

Understanding synthetic biology using the Q Exactive GC Orbitrap GC-MS/MS system and high-resolution, accurate-mass metabolomics library for untargeted metabolomics

Authors

Cristian Cojocariu,¹ Maria Vinaxia,² Mark Dunstan,² Adrian J. Jervis,² Paul Silcock,¹ and Nicholas J W Rattray²

¹Thermo Fisher Scientific, Runcorn, UK

²SYNBIOCHEM – Manchester Synthetic Biology Research Centre for Fine and Specialty Chemicals, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK

Keywords

Metabolomics, synthetic biology, Q Exactive GC, RBS-promotor, multivariate statistics, HRAM

Goal

To demonstrate the application of untargeted metabolomics to determine the effects of inducer type and concentration on the metabolic fingerprint of engineered bacteria under different growth conditions.

Introduction

Meeting the demand for specialty chemical compounds for the pharmaceutical, agricultural, and manufacturing industries is one of the grand challenges of the modern chemical industry. This demand must be met under increasing regulatory scrutiny using environmentally friendly methodologies. Biotechnological approaches, powered by the techniques and concepts of synthetic biology, have the potential to deliver the necessary sustainable solutions. At its core, synthetic biology applies a design-build-test framework¹ to the redesigning of natural biological systems for beneficial purposes. By inserting and fine-tuning genetic information within microbial bio-factories (such as *Escherichia coli*² and *Streptomyces* spp.³), it is possible to assemble complex enzymatic pathways for rapid and diverse chemical production. Inducible bacterial promoters, such as the isopropyl β -D-1-thiogalactopyranoside (IPTG)-inducible lac promoters or tetracycline-inducible *tet*-promoters, are commonly used to strongly activate gene transcription,

switching on engineered biosynthetic pathways. Yet, it is often not fully understood how these inducible systems interact with the global bacterial metabolome, potentially with toxic side effects.

The aim of this study was to investigate how the application of untargeted metabolomics can be used to understand the metabolic effects of inducer type and concentration on the metabolic fingerprint of engineered bacteria (*E. coli* DH5 α harboring an IPTG-inducible, red fluorescent protein expression plasmid pBbA1a-RFP) under different growth conditions. This phenotypic information has the potential to inform upstream genetic strategies while at the same time better defining the most efficient use of this promoter for biochemical pathway expression.

To track the metabolic shifts caused by promoter induction in the bacteria, a large number of metabolites from the cellular exo-metabolome (spent media) were detected and identified using the Thermo Scientific™ Q Exactive™ GC Orbitrap™ GC-MS/MS system. This innovative setup combines the advantages of high sensitivity (down to ppt levels), large dynamic range (to cover a wide range of metabolite concentrations), and high resolution (60,000 FWHM @ m/z 200) with a < 2 ppm mass accuracy as commonly needed to detect and identify individual metabolites within complex mixtures.

The untargeted workflow used in this study involved data acquisition of randomized biological samples and quality controls. Injections of quality controls, consisting of pooled samples, were performed after every five samples to ensure the analytical process is performed appropriately and to assess the system stability. This was followed by data processing for peak deconvolution, library search for putative compound ID, and analysis of variance for the selection of compounds with significant effects between the sample groups. Compound identification was made using both NIST2017 nominal mass library and the Thermo Scientific™ Orbitrap™ GC-MS high-resolution, accurate-mass (HRAM) metabolomics library (the first commercially available

high-resolution, accurate-mass metabolomics library for electron ionization GC-MS). The HRAM library contains more than 900 retention-indexed unique entries from more than 800 metabolites, covering both primary and secondary metabolites. Within this GC-MS analysis, the samples analyzed were spiked with a mixture of linear alkanes (C7–C30), which were used as internal retention markers.

Experimental

Growth conditions

Escherichia coli DH5 α from glycerol stocks were inoculated onto Lysogeny Broth (LB) agar plates followed by inoculation into either Terrific Broth (TB) or LB with 0.4% glucose and incubated overnight at 37 °C with shaking. Cultures were inoculated (1/100) into 1 mL of fresh broth in a 24-well plate and grown to mid-logarithmic phase, whereupon they were induced using a Hamilton® Multistar robotic system with variable levels of IPTG (25, 50, and 100 μ M final concentration) and incubated for a further 24 h at 37 °C with shaking.

