A metabolic network of a phosphate-accumulating organism provides new insights into enhanced biological phosphorous removal

Sergio Bordel

ABSTRACT

Here we present a metabolic network representing the central carbon metabolism as well as the synthesis of polyhydroxyalcanohates and the polyphosphate accumulation mechanisms of the bacterium Candidatus Accumulibacter phosphatis, which was previously identified from metagenomic studies in enhanced biological phosphorous removal sludges. The reconstructed metabolic network, together with flux balance analysis can be used to provide new insights into controversial aspects of the metabolism of phosphate-accumulating organisms and is also a tool that can be used in to help enhanced biological phosphorous removal (EBPR) process design and operation.

Key words | enhanced biological phosphorous removal, flux balance analysis, metabolic networks

INTRODUCTION

Enhanced biological phosphorous removal (EBPR) has been used for more than 30 years in waste water treatment. The core of this process is the accumulation of polyphosphate granules inside some of the cells present in the activated sludge. When this polyphosphate-enriched sludge is separated from the effluent water, phosphorous is also removed (Fuhs & Chen 1975; Wentzel et al. 1986). It has been observed (Meganck & Faup 1988; Toerien et al. 1990) that in order to achieve a successful phosphate removal, the activated sludge needs to be exposed to readily usable carbon sources (volatile fatty acids) in anaerobic conditions. It is very important for the phosphorous elimination that the conditions are strictly anaerobic and nitrates are absent in the medium. When the sludge is exposed to a subsequent aerobic phase, polyphosphates are accumulated in the biomass. During the anaerobic phase the volatile fatty acids are transformed into polyhydroxyalcanoates (PHAs) which are stored and used as carbon and energy sources during the aerobic phase (Lemos et al. 1998). Glycogen reserves are also utilized to supply reductive power (in the form of NADPH) for the PHA synthesis.

The identity of the phosphate-accumulating organisms (PAOs) and the metabolic processes involved remain open questions. An extensive review on the topic was published by Seviour and co-workers in 2003 (Seviour et al. 2003). The astonishing recent advances in gene-sequencing techniques have made possible to use shot-gun sequencing and metagenomics in order to identify PAOs and their genomes. García Martín and co-workers (García Martín et al. 2006) identified via metagenomic studies a putative PAO that accounted for 80 and 60% of the biomass in two different activated sludges from the USA and Australia respectively; this organism was named Candidatus Accumulibacter phosphatis. The genomic data collected by García Martín and co-workers can be used to deduce the presence of certain metabolic reactions and pathways and give a good insight into the metabolic bases of EBPR. In this work we have built a metabolic network based on the pathways identified by García Martín and co-workers. This metabolic network can be used to make quantitative predictions about the EBPR using flux balance analysis (FBA). The metabolic network here presented can be used as a guide for the design and operation of more efficient EBPR process and at the same time can explain and give new insights about some unclear features of EBPR.

METHODS

Reconstruction of the metabolic network

The genome of A. phosphatis allowed inferring the presence of certain metabolic enzymes and therefore the existence of
certain metabolic reactions (García Martín et al. 2006). García Martín and co-workers identified the existence of both high- and low-affinity phosphate transporters that allow for the uptake of phosphates against the concentration gradient (with the consumption of ATP). They also identified all the genes corresponding to the Embden-Meyerhof glycolytic pathway while several key genes in the Entner–Doudoroff pathway were missing. Therefore the Embden-Meyerhof pathway was added to our network. All the steps in the TCA cycle were identified as well as a gene coding for the malic enzyme (MalDH). A phosphatidic acid phosphatase (pap) involved in polyphosphate degradation by adding a phosphate to an AMP molecule was discovered as well as adenosine kinase (adk), which transforms two ADP molecules into one AMP and an ADP. These two enzymes are responsible for the supply of ATP in anaerobic conditions. Many enzymes involved in PHA biosynthesis were also identified, such as three β-ketothiolases (phaA, phaB and phaC) and an acetoacetyl-CoA reductase responsible for the synthesis of hydroxybutyryl-CoA and hydroxy-2-methylvaleryl-CoA from acetyl-CoA and propionyl-CoA respectively. Also a crotonase (croR), a crotonyl-CoA carboxylase reductase (ccR) and an ethyl-malonyl epimerase (ibd2) were identified. These last three enzymes are part of a pathway that transforms hydroxybutyryl-CoA into propionyl-CoA and glyoxylate (Šmejkalová et al. 2010), therefore we inferred the presence of the rest of the pathway from the existence of these three enzymes and added the full path to our network. Two isoenzymes (mcm and meaB) that catalyze the transformation of succinyl-CoA into methymalonyl-CoA were also found in the genome of A. phosphatis, as well as three isoenzymes transforming methymalonyl-CoA into propionyl-CoA (mmdA, mmdB and mmdC). The presence of pathways producing propionyl-CoA explains the fact that even in cases in which acetate is the only external carbon source, significant amounts of polyhydroxylvalerate (PHV) are also formed (Schuler & Jenkins 2003). A putative cytochrome able to transfer electrons from reduced quinone to NADP was also identified by García Martín and co-workers. The putative reaction is driven by a proton gradient across the membrane and was also added to our network.

