



OPEN

Streamlined identification of strain engineering targets for bioprocess improvement using metabolic pathway enrichment analysis

Joan Cortada-García¹, Rónán Daly², S. Alison Arnold³ & Karl Burgess^{1✉}

Metabolomics is a powerful tool for the identification of genetic targets for bioprocess optimisation. However, in most cases, only the biosynthetic pathway directed to product formation is analysed, limiting the identification of these targets. Some studies have used untargeted metabolomics, allowing a more unbiased approach, but data interpretation using multivariate analysis is usually not straightforward and requires time and effort. Here we show, for the first time, the application of metabolic pathway enrichment analysis using untargeted and targeted metabolomics data to identify genetic targets for bioprocess improvement in a more streamlined way. The analysis of an *Escherichia coli* succinate production bioprocess with this methodology revealed three significantly modulated pathways during the product formation phase: the pentose phosphate pathway, pantothenate and CoA biosynthesis and ascorbate and aldarate metabolism. From these, the two former pathways are consistent with previous efforts to improve succinate production in *Escherichia coli*. Furthermore, to the best of our knowledge, ascorbate and aldarate metabolism is a newly identified target that has so far never been explored for improving succinate production in this microorganism. This methodology therefore represents a powerful tool for the streamlined identification of strain engineering targets that can accelerate bioprocess optimisation.

Strain engineering and media optimisation have proven to be a successful tool for bioprocess improvement, allowing commercial feasibility of several biotechnological products. Two of the most well-known examples are penicillin—going from laboratory production soon after its discovery in 1928¹ to large-scale production in 1943 with 100- to 1000-fold increase in product titres (up to 1.8 g L⁻¹)²—and artemisinin acid—precursor of the antimalarial drug artemisinin, for which heterologous expression was demonstrated in *Saccharomyces cerevisiae* up to 100 mg L⁻¹ and was further improved to 25 g L⁻¹ titres^{3,4}.

With the advent of “-omics” technologies, system-wide analysis of biological samples is possible, increasing the potential for bioprocess improvement with better guided strain and media optimisation. The uses and advantages of each -omics discipline have been reviewed elsewhere⁵⁻⁷. In this work, the focus is placed on metabolomics, the global analysis of the metabolome⁸, which provides detailed information on the small molecules in a biological system, and therefore has the most relevance for small molecule bioprocess improvement. Some examples of different ways in which metabolomics has been used to enhance bioprocesses are given below.

Nitta et al. used unsupervised (principal component analysis) and supervised (orthogonal partial least squares regression) multivariate analysis on 74 targeted metabolites to find gene targets for the improvement of 1-butanol production in *Escherichia coli* (*E. coli*)^{9,10}. In the first of the two studies, acetyl-CoA was identified as a bottleneck, which was resolved with the overexpression and optimisation of the RBS region of the *atoB* gene, which converts acetyl-CoA into acetoacetyl-CoA, resulting in a significant improve in 1-butanol titres. In the second study, glyoxylate was found to accumulate. The knockout of *aceA*, the first gene in the glyoxylate shunt, resulted in a 39% increase in 1-butanol titres in the cultivation conditions tested. In a different study, Kawaguchi et al. used targeted metabolomics to find intracellular bottlenecks in the co-consumption of glucose and L-arabinose in *Corynebacterium glutamicum* ATCC 31831¹¹. The authors analysed 138 metabolites of central metabolism, with

¹Institute of Quantitative Biology, Biochemistry and Biotechnology, School of Biological Sciences, University of Edinburgh, Edinburgh EH8 9AB, UK. ²Institute of Infection, Immunity and Inflammation, Glasgow Polyomics, University of Glasgow, Glasgow G61 1QH, UK. ³Ingenza Ltd., Roslin Innovation Centre, Roslin EH25 9RG, UK. ✉email: karl.burgess@ed.ac.uk

their conclusions based on 29 of these metabolites. Their analysis led to the overexpression of *pyk* and deletion of the *araR* repressor, achieving the simultaneous consumption of both substrates at the same rate.

Other authors looked at a smaller set of targeted metabolites for strain optimisation. For example, George et al. used targeted metabolomics to look at eight compounds of the isoprenoid pathway in *E. coli* to improve the heterologous production of three C5 alcohols with potential use as biofuels, particularly by engineering the Shine-Dalgarno sequence of the *nudB* gene, significantly alleviating the bottleneck of isopentenyl diphosphate¹². Barton et al. performed a targeted ¹³C labelling analysis using LC-MS to look at the intracellular levels of two compounds of the 1,4-butanediol (BDO) formation pathway and five up-stream by-products¹³. This way, the authors found the last two steps to be the bottleneck of BDO formation. Subsequent genetic engineering of these two steps led to improved BDO formation.

Some studies have also looked at the availability of different target nutrients during the bioprocess in order to identify and prevent specific limitations to improve process performance. For example, Korneli et al. looked at 22 metabolites—19 of which were amino acids—and identified amino acid limitation during the production of GFP in *Bacillus megaterium*¹⁴, and Ho et al. looked at the composition of six fatty acids—and at least 11 other intracellular metabolites—produced by the marine microalga *Chlamydomonas* sp. JSC4 under different conditions of salinity and nitrogen starvation, allowing an increase in lipid accumulation by adjusting the cultivation conditions¹⁵.

In the examples mentioned above—and others^{16–19}—metabolomics was used as a targeted tool to look at specific metabolites and pathways, in most cases using unit mass resolution mass spectrometers—such as single- and triple-quadrupoles—thus limiting the potential of finding engineering targets for bioprocess improvement to those based on prior knowledge of the biological system. However, untargeted metabolomics offers the possibility to find targets for bioprocess optimisation in a wider metabolic context in a more unbiased fashion, particularly when used with high-resolution accurate mass (HRAM) mass spectrometry. For example, Xu et al. performed an untargeted metabolomics analysis to compare a *pho13Δ* mutant of *S. cerevisiae*—with a higher capacity to catabolise xylose—with its parental strain. A total of 134 intracellular metabolites were identified, and sedoheptulose-7P—a metabolite from the pentose phosphate pathway (PPP) outside carbohydrate metabolism—showed the most significant difference between both strains. Mutants overexpressing the PPP gene *TAL1* at different degrees were constructed, and their *TAL1* expression levels were positively correlated to their respective xylose consumption rates²⁰.

Another study used metabolite profiling to determine which nutrients were limiting in a bioprocess for recombinant IgG4 antibody production using Chinese hamster ovary (CHO) cells, a discovery experiment better suited to untargeted metabolomics due to the unknown composition of the medium²¹, and Xia et al. used untargeted and targeted metabolomics to compare the production by *Streptomyces tsukubaensis* of FK506—a polyketide used as immunosuppressant—in a high and low productivity media²². Multivariate analysis of the metabolomics results with partial least squares (PLS) allowed the authors to identify and supplement limiting nutrients in the medium, increasing FK506 production. The 13 nutrients most significantly correlated with FK506 biosynthesis included coenzyme A (CoA) esters, shikimate, amino acids, pyruvate, lactate and PPP intermediates. Once again, these would have been difficult to identify by targeting the product biosynthetic pathway.

As the examples above show, untargeted metabolomics offers the possibility to identify key metabolites for bioprocess improvement that are outside the product biosynthetic pathway, which are commonly missed when targeted methods are used focusing on prior biological knowledge about the system. However, untargeted metabolomics produces very large datasets from which it is challenging to prioritise metabolic reactions and pathways for modification. A useful tool for dealing with and obtaining useful information from untargeted metabolomics data is pathway enrichment analysis, which analyses groups of compounds that work together to carry out a biological process—like a metabolic pathway^{23,24}, and ranks them in terms of their statistical significance and importance. Although pathway enrichment analysis has historically been applied more frequently to genomics, transcriptomics and proteomics data^{25–27}, it has also been applied in the field of metabolomics. However, most of these publications are clinical studies^{28–32}.

