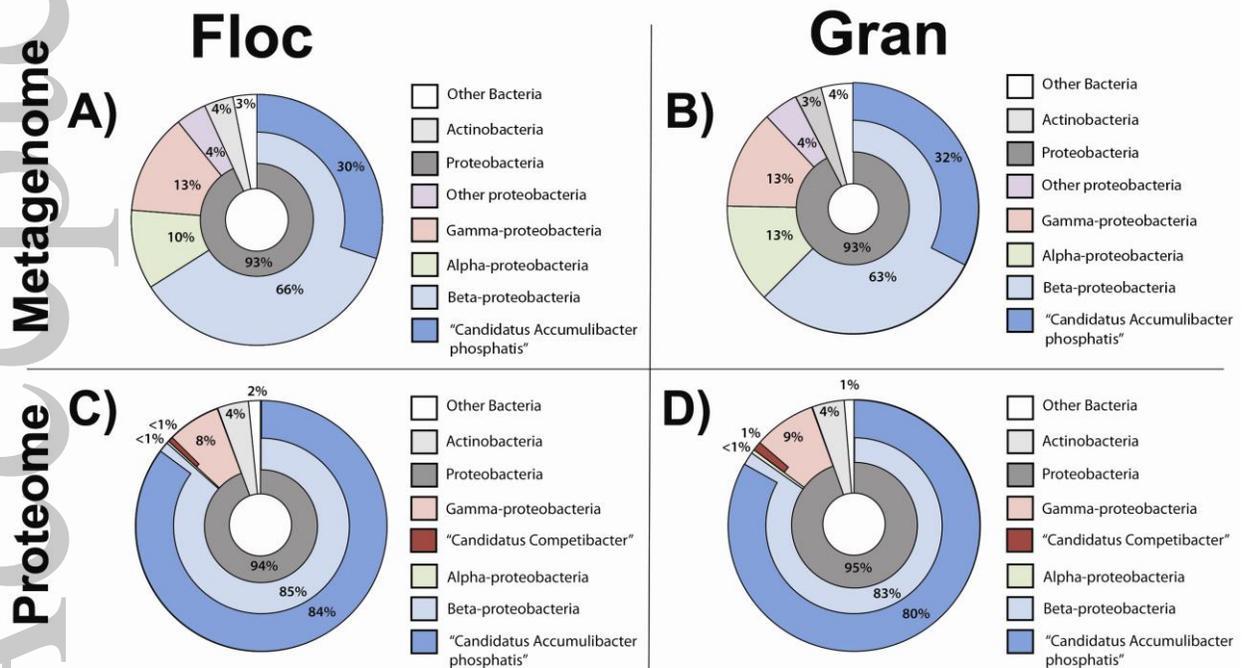
**Figure 7.** Granular biofilm schematic of proteins that were significantly enriched within the Gran metaproteomes (black arrows). Grey arrows represent core metabolic proteins that were present, but not sufficiently enriched in Gran. Abbreviations: aa, alpha-amylase; acct, acetyl-CoA acetyltransferase; acs, acetyl-CoA synthase; FRD, fumarate reductase; G1PA, glucose-1-phosphate adenylytransferase; GabT, 4-aminobutyrate aminotransferase; Moth, malto-oligosyltrehalose trehalohydrolase; NOS, nitrous oxide reductase; PDC, pyruvate dehydrogenase complex; PEP, phosphoenolpyruvate carboxykinase.