In the draft network generated by García Martín and co-workers (García Martín et al. 2006) the redox cofactors NAD and NADP are considered as equivalent. In reality it is not possible to transport electrons freely between NADH and NADP. Even in organisms that possess transhydrogenases, the ratio NADPH/NADP is higher, than the ratio NADH/NAD in physiological conditions. This means that the transfer of electrons is only thermodynamically favourable from NADPH to NAD and not from NADH to NADP (Sazanov & Jackson 1994). We included the typical cofactor that each of the included reactions uses in other microorganisms. Once the proper redox cofactors are introduced, the first simulations revealed that glycogen degradation cannot supply NADPH for PHA biosynthesis, unless the pentose phosphate pathway is used (which is the pathway that supplies NADPH for biomass formation in almost all the microorganisms), therefore the pentose phosphate pathway was included in our network, even if it was not mentioned by García Martín and co-workers.

In order to model the biomass growth in aerobic conditions, lumped reactions representing the drain of metabolites from the central carbon metabolism for the synthesis of proteins, nucleic acids, lipids and extracellular polysaccharides were included in the model, these equations were taken from the literature (Bordel & Nielsen 2010) and are only approximate. The units of the biomass components (proteins, lipids, etc.) in the stoichiometric equations are kg (all the other species are expressed in moles).

The resulting metabolic network has 73 reactions and 52 internal metabolites, which can be consulted in detail in the supplementary material.

**Flux balance analysis**

Flux balance analysis (FBA) is a common modelling technique used to predict the behaviour of metabolic networks (Edwards et al. 2002; Price et al. 2005). It is based on the assumption of steady state for the internal metabolites. In our case, the polymers that are being degraded or synthesized and accumulated are formally considered as external metabolites. The stoichiometric coefficients of each internal metabolite in each of the reactions in the network can be arranged forming a so-called stoichiometric matrix. The steady state condition implies that for each internal metabolite its production rate has to be equal to its consumption rate. In matrix notation this condition is expressed as follows:

$$\mathbf{S} \mathbf{v} = \mathbf{0}$$

The vector \( \mathbf{v} \) is the vector of rates for each of the reactions in the network and represents the distribution of metabolic fluxes. The previous equation imposes a limitation on the flux distributions attainable by the network, the flux distribution must belong to the null space of the stoichiometric matrix. Some of the reactions included in the

S. Bordel | Insights into phosphate-accumulating organism network
network are considered as irreversible under physiological conditions. The irreversibility condition restricts even more the set of attainable flux distributions in the metabolic network.

\[ v_i \geq 0 \] (2)

More restrictions can be added depending on the environmental conditions in which the metabolic network is operating, for example, under anaerobic conditions, the reaction corresponding to oxygen uptake should be set to zero.

Once all the restrictions have been specified, the metabolic network can be interrogated in order to obtain the theoretical yields achievable by the organism. For example, if the acetate uptake rate is set to \(1\ \text{mol h}^{-1}\) per gram of biomass dry weight and the propionate uptake rate is set to zero, we can maximize the production of PHB subject to all the previous restrictions and obtain the corresponding flux distribution. These optimizations were solved using the Mosek optimization package.