There have been some recent studies where metabolic pathway enrichment analysis (MPEA) has been used with metabolomics data for bioprocess improvements. For example, Morris et al. compared four different fed-batch cultivation conditions resulting in differences in monoclonal antibody titres in a CHO bioprocess³³. The authors then manually looked in the KEGG database³⁴ at the pathways of the 10 most significant metabolites, as identified by PLS-discriminant analysis (PLS-DA), to find titre inhibitors and promoters that might be modulated by changing the bioprocess feeding strategy. With this approach, however, constraining pathway analysis to the 10 most significant metabolites significantly limits the vast analytical power of metabolomics. In another example, Alden et al. used untargeted metabolomics to find metabolites that accumulated in the culture medium of CHO fed-batch processes. Using MPEA with a modified Fisher's exact test, the authors found three pathways that were enriched in the cell line with the lowest cell density profile of their study, including aminoacyl-tRNA biosynthesis, tryptophan and histidine metabolism. Further investigation into 11 putatively annotated metabolites that accumulated in the culture medium led the researchers to hypothesise that products of tryptophan metabolism could behave as inhibitors of cell growth. The authors finished by suggesting targeting the tryptophan pathway with genetic engineering, or lowering the concentration of tryptophan in the cultivation medium³⁵. These examples show how MPEA can be used for bioprocess improvement, however, this is still a line of research that is underexplored, particularly for the identification and prioritising of genetic targets for bioprocess optimisation.

MPEA can be performed in different ways, similar to the examples described above for targeted and untargeted metabolomics. Mainly, the analysis can focus on comparing a case and a control group, such as high- and low-productivity conditions or strains^{9,10,15,20,22}, or focus on the dynamic changes of different metabolites throughout the course of the fermentation^{11,12,14,21}. Although not mutually exclusive, the first type of analysis can be more tailored to identifying media deficiencies, performance biomarkers and genetic differences between differently

performing strains, whereas the second type can be used to improve an established working fermentation process further by, for example, identifying the accumulation of by-products, the presence of inhibitors, the (in)activation of specific metabolic pathways or the depletion of substrates.

In this study, a commercial *E. coli* succinate production process was analysed with a combined targeted and untargeted metabolomics method using HRAM mass spectrometry, and the results were used to perform MPEA throughout the time course of the fermentation to identify potential targets for bioprocess optimisation in an unbiased fashion.

Results

Escherichia coli succinate fermentation process. Three *E. coli* dual-phase succinate fermentation replicates were performed and samples were taken throughout the course of the fermentation for metabolomics analysis with LC–MS. Furthermore, the extracellular concentration of glucose and the main fermentative products was also determined by HPLC–UV/Vis-RI analysis (Fig. 1). The fermentation process is split into two distinctive parts. The first phase is an aerobic batch process, where all the glucose from the medium is consumed for biomass growth, with little production of fermentative products. Under aerobic conditions with glucose excess, acetate is formed in *E. coli* due to overflow metabolism, and part of this acetate is excreted out of the cell to prevent osmotic stress due to the accumulation of negative charges in the cytoplasm. After glucose depletion, this acetate can be reimported inside the cell and be consumed as a source of carbon and energy^{36–38}. A small amount of lactate, malate, pyruvate and succinate was measured in one of the three replicates in the middle of the biomass exponential growth phase, potentially due to a temporary limited supply of oxygen. However, the concentration of these organic acids went back down before the end of the biomass growth phase. The second phase of the bioprocess is an anaerobic production phase (biotransformation), where additional glucose is added to the medium and this is converted into succinate and other fermentative products, being pyruvate the main by-product. The *E. coli* strain used in this process cannot sustain cell growth on glucose under anaerobic conditions, which can be appreciated by the plateau and slight decrease in the biomass wet cell weight (WCW) measurements during the anaerobic phase of the process.

LC–MS analysis of fermentation samples. The three fermentation experiments were analysed by LC–MS to try to identify targets for improving succinate production. A total of 13 samples from each replicate fermentation were analysed using both intracellular and extracellular fractions of the samples (see Supplementary file “Metabolomics data.xlsx”). The Principal Component Analysis (PCA) plot of the samples for both intra- and extracellular analysis shows good clustering of the different time points of the three fermentation replicates,

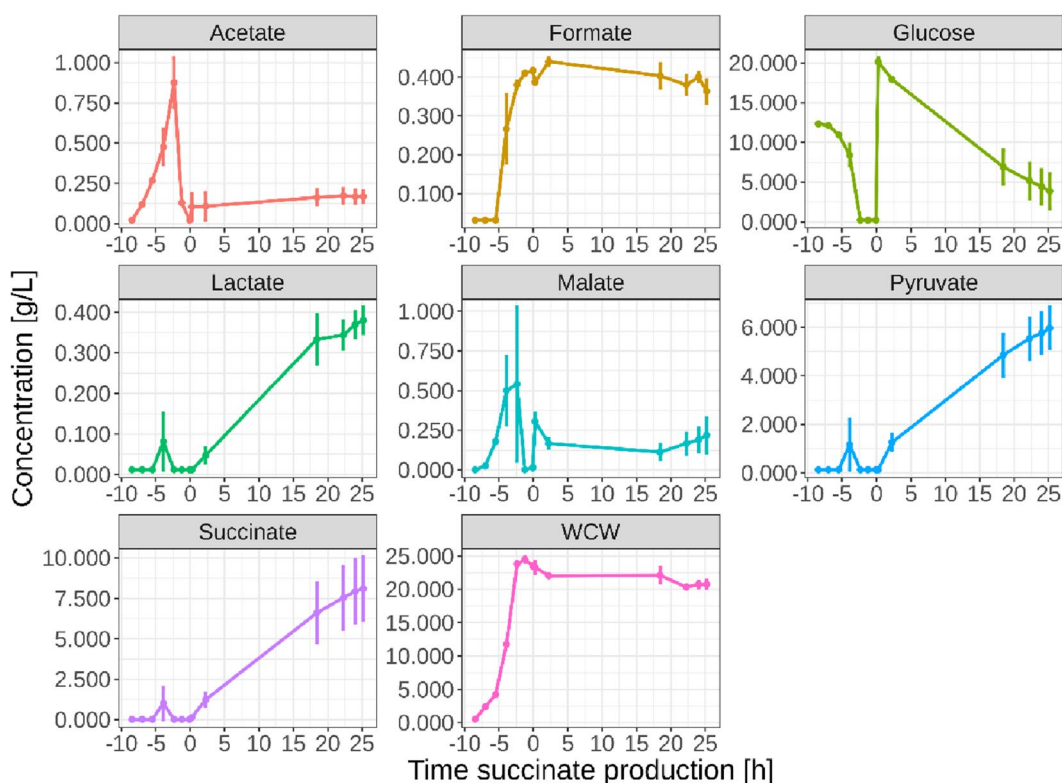


Figure 1. Profile of the *E. coli* succinate fermentation process showing the main parameters measured by HPLC–UV/Vis-RI. Time is indicated with respect to the beginning of the succinate production phase. Error bars represent the standard error of the mean of the three replicates ($n = 3$).

demonstrating reproducible data (Fig. S1 in the Supplementary materials), despite the lower succinate production for one of the three fermentation runs. The combined untargeted and targeted analysis resulted in a total of 6341 annotated and 92 identified metabolites (Table S1 in the Supplementary materials and see also the Supplementary file “Metabolomics data.xlsx”). Metabolite identification was performed by matching the retention time and accurate mass to reference standards following the Metabolomics Standards Initiative (MSI) classification system³⁹.

Metabolic pathway enrichment analysis. The metabolomics results were used to perform pathway enrichment analysis in order to find potential genetic targets for bioprocess optimisation without bias from preconceived process-specific biological knowledge. MPEA was performed using the Pathway Activity Level Scoring (PALS) application in the Polyomics integrated Metabolomics Pipeline (PiMP) online platform⁴⁰ (date of use: 08 Jul 2020). PALS ranks significantly changing metabolite groups over different sample sets using the pathway level analysis of gene expression for metabolomics (mPLAGE) algorithm, which converts m/z features into formulae and matches them to pathways⁴¹ and the KEGG library³⁴ (see Supplementary file “Pathway enrichment analysis data.xlsx”). The output of the mPLAGE algorithm is a p-value that has taken into account the multiple comparisons over all the pathways in the KEGG library. The different time points of the fermentation data were grouped by triplicate measurements and used for pathway enrichment analysis. For every triplicate time point of the fermentation, the algorithm evaluates pathways that are significantly different when compared to the first triplicate time point. This way, every metabolic pathway assessed is assigned a significance level (p-value) for every fermentation sample (except for the first time point). Each p-value indicates the probability that there is no difference between the time points for the metabolites on the pathway. The first sample analysed is a time point at the early stages of the aerobic phase of the fermentation (right after inoculation for extracellular samples and 3 h after inoculation for intracellular samples. See the “Methods” section for more details). By doing a time-course comparison to an early time point in the fermentation where the cells had abundant access to glucose and oxygen, it is possible to identify changes in metabolic regulation at the different stages of the fermentation (e.g. end of biomass growth, beginning or end of succinate production phase, etc.). This information can be used to put the results in the context of the experiment and aid in the identification of engineering targets.