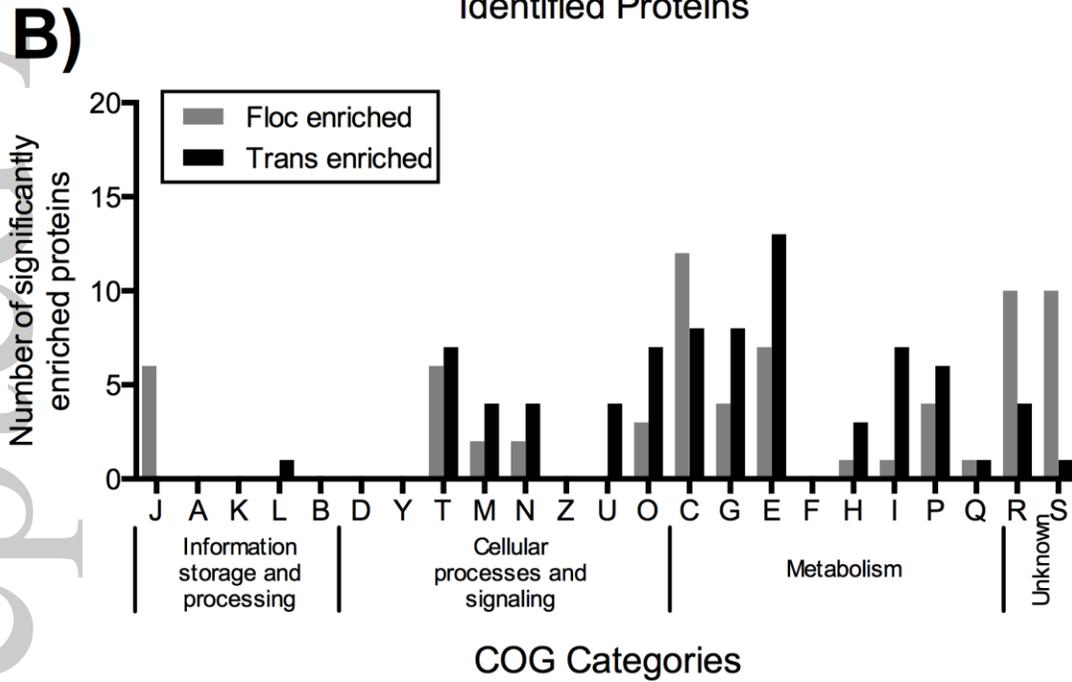
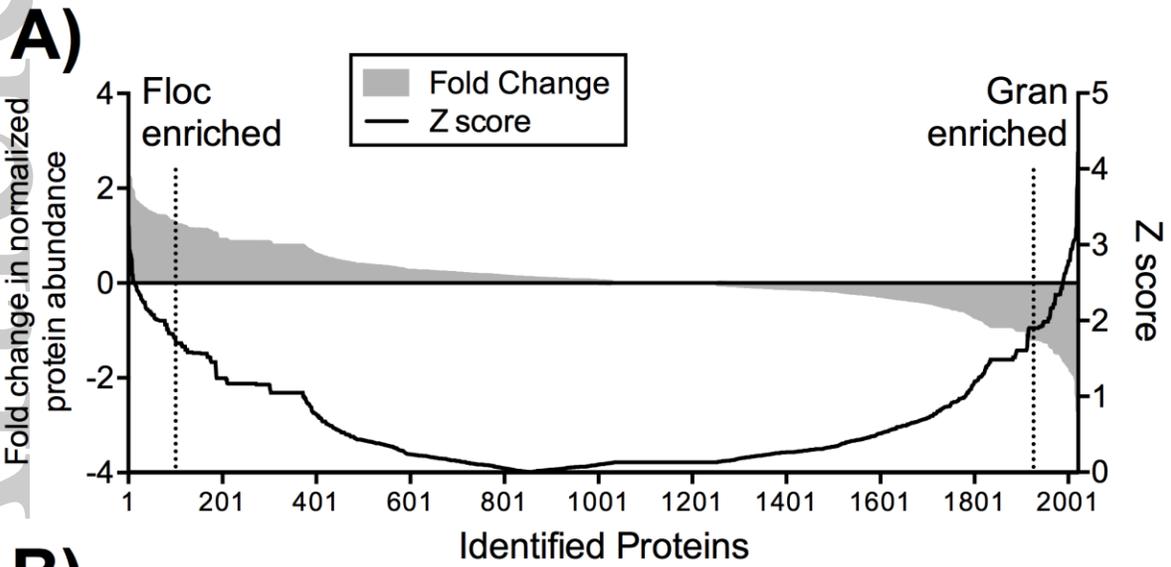
EMI\_13019\_f1