**RESULTS AND DISCUSSION**

Production of PHAs from acetate and propionate in anaerobic conditions

In our metabolic network, we have included three PHAs, namely polyhydroxybutyrate (PHB), polyhydroxyvalerate (PHV) and polyhydroxy-2-methylvalerate (PH2MV). Using FBA we have calculated the maximal amounts of each of these polymers that can by synthesized per unit of acetate and propionate consumed, as well as the amounts of glycogen and polyphosphate that are consumed in each of the cases.

The results can be seen in Table 1, all the units are moles of monomer per mol of the corresponding carbon source.

**Table 1**

<table>
<thead>
<tr>
<th>PHB produced</th>
<th>PHV produced</th>
<th>PH2MV produced</th>
<th>Glycogen consumed</th>
<th>Polyphosphate consumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal PHB from acetate</td>
<td>0.5333</td>
<td>0</td>
<td>0</td>
<td>0.0667</td>
</tr>
<tr>
<td>Maximal PHV from acetate</td>
<td>0</td>
<td>0.333</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximal PH2MV from acetate</td>
<td>0.1667</td>
<td>0</td>
<td>0.1667</td>
<td>0</td>
</tr>
<tr>
<td>Maximal PHB from propionate</td>
<td>0.0333</td>
<td>0</td>
<td>0.5</td>
<td>0.0667</td>
</tr>
<tr>
<td>Maximal PHV from propionate</td>
<td>0</td>
<td>0.1667</td>
<td>0.333</td>
<td>0</td>
</tr>
<tr>
<td>Maximal PH2MV from propionate</td>
<td>0</td>
<td>0</td>
<td>0.5152</td>
<td>0.0606</td>
</tr>
</tbody>
</table>

Origin of the NADPH necessary for PHA biosynthesis

The obtained results give some relevant insights into some controversial metabolic aspects of EBPR. The source of the reductive power (in form of NADPH) necessary for PHA synthesis is not completely clear (Seviour et al. 2003). It has been traditionally assumed that it comes from the degradation of glycogen; however the involved pathways are unclear.

García Martín and co-workers postulate that part of the reducing power is comes from the Embden-Meyerhof glycolytic pathway, which is present in the genome of *A. phosphatis*. The Embden-Meyerhof pathway generates two NADH molecules per mol of glucose and no NADPH, which on one hand is not enough reducing power to explain the observed formation of PHAs (Seviour et al. 2003) and on the other hand would require an extra mechanism of transformation of NADH into NADPH. It has been proposed (Lemos et al. 2003) that the remaining reductive power is generated by the TCA cycle. García Martín and co-workers pointed out that the reduced quinone generated by the succinate dehydrogenase in the TCA cycle cannot be reoxidized in the respiratory chain in anaerobic conditions. Therefore they propose the function of a putative cytochrome able to transfer electrons from the reduced quinone (\(\text{QH}_2\)) to NADPH or NADH (which are treated without distinction by García Martín and co-workers). We would like to stress that not only the reduction of \(\text{QH}_2\) generated in the TCA cycle is problematic but also the NADH generated by malate dehydrogenase and oxoglutarate dehydrogenase. Only isocitrate dehydrogenase has an isoenzyme that uses NADP as a cofactor and generates NADPH that can be directly used in PHA production. An alternative scenario suggested by García Martín and co-workers is the operation of a split TCA cycle, something typical in other organisms in anaerobic conditions. The FBA simulations that we have performed provide new insights into this topic.

When the production of PHB is maximized using acetate as the only external carbon source, most of the...
necessary NADPH for its synthesis comes from glycogen degradation via the pentose phosphate pathway. The NADH formed by the glyceraldehyde-3-phosphate dehydrogenase can be oxidized in two different energy consuming cycles. The first one involves the putative cytochrome proposed by García Martín and co-workers and is driven by the proton gradient across the cell membrane.

\[
\text{SUC} + Q \rightarrow \text{FUM} + \text{QH}_2 \\
\text{FUM} + \text{NADH} \Rightarrow \text{SUC} + \text{NAD} \\
\text{H}^+ \text{(ex)} + \text{QH}_2 + \text{NADP} \Rightarrow \text{NADPH} + \text{Q} + \text{H}^+ \text{(in)}
\]

The second cycle involves the malic enzyme (MalDH) and consumes ATP.

\[
\text{OAA} + \text{NADH} \Rightarrow \text{MAL} + \text{NAD} \\
\text{Pyr} + \text{CO}_2 + \text{ATP} \Rightarrow \text{OAA} + \text{ADP} + \text{P} \\
\text{MAL} + \text{NADP} \Rightarrow \text{NADPH} + \text{CO}_2 + \text{Pyr}
\]

The last cycle is consistent with a branched operation of the TCA cycle and does not involve putative or unclear reactions; therefore we consider it to be more likely than the first one.