To find genetic targets for improved succinate production, the mean p-value of each pathway (as compared to the first time point) was calculated across all time points in the succinate production phase. This mean p-value was used as ranking criteria (rather than focusing on extremely small p-values at individual time points) in order to find pathways with consistent statistical significance throughout the production phase and avoid the impact of potential outliers. Intracellular and extracellular metabolite fractions were analysed separately. The ten pathways with the highest level of significance (lowest mean p-value) resulting from analysing intracellular metabolites are represented as a heatmap in Fig. 2. These include several pathways from amino acid, carbohydrate, nucleotide and vitamin metabolism.

More significant changes were observed in the extracellular fraction samples (Fig. S2 in the Supplementary materials). This is caused by a significant increase in the number of metabolites found in the extracellular fraction as the fermentation progresses—potentially due to cell lysis, particularly under oxygen limitation or absence, because the *E. coli* strain used cannot sustain biomass growth on glucose anaerobically. The large number of significant pathways makes the selection of specific ones with significant changes in metabolic levels based on the extracellular fraction more challenging. For this reason, the selection of potential genetic targets was done purely based on the MPEA performed with the intracellular metabolites.

Selection of potential genetic targets. Pathways showing high metabolic changes during the succinate production phase could be potential targets for metabolic reshuffling. Either due to the removal of by-products or the elimination of metabolic bottlenecks, the modulation of these pathways could potentially lead to higher titres or yields. From the 10 pathways in Fig. 2, three pathways were identified to have particularly low p-values during the succinate production phase: the pentose phosphate pathway, ascorbate and aldarate metabolism and pantothenate and CoA biosynthesis (marked with red boxes in Fig. 2). The criteria to select these three particularly was that they all had at least three time points in the succinate production phase with a p-value ≤ 0.01 .

The metabolomics results for these pathways are shown in Fig. 3. Several metabolites from the PPP show a sudden increase in intracellular levels at the beginning of the succinate production phase and in many cases the levels remain high, with the exception of sedoheptulose-7P. The sudden increase at the beginning of the production phase is also observed for the precursors of pantothenate. However, pantothenate itself shows a decrease in intracellular intensity at the transition from the aerobic growth phase to the anaerobic production phase. Similarly, CoA levels drop significantly between 2 and 18 h of succinate production, indicating that CoA availability can be a potential bottleneck for succinate production, as suggested by Lin et al.⁴² Finally, 5-dehydro-4-deoxy-D-glucarate and 2-oxoglutarate—both involved in ascorbate and aldarate metabolism—show a sudden increase in intracellular levels at the beginning of the production phase, followed by a sudden decrease. However, L-gulonono-1,4-lactone shows a progressive increase in intracellular intensity throughout the succinate production phase, indicating intracellular accumulation. Trying to minimise this accumulation can, therefore, be another potential genetic target to improve succinate production.

PLS-DA for identifying targets of succinate improvement. Other authors have used PLS regression to identify targets for bioprocess engineering^{9,22}. Therefore, PLS-DA was chosen as a benchmark analytical method to compare to the findings of the metabolic pathway enrichment analysis. For this, the three fermentation replicates described in this work were split into two groups, the first group containing the two experiments with higher succinate titres and the second group containing the experiment with a lower titre (10.05 ± 0.64 SD

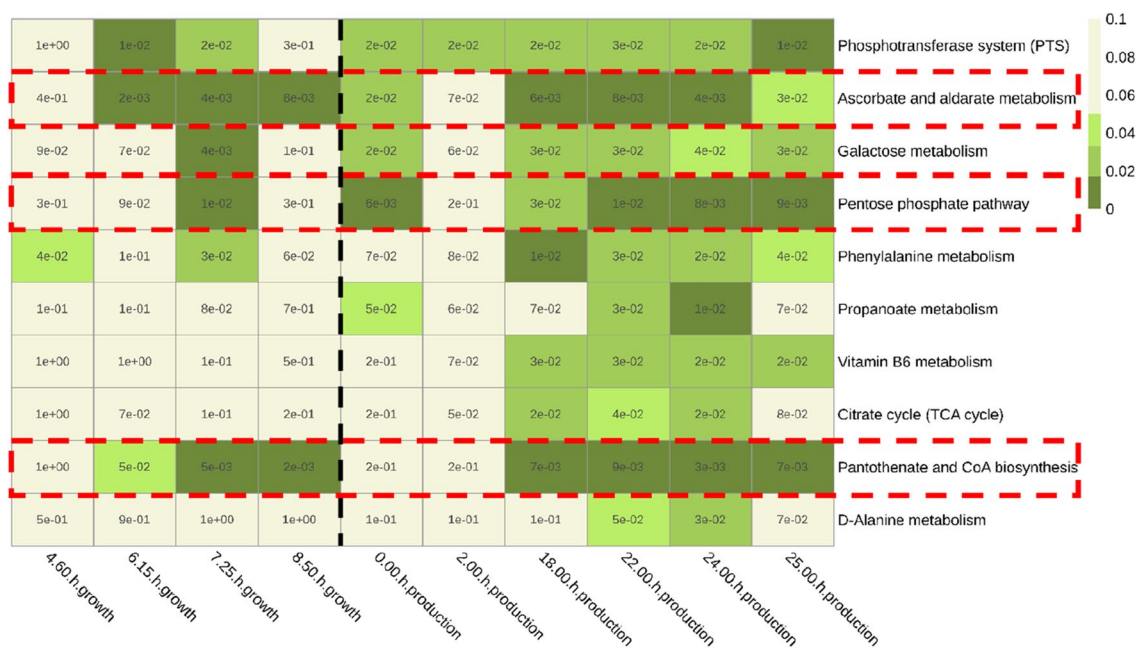


Figure 2. Top 10 pathways with the lowest mean p-value in the succinate production phase based on MPEA using the mPLAGE algorithm for intracellular metabolites found with a combined targeted and untargeted metabolomics method. Pathway names are shown on the y-axis (in increasing average p-value in the production phase going down) and time points of the fermentation process on the x-axis. The values of the cells are the p-values ($n = 3$) comparing each time point to the first time point of the fermentation analysed (3 h after the beginning of the aerobic batch phase). The resulting colour corresponds to the level of significance: green (p -values ≤ 0.05) and beige (p -values > 0.05). The vertical black dashed line indicates the transition from the aerobic growth phase to the anaerobic succinate production phase. The horizontal red dashed lines indicate the three pathways highlighted as potential targets for strain engineering.

and 4.23 g/L final succinate concentration, respectively). Then, the samples from the succinate production phase of the bioprocess were analysed by PLS-DA to identify which metabolites contribute the most to the differences in succinate titre (Fig. 4). The samples from the aerobic growth phase were not included in the analysis to avoid discrimination based on the large metabolic differences between the aerobic biomass growth phase and the anaerobic succinate production phase. Succinate was also removed from the data set before performing PLS-DA in order to prevent biasing group discrimination. Good separation of the two sample groups was observed (Fig. 4, top), and the four first latent variables were selected to build the model, based on maximising the cross-validated coefficient of determination ($Q^2 = 0.7123$, $R^2 = 0.9738$, Fig. 4, bottom).