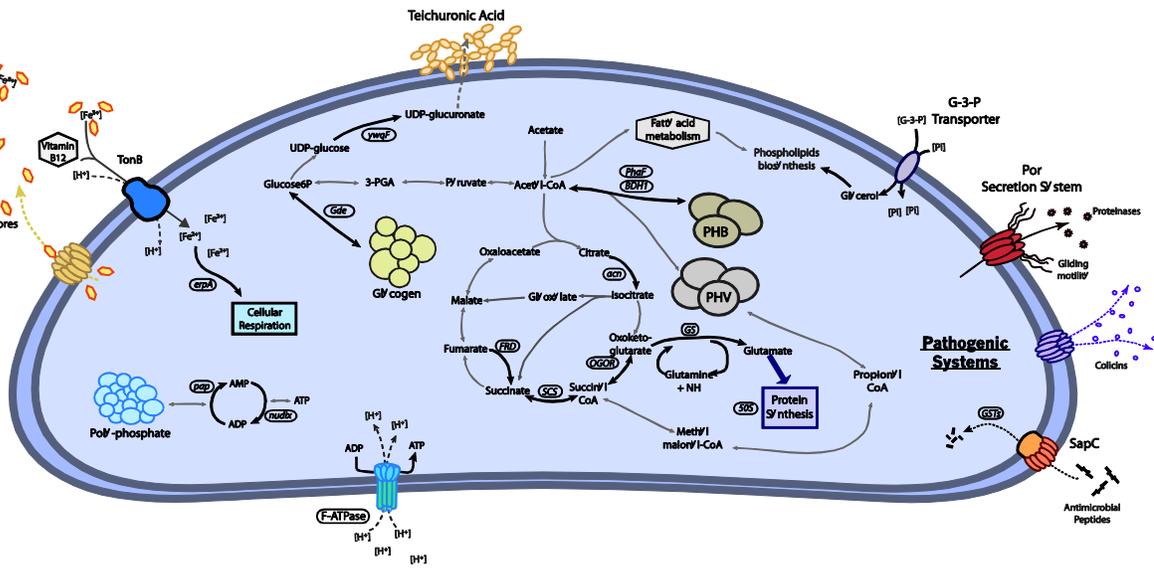
EMI\_13019\_f2



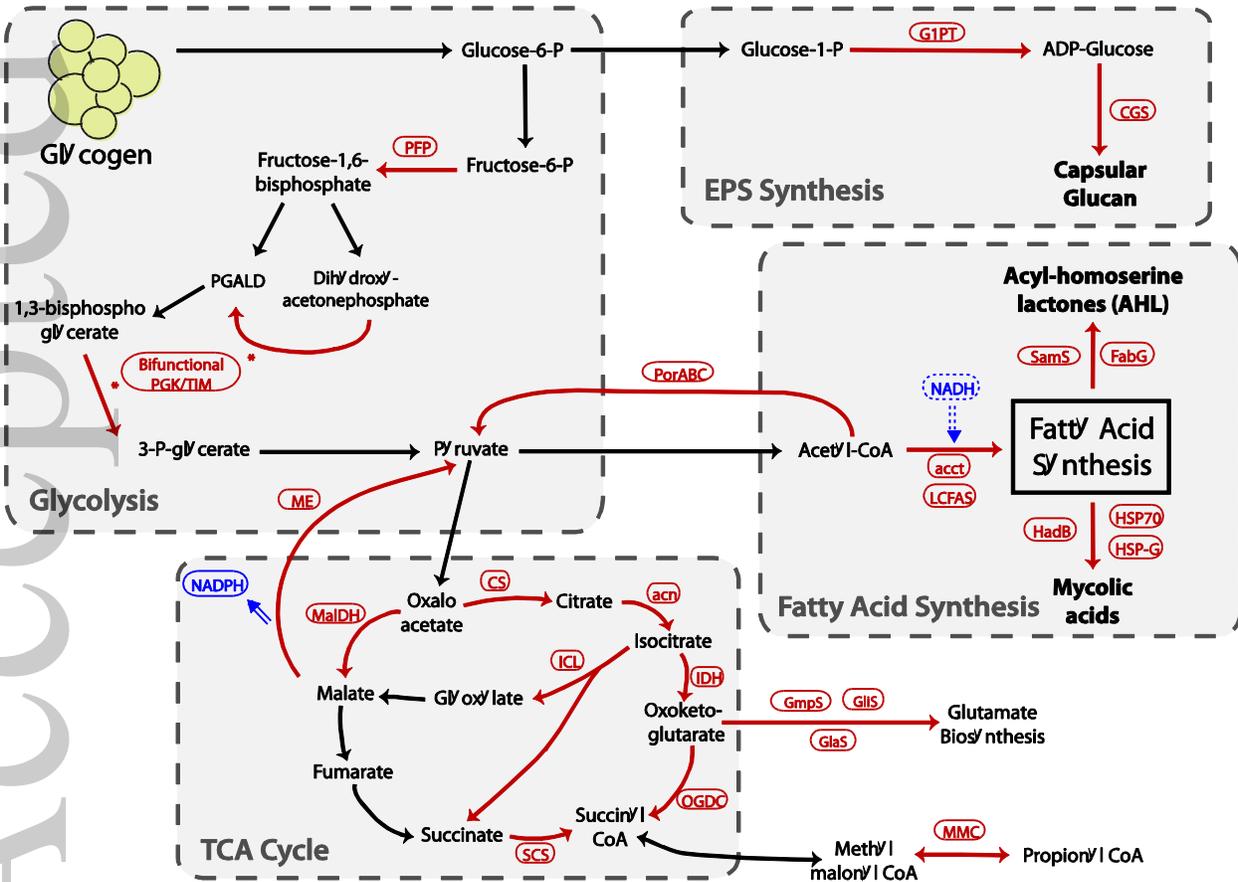
EMI\_13019\_f3