In order to test the necessity of the pentose phosphate pathway to provide the reductive power supplied by glycogen degradation, we removed it from our network and repeated the same simulation. Without the pentose phosphate pathway the rate of glycogen degradation was zero. Also zero glycogen consumption is observed when the production of PHV or PH2MV is maximized using acetate as a sole carbon source. In order to produce PHV or PH2MV from acetate, propionyl-CoA needs to be synthesized. This is done in two steps from succinyl-CoA, which is generated in the oxidative branch of a branched TCA cycle. Succinyl-CoA is transformed into methylmalonyl-CoA by the enzymes mcm or meaB and is then transformed into propionyl-CoA by mmd.

In the oxidative branch of the TCA cycle NADPH is generated by the isocitrate dehydrogenase and NADH by the oxoglutarate dehydrogenase. This NADH can be oxidized in the same two cycles mentioned before. As it was mentioned before, we consider more likely the cycle involving the malic enzyme.

Production of PHV in reactors using acetate as the sole carbon source

As it was previously mentioned, the formation of amounts of between 5% and 20% of the total PHA content in reactors fed only with acetate has been reported (Schuler & Jenkins 2003). Here we aim to provide a kinetic explanation for this fact. As we have seen in our simulations, PHV from acetate is formed without glycogen consumption.

From the previous simulations we obtained the amounts of glycogen and polyphosphates consumed for the production of PHB and PHV. For a system in which the production rates of PHB and PHV are \( r_{\text{PHB}} \) and \( r_{\text{PHV}} \) respectively, the glycogen and polyphosphate degradation rates are the following ones:

\[
-r_{\text{gly}} = \frac{0.0667}{0.333} r_{\text{PHB}} \tag{3}
\]

\[
-r_{\text{PolyP}} = \frac{1.0667}{0.333} r_{\text{PHB}} + \frac{1}{0.333} r_{\text{PHV}} \tag{4}
\]

If we divide Equation (4) by Equation (3) we obtain the following relation:

\[
\frac{r_{\text{PolyP}}}{r_{\text{gly}}} = \frac{1.0667}{0.0667} + \frac{0.333}{0.333}(0.0667) \frac{r_{\text{PHV}}}{r_{\text{PHB}}} \tag{5}
\]

This equation allows us to relate the ratio of PHV and PHB production to the ratio between the kinetics of glycogen degradation and the utilization of polyphosphates. For example, in a sludge fed with acetate, in which a mole of PHV is made per 10 mol of PHB, 18.39 mol of polyphosphate are used per mole of glycogen (when we talk of moles of polymers we mean the moles of the corresponding monomer). In order for only PHB being produced, the rate of polyphosphate utilization should be 15.99 times higher than the rate of glycogen utilization. A possible explanation of the accumulation of PHV could be that the maximal rate for the hydrolysis and utilization of glycogen is lower than 1/15.99 times the polyphosphate utilization rate and therefore the extra ATP generated from the polyphosphate degradation is used in PHV synthesis instead of PHB (Figure 1).

Some insights into process design and operation

Process design is fully out of the scope of this paper; however we would like to provide an example of how reconstructed metabolic networks and FBA can help the design and operation of activated sludge processes. First of all we should note that kinetics is absent from our model and the metabolic network contains only stoichiometric information, therefore it can predict yields but not rates. Kinetic information could be added once the uptake or
production rate of one or more compounds have been measured and is introduced into the model as a constraint. The absence of kinetic information does not allow predicting residence times but the stoichiometric information contained in the metabolic network can for example help to decide the optimal recirculation of biomass from the aerobic to the anaerobic reactor (Figure 2).