From a total of 3989 mass spectrometry signals analysed in PLS-DA, 495 were found significant with a variable importance in projection (VIP) score > 1 ⁴³ in the first latent variable (Supplementary file PLSDA summary.xlsx). The VIP score is a common measure in PLS-DA of the relative importance of each metabolite with respect to the total variation of the latent variable. From these significant signals, 270 had unknown annotations and many of the 225 remaining had multiple possible metabolic annotations. Those metabolites with the highest level of annotation confidence and a VIP score above 1 in the four first latent variables were investigated. Metabolites with a high level of annotation confidence were defined as either having retention time and accurate mass matched to a reference standard, or accurate mass and fragmentation pattern matched to metabolomics library data. The 38 metabolites that met the criteria above are gathered in Table S2 (see Supplementary file PLSDA summary.xlsx), and most of them belong to amino acid metabolism (18 out of the 38 metabolites) and nucleotide metabolism (10 out of 38), which is a much more reduced breadth of metabolic pathways than the findings from Fig. 2. Nevertheless, coinciding with the results from MPEA, three metabolites involved in pantothenate and CoA biosynthesis were found important for discrimination of a successful *E. coli* succinate fermentation run according to the PLS-DA model, namely D-pantothenic acid, L-alanine and L-valine (Fig. 5), strengthening the hypothesis that this pathway could be a potential target of interest for bioprocess optimisation. While L-alanine and L-valine show clearer differences between both sample groups (higher levels for the higher succinate titre group), the levels of D-pantothenic acid are more similar in both groups, which is also reflected in the latter having a lower VIP score in the first two latent variables (see Supplementary file PLSDA summary.xlsx). However, it is worth mentioning that PLS-DA looks at individual metabolites and does not reveal any biological information of the relationship between them, whereas enrichment analysis groups features (metabolites in this case) under common biological themes, making it more robust to reveal real biological differences between sample groups⁴⁴. That is, for analysing the PLS-DA results, the three abovementioned metabolites were manually linked to the same metabolic pathway assisted by the findings from MPEA from Fig. 3. In other words, MPEA makes it easier to put metabolomics results in a biological context.

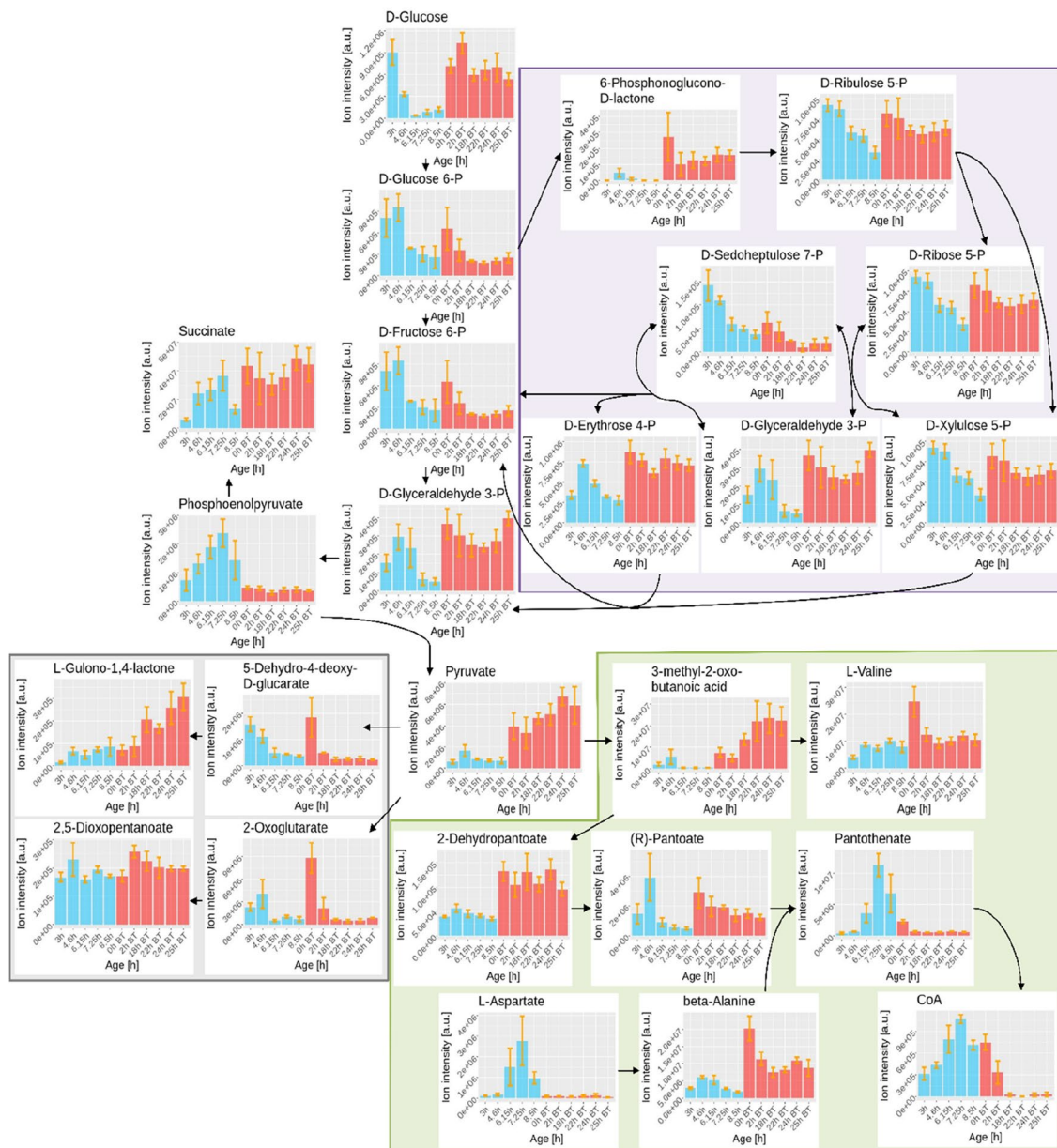


Figure 3. Dynamic evolution of intracellular metabolites of the three pathways identified as potential targets for genetic engineering: the pentose phosphate pathway (purple), pantothenate and CoA biosynthesis (green) and ascorbate and aldarate metabolism (grey). A simplified version of the pathways is shown for ease of reference. The results show both phases of the bioprocess: the aerobic bacterial growth phase (blue) and the succinate production phase (red). Orange error bars represent the standard error of the mean ($n = 3$).

Discussion

Metabolomics is an expanding field that has been previously applied to identify potential targets for genetic engineering to improve bioprocesses^{9–13,20}. However, in most cases, this is still performed with targeted metabolomics using unit mass resolution single- or triple-quadrupole instruments looking at pre-selected intermediates from the biosynthetic pathway of the product of interest, thus limiting the vast detection capacity of mass spectrometry. Some examples are available using untargeted metabolomics^{20–22}, however, these tend to identify potential genetic targets using multivariate statistical tools—often PCA or PLS—applied directly to the metabolomics data without using information from metabolic pathways. As this study shows, without the context of biological pathways it can be very hard to interpret and extract useful information from untargeted metabolomics data using multivariate statistical analysis. This was demonstrated by developing a PLS-DA model to discriminate between higher and lower succinate titre runs in an *E. coli* bioprocess, based on targeted and untargeted intracellular metabolomics data. Overall, although the PLS-DA model was able to achieve good separation of the two sample groups, the large number of significant signals and the high level of uncertainty in metabolite annotation makes it difficult to identify genetic targets for succinate production improvement. Focusing on those metabolites with a higher confidence in annotation it was possible to identify metabolites from amino acid, nucleotide and pantothenate