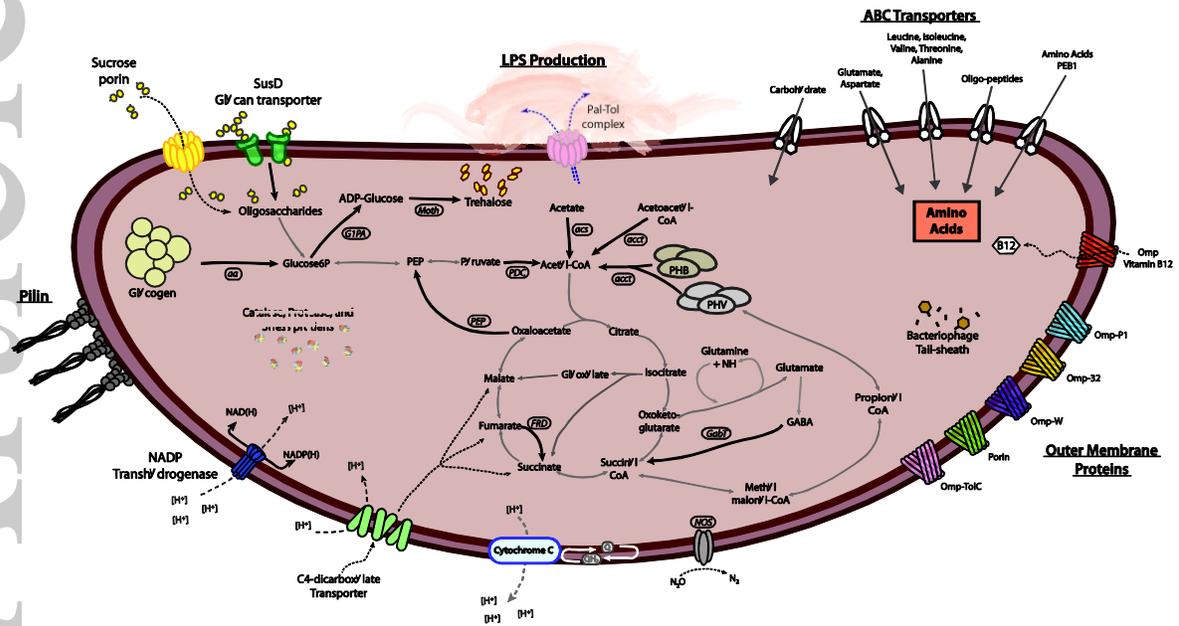
EMI\_13019\_f4



EMI\_13019\_f5



EMI\_13019\_f6



EMI\_13019\_f7