Let’s assume that 1 mol of acetate and 1 mole of propionate are fed to the anaerobic tank. The residence time in the reactor is considered to be enough for the total utilization of both carbon sources. Our metabolic network allows us to estimate the amounts of PHAs that can be obtained from these carbon sources as well as the amount of phosphorous released by the biomass during the anaerobic phase.

A first approach to estimate the amounts of PHAs accumulated could be maximizing the sum of PHB and PH2MV subject to the restrictions of zero oxygen uptake and maximal acetate and propionate uptakes equal to 1. The results of this simulation are shown in Table 2.

The PHA amounts that enter the aerobic tank are 0.5333 mol of PHB and 0.5152 mol of PH2MV. We also found that 2.1578 mol of phosphate were released in order to consume 1 mol of acetate and 1 mole of propionate. Therefore the biomass recirculated from the aerobic tank must contain at least this amount of polyphosphate if we want to use completely both carbon sources. García Martín and co-workers reported 18% (in mass) phosphorous content in the activated sludge that they used for their study. If we take the same number and considering that 2.1578 mol of phosphorous correspond to 66.82 g of phosphorous, the minimal amount of biomass recirculated from the aerobic tank in order to assimilate completely both carbon sources should be 371.2 g of biomass. In an activated sludge, not all the biomass will belong to the PAOs. For our example

![Figure 1](https://iwaponline.com/wst/article-pdf/64/12/2410/443426/2410.pdf) | The synthesis of PHAs requires ATP and NADPH, the ATP can be generated in anaerobic conditions from the degradation of polyphosphate, the origin of NADPH is supposed to be glycogen degradation but the mechanisms are unclear.

![Figure 2](https://iwaponline.com/wst/article-pdf/64/12/2410/443426/2410.pdf) | A possible configuration of an EBPR process. Some biomass rich in polyphosphates is recirculated from the aerobic tank to the anaerobic tank. This biomass assimilates acetate and propionate and accumulates PHAs in the anaerobic tank. In the anaerobic tank the PHAs are used as carbon and energy sources for cell division and polyphosphate accumulation. Part of the biomass rich in polyphosphates is removed from the system together with the phosphorous that it contains.
we assumed that the percent of biomass belonging to the *A. phosphatis* is 60% (this corresponds to the estimation of García Martín and co-workers for one of their DNA samples). Therefore the amount of *A. phosphatis* biomass that needs to be recirculated will be 222.8 g.

Once we have estimated the biomass that needs to be recirculated into the anaerobic tank, we can estimate how much biomass can be removed from the system and therefore the amount of phosphate that can be eliminated per mol of acetate and propionate consumed.

In order to estimate the amount of phosphate that can be removed we need to have an estimation of the biomass composition. In our example we will use approximate numbers typical for many microorganisms. The phosphorous percent has been already mentioned to be 18% of the biomass dry weight. The phosphorous percent in the *A. phosphatis* (which counts for 60% of the total biomass) will be 30%. If we consider the total mass of the polyphosphates (adding the oxygen and hydrogen) they would make up to 60% of the total *A. phosphatis* biomass. The protein content is typically around 50% of the total dry weight of microorganisms. In this case we considered it to be 50% of the remaining biomass, therefore a total of 20%. In the same approximate way we proposed a lipid content of 3%, a glycogen content of 7% and a content of extracellular polysaccharides of 10%. Following the previous considerations the following biomass equation was added to our metabolic network:

\[
0.2 \text{ PROT} + 0.03 \text{ LIPID} + 0.1 \text{ EPS} + 6.818 \text{ PolyP} + 0.388 \text{ Glycogen} \Rightarrow \text{ BIOMASS}
\]

The glycogen and polyphosphates are expressed in moles to be consistent with the rest of the stoichiometric equations, the rest of the species are expressed in kg.

We should keep in mind that not only the glycogen and the polyphosphates to be added to the newly formed biomass have to be synthesized in the aerobic tank, but also the amounts that were consumed in the anaerobic phase have to be replenished. Therefore the optimization problem to be solved is the maximization of the biomass production subject to the restrictions in Table 3.

The result was 31.3 g of new *A. phosphatis* biomass formed. From this number we can estimate the fraction of biomass that needs to be recirculated as well as the phosphate removed.