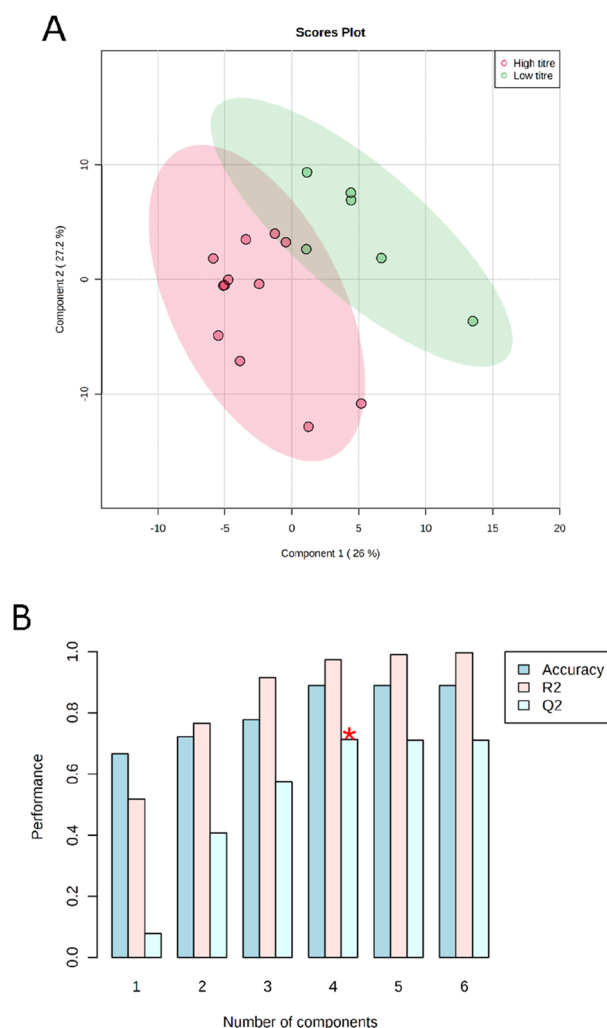


Figure 4. (A) Scores plot between the latent variables (components) 1 and 2 for the PLS-DA of the three fermentation replicates analysed, which were split into two experiments with a high succinate titre ($n = 12$) and one with low succinate titre ($n = 6$). The explained variances of each latent variable are shown in brackets. (B) PLS-DA classification using different number of latent variables (components). The red star indicates the best classifier based on maximising Q^2 .

and CoA metabolism as being important for group discrimination. However, this is a time- and effort-intensive exercise that left behind 448 of the 495 (90.51%) significant features according to the PLS-DA model.

Conversely, the use of MPEA allows to exploit metabolomics data further. By considering correlated changes in different compounds of specific pathways, certain enrichment analysis tools such as PALS can be used to analyse targeted and untargeted metabolomics to identify significantly modulated pathways during the bioprocess in a more streamlined way. Importantly, as MPEA highlights sets of metabolites that are consistently regulated, it is more robust to individually misannotated metabolites than techniques that look at individual metabolites without considering any biological context (such as PLS-DA). Furthermore, the PALS tool used in this work uses formulae rather than identifications to safeguard even more against the impact of potential misannotations.

The *E. coli* succinate production bioprocess was analysed with MPEA and the results pointed at three different pathways with significantly different levels of metabolites during the bioprocess: the pentose phosphate pathway, pantothenate and CoA biosynthesis and ascorbate and aldarate metabolism.

Interestingly, the PPP has previously been reported to play an important role during succinate production in *E. coli*. For example, Lu et al.⁴⁵ calculated the carbon flux down the PPP during succinate production using LC-MS and HPLC measurements and using a stoichiometric model in an *E. coli* AFP111 dual phase bioprocess using ^{13}C -labelled glucose as a carbon source. Using a higher percentage of CO_2 during the succinate production phase increased both succinate production and the percentage of carbon flux channelled towards the PPP. In another study, Zhu et al. found that increased activity of the enzyme transketolase in the PPP led to an increase in succinate titre and yield in *E. coli* ATCC 8739 derived strains. The succinate titre and yield improved even further increasing the activity of the enzyme transhydrogenase, which can convert NADPH generated from the PPP into NADH—which can then be used for succinate production⁴⁶. The same research group performed a

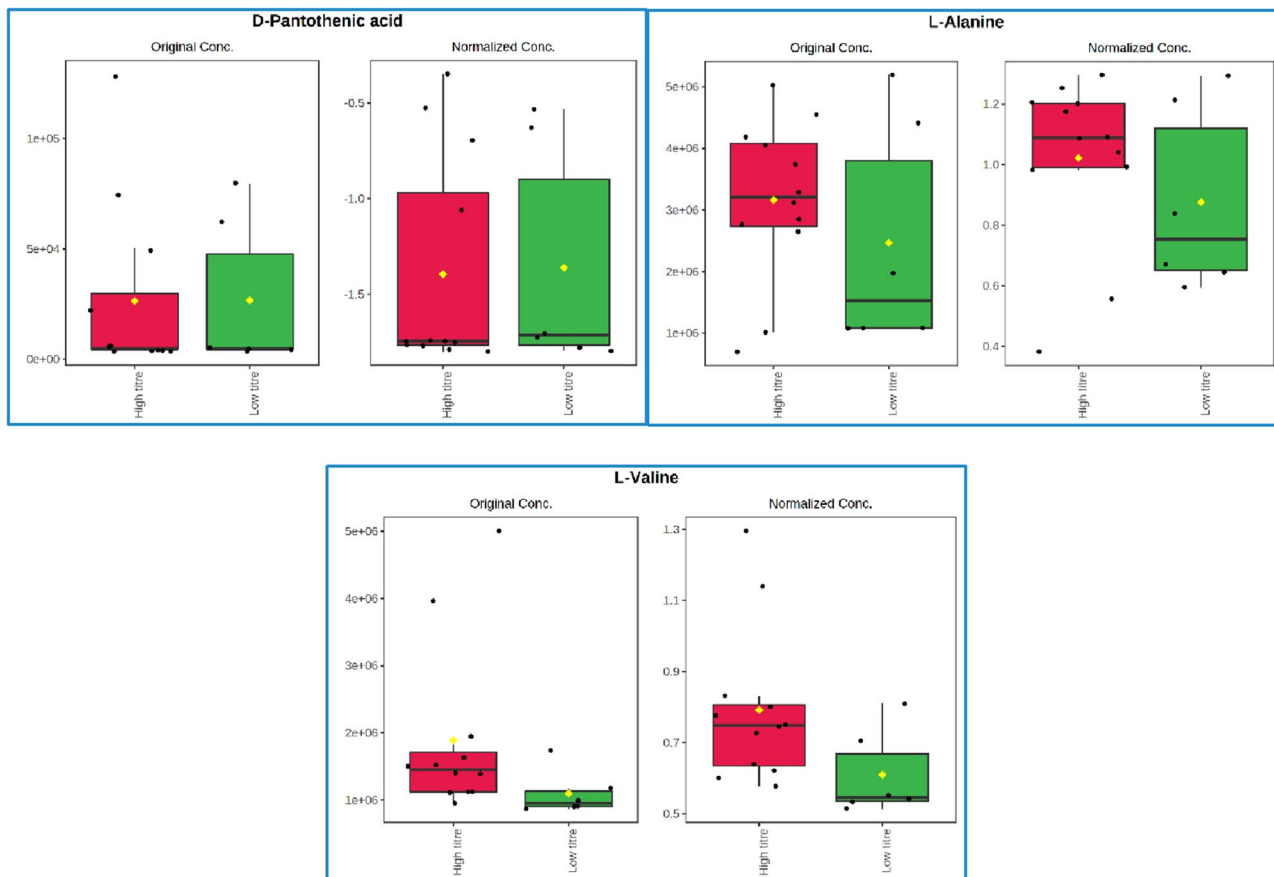


Figure 5. Three metabolites involved in pantothenate and CoA biosynthesis found important for discrimination of a successful *E. coli* succinate fermentation run according to the PLS-DA model (VIP score > 1 in at least one of the four latent variables) and with a high level of annotation confidence. The boxplots show the peak intensity of the three metabolites in the two sample groups: high (red) and low (green) succinate titre ($n = 12$ and 6 , respectively). The peak intensity values are shown as black dots, the black horizontal line represents the population median and the yellow diamond represents the population mean. For each metabolite, the subplot on the left shows the results with the original peak intensity data, and the subplot on the right the results after \log_{10} transformation and normalisation of the data.

further study where the authors generated a library of PPP genes under the regulation of a constitutive M1-93 promoter and different ribosome binding sites to generate different *E. coli* ATCC 8739 derived variants with different expression levels of the PPP genes. Their results indicate that increasing the expression levels of PPP genes can lead to higher succinate yields and titres, but that a fine tuning of the expression levels gives the best results⁴⁷. These examples are in agreement with the identification, in this work, of the PPP as a potential target for succinate improvement using MPEA, validating this methodology for the identification of pathways for bioprocess improvement.

There is less prior literature about manipulating the pantothenate and CoA biosynthetic pathways to increase succinate production. However, Lin et al. showed increased succinate production in *E. coli* GJT001 derived strains by overexpressing the enzyme pantothenate kinase, which catalyses the first step in the conversion of pantothenate to CoA⁴². Succinate production was increased even further with the co-overexpression of either phosphoenolpyruvate carboxylase or pyruvate carboxylase, which catalyse the conversion of phosphoenolpyruvate and pyruvate into oxaloacetate. The authors attributed the increase in succinate production to a higher intracellular availability of acetyl-CoA and CoA. In all experiments, the fermentation medium was supplemented with 5 mM pantothenate, therefore, genetic engineering to increase the levels of intracellular pantothenate would also be required on top of the changes indicated by the authors to increase succinate production. This study suggests that genetic manipulation of the pantothenate and CoA pathway can, indeed, lead to an increase in succinate production.

Finally, as far as ascorbate and aldarate metabolism is concerned, no examples were found in the literature showing the manipulation of this pathway in *E. coli* for increased succinate production. Therefore, this study might be the first to identify this pathway as a potential target for genetic engineering to achieve improved succinate production in *E. coli*.

This work is the first example of the application of MPEA using both targeted and untargeted metabolomics to identify potential strain engineering targets for bioprocess improvement. The examples above show that the targets identified with this methodology are relevant and coincide with previous successful attempts to improve

succinate production in *E. coli*. The three pathways identified in this work are outside the succinate biosynthetic pathway, showing how untargeted metabolomics can identify important pathways for product formation, even if these are initially not known to have an impact on the bioprocess. Untargeted metabolomics has the potential to accelerate bioprocess optimisation, and pathway enrichment analysis is a useful tool to help analyse and interpret the results obtained with this technology putting them in a biological context.

Looking forward, it is important to mention that, although MPEA is a powerful tool for the identification of significantly regulated pathways, it does not indicate how these need to be modified to achieve the desired strain phenotype. Indeed, biological regulation is complex, with many factors affecting metabolic and protein levels. Therefore, predicting and modulating phenotypes is still difficult even when knowing which metabolic pathways are highly regulated in a bioprocess. Nevertheless, MPEA can be a powerful tool to streamline and speed up the Design-Build-Test-Learn cycle of strain engineering, i.e. the identification of engineering targets.

Methods

Bacterial strain. All experiments described in this article were carried out using a proprietary industrial *E. coli* strain (Ingenza Ltd., UK) previously described⁴⁸, based on the *E. coli* NZN111 strain with deletions of the pyruvate-formate lyase (*pflB*) and lactate dehydrogenase (*ldhA*) genes⁴⁹.

Growth media. All 5 L scale fermentation experiments were carried out with a batch phase for biomass formation using a defined minimal medium containing 11.90 g/L glucose as the sole carbon source, 2.00 mM MgSO₄, a mix of salts solution (2.00 g/L (NH₄)₂SO₄, 14.60 g/L K₂HPO₄, 3.60 g/L NaH₂PO₄·2H₂O, 0.50 g/L (NH₄)₂H-citrate), a mix of trace elements (1.0 mg/L CaCl₂·2H₂O, 20.06 mg/L FeCl₃, 0.36 mg/L ZnSO₄·7H₂O, 0.32 mg/L CuSO₄·5H₂O, 0.30 mg/L MnSO₄·H₂O, 0.36 mg/L CoCl₂·6H₂O, 44.60 mg/L Na₂EDTA·2H₂O), antibiotics (100 mg/L kanamycin, 34 mg/L chloramphenicol) and antifoam (33.33 µL/L polypropylene glycol P-2000). Shake flask overnight cultures were prepared using the same medium but with 10.00 g/L glucose and no antifoam.

Fermentation process conditions. All fermentation experiments were carried out in a 5 L Applikon stirred tank fermenter (ADI 1030 Bio Controller, 1035 Bio Console), and the process consisted of an initial aerobic batch phase where the minimal medium was primarily used for biomass formation, followed by a 24 h anaerobic succinate production phase as previously described⁴⁸.

Inoculum. Fermentation inocula were prepared by inoculating 50 µL of cell bank into 100 mL of growth medium in a 500 mL baffled shake flask and incubated at 37 °C and 165 rpm for 17–17.5 h.

Aerobic batch phase for biomass growth. The fermentation was started by inoculating 100 mL of overnight culture into 3 L of growth medium in the 5 L fermenter for a starting OD₆₀₀ of 0.21 ± 0.025. During biomass growth the conditions were maintained at 37 °C temperature, 500–900 rpm agitation (controlled to keep the dissolved oxygen (DO) > 30%), 4.00 L/min air (1.33 vvm) and pH 7.0 ± 0.1, controlled with 2.00 M H₂SO₄ and 28% w/v NH₄OH.

Anaerobic succinate production phase. At the beginning of the production phase, glucose from a 500 g/L solution and sodium bicarbonate from a 100 g/L solution were added to the fermenter as a single bolus addition to a final concentration of 20 g/L and 5 g/L respectively in the vessel, as previously described⁵⁰. The sodium bicarbonate provides soluble CO₂, which is required for the conversion of PEP to oxaloacetate⁵¹. Once the glucose and sodium bicarbonate were added to the fermenter, the sparged air was replaced by pure (99.8%) CO₂ at 0.50 L/min (0.17 vvm), agitation was set to 300 rpm, temperature at 37 °C and pH at 7.0 ± 0.1 controlled with 2.00 M H₂SO₄ and 28% w/v NH₄OH.

Biomass measurement. Biomass levels were reported as OD₆₀₀ and wet cell weight (WCW). The former was the measured optical density at 600 nm wavelength. The latter was determined by spinning down 1 mL of sample for 5 min at 14,462 g twice in a pre-weighed Eppendorf tube, removing the supernatant and weighing the resulting pellet. The weight of the pellet in g/L was calculated from gravimetric difference.

Metabolites extraction for LC–MS analysis. Samples for off-line liquid chromatography–mass spectrometry (LC–MS) analysis were removed from the bioreactor and collected into dry-ice-cold universal vials, placed briefly on an ethanol dry ice bath for a fast cold sample quenching and immediately spun down twice at 4 °C and 13,000 g for 10 min. The supernatant and cell pellet were collected as extracellular and intracellular fractions respectively and stored at – 80 °C until further extraction for LC–MS analysis.

Extracellular fractions. Extracellular fraction extractions were prepared by diluting 10 µL of sample into 400 µL of 1:3:1 chloroform:methanol:water (C:M:W). The samples were then mixed vigorously in a chilled microtube mixer for 5 min and then centrifuged for 3 min at 13,000g and 4 °C. At this point, 360 µL of supernatant were transferred into a new microtube and stored at – 80 °C until LC–MS analysis. 25 µL of supernatant of each extracted sample were combined into one single vial to generate a pooled sample. During handling, the 1:3:1 C:M:W extraction solvent and the samples were kept on an ethanol dry ice bath.

Intracellular fractions. Prior to extraction, intracellular fractions were washed by resuspending the cell pellets in 1 mL of sterile phosphate buffer solution. The phosphate buffer was removed by spinning down the samples twice for 10 min at 13,000 g and 4 °C. For metabolite extraction, 200 µL of 1:3:1 C:M:W were added for every 5 mg of WCW pellet, thus normalising all samples by their biomass concentration. Cell pellets were resuspended by pipetting, and then the samples were mixed vigorously in a chilled microtube mixer for 1 h, before being centrifuged for 3 min at 13,000g and 4 °C. At this point, 200 µL of supernatant were transferred into a new microtube. The samples were further diluted by adding 200 µL 1:3:1 C:M:W and stored at – 80 °C until LC–MS analysis. 25 µL of supernatant of each extracted sample were combined into one single vial to generate a pooled sample. During handling, the 1:3:1 C:M:W extraction solvent and the samples were kept on an ethanol dry ice bath. The first two fermentation samples had a very small biomass pellet, which made the extraction process impractical, particularly the resuspension of the pellet in the amount of extraction solvent required for sample normalisation. For this reason, these first two time points from the intracellular fraction were not included in the analysis.