If the amounts of phosphate that are entering the system are higher, more acetate and propionate should be added to the anaerobic reactor in order to achieve its full assimilation into the newly produced biomass.

In order to study the effects of carbon sources, the same simulations have been repeated using 2.5 mol of acetate and 1.66 mol of propionate (the quantities have been selected in order to have the same number of carbon atoms than in the previous example).

The results in Tables 4–6 reveal that the performance is lower when only acetate is fed to the system.

Propionate utilization results in a more efficient phosphorous removal, compared to acetate.

It is important to highlight that the previous calculations have been made under the assumption that PHAs are the only energy and carbon source used by the cells during the aerobic phase. In the real scenario, the *A. phosphatis* cells could be taking up a significant part of their biomass constituents (aminoacids, lipids, etc.) from the lysis products.

### Table 2 | Estimations of the PHA accumulation in an anaerobic reactor fed with a unit of acetate and a unit of propionate, all the units are expressed in moles

<table>
<thead>
<tr>
<th>Acetate consumption</th>
<th>Propionate consumption</th>
<th>PHB production</th>
<th>PHZMV production</th>
<th>PHV production</th>
<th>Polyphosphate consumption</th>
<th>Glycogen consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5333</td>
<td>0.5152</td>
<td>0</td>
<td>2.1578</td>
<td>0.1273</td>
</tr>
</tbody>
</table>

### Table 3 | Constraints used to calculate the biomass of *A. phosphatis* produced when 1 mol of acetate and 1 mol of propionate are supplied to the system

<table>
<thead>
<tr>
<th>Maximal PHB consumption</th>
<th>Maximal PHZMV consumption</th>
<th>Minimal polyphosphate production</th>
<th>Minimal glycogen production</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5333</td>
<td>0.5152</td>
<td>2.1578</td>
<td>0.1273</td>
</tr>
</tbody>
</table>

### Table 4 | Recirculation necessities, production of *A. phosphatis* biomass and phosphorous removed, per mol of acetate and propionate supplied to the system

<table>
<thead>
<tr>
<th>Biomass of <em>A. phosphatis</em> recirculated</th>
<th>Biomass of <em>A. phosphatis</em> removed</th>
<th>Fraction recirculated</th>
<th>Mass of phosphorous removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>222.8 g</td>
<td>31.3 g</td>
<td>0.88</td>
<td>9.39 g</td>
</tr>
</tbody>
</table>

### Table 5 | Recirculation necessities, production of *A. phosphatis* biomass and phosphorous removed per 2.5 mol of acetate supplied to the system

<table>
<thead>
<tr>
<th>Biomass of <em>A. phosphatis</em> recirculated</th>
<th>Biomass of <em>A. phosphatis</em> removed</th>
<th>Fraction recirculated</th>
<th>Mass of phosphorous removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>272.7 g</td>
<td>25.8 g</td>
<td>0.91</td>
<td>7.74 g</td>
</tr>
</tbody>
</table>
and would be interesting to test experimentally to what extent this phenomenon is taking place.

**CONCLUSIONS**

A metabolic network of *A. phosphatis* based on the metagenomic study published by García Martín and co-workers was reconstructed. After using this network to perform FBA simulations we observed that the NADPH necessary for the synthesis of PHAs is likely to be supplied by glycogen degradation via the pentose phosphate pathway or alternatively by the oxidative branch of the TCA cycle. The NADH generated is likely to be oxidized in an ATP consuming cycle that involves the malic enzyme and generates NADPH. PHV can be synthesized from acetate without glycogen utilization. This result suggests the possibility that the observed PHV formation in sludges fed only with acetate is due to a lower utilization rate of glycogen compared to the utilization rate of polyphosphates. In order to illustrate how metabolic networks and FBA could be used in the design and operation of activated sludge processes we estimated the biomass recirculation and the phosphorous removed for a simple process configuration. The process was revealed to be more efficient when propionate was fed as a carbon source, in comparison to acetate. The metabolic network that we present could be easily extended to include other possible carbon sources other than acetate and it could be used as a simulation tool by researchers specialized in EBPR.

**REFERENCES**


First received 1 June 2011; accepted in revised form 25 July 2011