LC–MS method. Metabolite separation was performed using a zwitterionic hydrophilic interaction liquid chromatography (ZIC⁻-pHILIC) column (Merck SeQuant[®]) (150 mm × 4.6 mm, 5 µm particle size) equipped with the corresponding guard column (20 mm × 2.1 mm, 5 µm particle size) (Merck SeQuant[®]). A linear gradient was applied to the column, running from 80 to 20% solvent B over 15 min, followed by a 2 min wash with 5% solvent B, and 9 min re-equilibration with 80% solvent B, where solvent B was acetonitrile and solvent A (the remaining percentage) was 20 mM ammonium carbonate in water. The total flow rate was 300 µL/min, column temperature was maintained at 25 °C, sample injection volume was 10 µL, samples were maintained at 4 °C for the duration of the analysis and a HESI probe was used on the ion source.

Metabolite detection was done in a high-resolution Thermo Scientific[™] Q Exactive[™] Orbitrap mass spectrometer at 70,000 resolution, mass range 70–1050 m/z in polarity switching mode with a spray voltage of ± 3.8 kV. Capillary temperature was set to 320 °C, sheath gas 40 a.u., AGC target 1×10^6 a.u. and the lock masses in positive and negative mode were 144.9822 m/z and 100.9856 m/z, respectively.

Fragmentation was performed on pooled samples by isolating ions in a 1.2 m/z window and fragmentation with stepped HCD collision energy of 24.8, 60.0 and 94.8% for both polarities with 17,500 resolution and AGC target 1×10^5 a.u. Top 10 ions (intensity threshold 1.3×10^5) were selected for fragmentation and then added to a dynamic exclusion window for 15 s.

Metabolomics data processing and analysis. Raw mass spectrometry files were converted into .mzXML files in profile mode with the open-source software ProteoWizard⁵² (Version 3.0). Further data processing and analysis was performed using the PiMP online platform⁴⁰ (date of use: 08 Jul 2020). Peak detection and filtering were set to 3 ppm of the theoretical monoisotopic mass, minimum intensity to 5000, noise to 0.8, retention window to 0.05 and minimum number of detections to 3. Peak retention time was corrected using the Obiwrap algorithm⁵³ from the xcms package (Version 1.48.0).

Metabolite annotation and identification. Metabolite annotation was performed using a local copy of the KEGG database³⁴ (Metabolomics Standards Initiative (MSI) class 2/3 annotation³⁹), the fragmentation library (MSI class 2 annotation) and by matching the retention time and accurate mass to reference standards, which are listed in the Supplementary file Std_list_Metabolites.xlsx.

HPLC–UV/Vis-RI analysis. HPLC coupled to UV/Vis and refractive index detectors (HPLC–UV/Vis-RI) analysis was carried out using a Rezex[™] ROA Organic Acid H⁺ ion-exclusion column (Phenomenex[®]) (300 mm × 7.8 mm) equipped with a Carbo-H4 guard column (SecurityGuard[™]) (3.0 mm i.d.). An isocratic method was applied to the column, running a 5 mM H₂SO₄ mobile phase solution for 30 min. The total flow rate was 800 µL/min, column temperature was maintained at 65 °C, sample injection volume was 10 µL and samples were maintained at 4 °C for the duration of the analysis. The HPLC–UV/Vis-RI data was extracted as a .csv file and was further analysed using the ggplot2 package⁵⁴ (Version 3.3.3) in the statistical software environment R (Version 3.6.1). Metabolite levels below the limit of quantification were replaced by 1/5 of the minimum metabolite concentration quantified.

Metabolic pathway enrichment analysis. MPEA was performed using the PALS application in PiMP (date of use: 08 Jul 2020) using the mPLAGE algorithm and a local copy of the KEGG database³⁴. Pathways with p-values of 0.05 or below were considered statistically significant across experimental groups compared. The experimental groups compared were the different time points, and each pairwise comparison was done comparing any given time point with the first sample of the fermentation. In order to reduce the dimensionality of the analysis, for each pathway, the p-values of all the samples in the succinate production phase were averaged together and this mean p-value was used to find the 10 “most significant” pathways (lowest mean p-value). These 10 “most significant” selected pathways were displayed as a heatmap showing the p-value of each pathway for each sample. Heatmaps were created using the gplots package⁵⁵ (Version 3.1.3) in the statistical software environment R (Version 4.0.4). MPEA for intracellular and extracellular samples was performed separately.

PLS-DA. PLS-DA was performed in MetaboAnalyst (date of use: 08 Sep 2022) with R (Version 4.1.3). Constant features across all samples were deleted and missing values were replaced by 1/5 of the minimum positive values of their corresponding variables. Variables with near-constant values throughout the experiment con-

ditions were filtered using the interquartile range, and the data was \log_{10} transformed and normalised by the median to adjust for systematic differences among samples (Figs. S3 and S4 in the Supplementary materials). The number of latent variables for the PLS-DA model was chosen by maximising Q^2 using tenfold cross validation. Those features with a VIP score above 1 in the first latent variable were considered as significant for group discrimination when analysing the results.

Data availability

The data that support the findings of this study are available in the Supplementary materials and in MetaboLights at <http://www.ebi.ac.uk/metabolights/MTBLS6667>, reference number MTBLS6667.

Received: 16 February 2023; Accepted: 28 July 2023

Published online: 10 August 2023

References

- Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**, 236 (1929).
- Demain, A. L., Vandamme, E. J., Collins, J. & Buchholz, K. History of industrial biotechnology. In *Industrial Biotechnology: Microorganisms* (eds Wittmann, C. & Liao, J. C.) 3–84 (Wiley, 2016).
- Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**(7086), 940–943 (2006).
- Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
- Adamczyk, P. A. & Reed, J. L. *Escherichia coli* as a model organism for systems metabolic engineering. *Curr. Opin. Syst. Biol.* **6**, 80–88 (2017).
- Becker, J. & Wittmann, C. From systems biology to metabolically engineered cells—an omics perspective on the development of industrial microbes. *Curr. Opin. Microbiol.* **45**, 180–188 (2018).
- Zahrl, R. J., Peña, D. A., Mattanovich, D. & Gasser, B. Systems biotechnology for protein production in *Pichia pastoris*. *FEMS Yeast Res.* **17**, 68 (2017).
- Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378 (1998).
- Nitta, K. *et al.* Orthogonal partial least squares/projections to latent structures regression-based metabolomics approach for identification of gene targets for improvement of 1-butanol production in *Escherichia coli*. *J. Biosci. Bioeng.* **124**, 498–505 (2017).
- Nitta, K. *et al.* Metabolome analysis revealed the knockout of glyoxylate shunt as an effective strategy for improvement of 1-butanol production in transgenic *Escherichia coli*. *J. Biosci. Bioeng.* **127**, 301–308 (2019).
- Kawaguchi, H. *et al.* Metabolome analysis-based design and engineering of a metabolic pathway in *Corynebacterium glutamicum* to match rates of simultaneous utilization of D-glucose and L-arabinose. *Microb. Cell Fact.* **17**, 76–91 (2018).
- George, K. W. *et al.* Metabolic engineering for the high-yield production of isoprenoid-based C5 alcohols in *E. coli*. *Sci. Rep.* **5**(1), 1–12 (2015).
- Barton, N. R. *et al.* An integrated biotechnology platform for developing sustainable chemical processes. *J. Ind. Microbiol. Biotechnol.* **42**, 349–360 (2015).
- Korneli, C., Bolten, C. J., Godard, T., Franco-Lara, E. & Wittmann, C. Debottlenecking recombinant protein production in *Bacillus megaterium* under large-scale conditions—targeted precursor feeding designed from metabolomics. *Biotechnol. Bioeng.* **109**, 1538–1550 (2012).
- Ho, S. H. *et al.* Optimizing biodiesel production in marine *Chlamydomonas* sp. JSC4 through metabolic profiling and an innovative salinity-gradient strategy. *Biotechnol. Biofuels* **7**, 97–112 (2014).
- Jordà, J. *et al.* Quantitative metabolomics and instationary ^{13}C -metabolic flux analysis reveals impact of recombinant protein production on trehalose and energy metabolism in *Pichia pastoris*. *Metabolites* **4**, 281–299 (2014).
- Ohtake, T. *et al.* Metabolomics-driven approach to solving a CoA imbalance for improved 1-butanol production in *Escherichia coli*. *Metab. Eng.* **41**, 135–143 (2017).
- Liao, J. C., Hou, S.-Y. & Chao, Y.-P. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* **52**, 129–140 (1996).
- Hasunuma, T. *et al.* Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of *Saccharomyces cerevisiae*. *Microb. Cell Fact.* **10**, 1–13 (2011).
- Xu, H. *et al.* PHO13 deletion-induced transcriptional activation prevents sedoheptulose accumulation during xylose metabolism in engineered *Saccharomyces cerevisiae*. *Metab. Eng.* **34**, 88–96 (2016).
- Sellick, C. A. *et al.* Metabolite profiling of recombinant CHO cells: Designing tailored feeding regimes that enhance recombinant antibody production. *Biotechnol. Bioeng.* **108**, 3025–3031 (2011).
- Xia, M. *et al.* Enhanced FK506 production in *Streptomyces tsukubaensis* by rational feeding strategies based on comparative metabolic profiling analysis. *Biotechnol. Bioeng.* **110**, 2717–2730 (2013).
- Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and Enrichment-Map. *Nat. Protoc.* **14**, 482–517 (2019).
- García-Campos, M. A., Espinal-Enríquez, J. & Hernández-Lemus, E. Pathway analysis: State of the art. *Front. Physiol.* **6**, 383 (2015).
- Khatri, P., Sirota, M. & Ten Butte, A. J. Years of pathway analysis: Current approaches and outstanding challenges. *PLoS Comput. Biol.* **8**, e1002375 (2012).
- Qiu, W. *et al.* Single and joint toxic effects of four antibiotics on some metabolic pathways of zebrafish (*Danio rerio*) larvae. *Sci. Total Environ.* **716**, 137062 (2020).
- Qi, D. & Chen, K. Bioinformatics analysis of potential biomarkers and pathway identification for major depressive disorder. *Comput. Methods Med.* **2021**, 3036741 (2021).
- Yang, Q. J. *et al.* Serum and urine metabolomics study reveals a distinct diagnostic model for cancer cachexia. *J. Cachexia Sarcopenia Muscle* **9**, 71–85 (2018).
- Long, N. P. *et al.* Metabolomics-guided global pathway analysis reveals better insights into the metabolic alterations of breast cancer. *J. Pharm. Biomed. Anal.* **202**, 114134 (2021).
- Chen, H. H. *et al.* The metabolome profiling and pathway analysis in metabolic healthy and abnormal obesity. *Int. J. Obes.* **39**, 1241–1248 (2015).
- Wei, Z. *et al.* Metabolomics coupled with pathway analysis characterizes metabolic changes in response to BDE-3 induced reproductive toxicity in mice. *Sci. Rep.* **8**, 5423–5438 (2018).
- He, Z., Liu, Z. & Gong, L. Biomarker identification and pathway analysis of rheumatoid arthritis based on metabolomics in combination with ingenuity pathway analysis. *Proteomics* **21**, 2100037 (2021).

33. Morris, C. *et al.* Bigdata analytics identifies metabolic inhibitors and promoters for productivity improvement and optimization of monoclonal antibody (mAb) production process. *Bioresour. Bioprocess.* **7**, 31 (2020).
34. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 30 (2000).
35. Alden, N. *et al.* Using metabolomics to identify cell line-independent indicators of growth inhibition for Chinese hamster ovary cell-based bioprocesses. *Metabolites* **10**, 199 (2020).
36. Akesson, M., Hagander, P. & Axelsson, J. P. Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding. *Biotechnol. Bioeng.* **73**, 223–230 (2001).
37. De Mey, M., De Maeseneire, S., Soetaert, W. & Vandamme, E. Minimizing acetate formation in *E. coli* fermentations. *J. Ind. Microbiol. Biotechnol.* **34**, 689–700 (2007).
38. Wolfe, A. J. The acetate switch. *Microbiol. Mol. Biol. Rev.* **69**, 50 (2005).
39. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
40. Gloaguen, Y. *et al.* PiMP my metabolome: An integrated, web-based tool for LC-MS metabolomics data. *Bioinformatics* **33**, 4007–4009 (2017).
41. McLuskey, K. *et al.* Ranking metabolite sets by their activity levels. *Metabolites* **11**, 1–15 (2021).
42. Lin, H., Vadali, R. V., Bennett, G. N. & San, K. Y. Increasing the acetyl-CoA pool in the presence of overexpressed phosphoenolpyruvate carboxylase or pyruvate carboxylase enhances succinate production in *Escherichia coli*. *Biotechnol. Prog.* **20**, 1599–1604 (2004).
43. Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **78**, 103–112 (2005).
44. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
45. Lu, S., Eiteman, M. A. & Altman, E. Effect of CO₂ on succinate production in dual-phase *Escherichia coli* fermentations. *J. Biotechnol.* **143**, 213–223 (2009).
46. Zhu, X. *et al.* Metabolic evolution of two reducing equivalent-conserving pathways for high-yield succinate production in *Escherichia coli*. *Metab. Eng.* **24**, 87–96 (2014).
47. Tan, Z., Chen, J. & Zhang, X. Systematic engineering of pentose phosphate pathway improves *Escherichia coli* succinate production. *Biotechnol. Biofuels* **9**, 262–274 (2016).
48. Cortada-Garcia, J. *et al.* On-line untargeted metabolomics monitoring of an *Escherichia coli* succinate fermentation process. *Biotechnol. Bioeng.* <https://doi.org/10.1002/BIT.28173> (2022).
49. Chatterjee, R., Millard, C. S., Champion, K., Clark, D. P. & Donnelly, M. I. Mutation of the ptsG gene results in increased production of succinate in fermentation of glucose by *Escherichia coli*. *Appl. Environ. Microbiol.* **67**, 148–154 (2001).
50. Wu, H., Li, Z. M., Zhou, L. & Ye, Q. Improved succinic acid production in the anaerobic culture of an *Escherichia coli* pflB ldhA double mutant as a result of enhanced anaerobic activities in the preceding aerobic culture. *Appl. Environ. Microbiol.* **73**, 7837–7843 (2007).
51. Thakker, C., Martínez, I., San, K. Y. & Bennett, G. N. Succinate production in *Escherichia coli*. *Biotechnol. J.* **7**, 213–224 (2012).
52. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
53. Prince, J. T. & Marcotte, E. M. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Anal. Chem.* **78**, 6140–6152 (2006).
54. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Preprint at <https://ggplot2.tidyverse.org> (2016).
55. Warnes, G. *et al.* *ggplots: Various R Programming Tools for Plotting Data*. R package version 3.1.3 (2022).

Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) grant number BB/R505523/1 and the Industrial Biotechnology Innovation Centre (IBioIC), grant number 2016-149. We want to thank Ingenza Ltd. for kindly providing the process details and medium recipe, and particularly to Scott Baxter for providing the succinate producer BW25113 *E. coli* strain and Fraser Brown, Iain McKean and Philip Weyrauch for running the off-line HPLC-UV/Vis-RI analysis of the fermentation samples. We also want to thank Glasgow Polyomics, especially Erin Manson, Hannah Rankin, Suzanne McGill and Stefan Weidt for running the LC-MS samples on the Q Exactive™ Orbitrap mass spectrometer.

Author contributions

K.B. and J.C.-G. conceived the idea. K.B. obtained the funding. J.C.-G. did the experimental work and analysed the data. J.C.-G. wrote the manuscript and drew the figures. S.A.A., K.B. and R.D. provided critical revision of the manuscript and made substantial contributions to its writing. S.A.A., K.B. and R.D. supervised the project and provided intellectual contributions.

Funding

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) Grant number: BB/R505523/1 and Industrial Biotechnology Innovation Centre grant number IBioIC grant number 2016-149.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-39661-x>.

Correspondence and requests for materials should be addressed to K.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023