Sample preparation and derivatization

Following incubation, samples were quenched with 1 mL of cold methanol (–48 °C) to halt any enzymatic action within the bacteria and centrifuged for 15 min at 12,225 RCF to remove cellular debris from the media. Then, 100 μ L of supernatant was filtered using a 0.45 μ m syringe filter, combined with 100 μ L of a 100 μ g/mL internal standard solution of D-glucose and L-alanine- d_3 , and dried down under vacuum. Lyophilized pellets were then subjected to a common two-step sample derivatization method carried out by the initial addition of 50 μ L of a 20 mg/mL methoxyamine/pyridine solution to enable the methoximation of any potentially labile ketone groups. Incubation at 65 °C for 40 min was followed by silylation in which 50 μ L of MSTFA + 1% TMCS (N-methyl-N-(trimethylsilyl)trifluoroacetamide + 1% trimethylchlorosilane) was added. Subsequent heating at 65 °C for 40 min afforded volatility to any labile hydroxyl and amine groups and the addition of the TMS (trimethylsilyl) moiety. The TMCS acted as a catalyst to ensure optimal TMS addition.

GC-MS analysis

In all experiments, a Q Exactive GC-MS/MS Orbitrap mass spectrometer was used. Sample injection into a hot split/splitless injector (280 °C) was performed using a Thermo Scientific™ TriPlus™ RSH™ autosampler, and chromatographic separation was obtained with a Thermo Scientific™ TRACE™ 1310 GC system and a Thermo Scientific™ TraceGOLD™ TG-5SilMS 30 m × 0.25 mm I.D. × 0.25 µm film capillary column (P/N 26096-1425). A total GC run time of 33 min per sample was used. Additional details of instrument parameters are shown in Table 1 and Table 2.

Table 1. GC and injector conditions.

| TRACE 1310 GC System Parameters | |
|---------------------------------|-----------------------------|
| Injection Volume (µL): | 1.0 |
| Liner: | Single taper (P/N 453A1345) |
| Inlet (°C): | 280 |
| Inlet Module and Mode: | SSL/SL, split 40:1 |
| Carrier Gas (mL/min): | He, 1.2 |
| Oven Temperature Program | |
| Temperature 1 (°C): | 70 |
| Hold Time (min): | 2 |
| Temperature 2 (°C): | 325 |
| Rate (°C/min): | 10 |
| Hold Time (min): | 6 |

Table 2. Mass spectrometer parameters.

| Q Exactive GC Mass Spectrometer Parameters | |
|--|---------------------|
| Transfer Line (°C): | 280 |
| Ionization Type: | EI |
| Ion Source (°C): | 250 |
| Electron Energy (eV): | 70 |
| Acquisition Mode: | Full-scan |
| Mass Range (Da): | 50–550 |
| Mass Resolution (FWHM): | 60,000 at m/z 200 |
| Lockmass (m/z): | 207.03235 |

Data processing workflow for unknown metabolite detection and identification

Full-scan, lock mass corrected data were imported into Thermo Scientific™ Compound Discoverer™ software and subjected to a qualitative untargeted workflow that involved retention time alignment, normalization, and statistical analysis (principal component analysis and differential analysis).⁴ Compound identification was achieved using Thermo Scientific™ TraceFinder™ software following spectral deconvolution and using the Thermo Scientific Orbitrap GC-MS HRAM metabolomics library. In addition to this, the NIST2017 nominal mass library was used to further extend the number of annotations assigned to putatively detected metabolites.

Results and discussion

Relative metabolite levels were determined from *E. coli* culture media (LB and TB) following incubation in the presence and absence of IPTG at various concentration levels. In addition to these samples, pooled quality control (QC) samples were also analyzed. Raw data files were imported into Compound Discoverer software and grouped according to the treatment (IPTG) and media (LB and TB) (Figure 1). Data processing in Compound Discoverer software involved an alignment step used to compensate for small differences in the retention times of the components in the sequence (Figure 1).

The component extraction (unknown detection using a ± 5 ppm extraction window and a signal intensity threshold of 500,000 peak area counts) step was followed by data normalization to correct for potential batch effects. To identify class differences, data was subjected to principal component analysis (PCA) (Figure 2). In this case, the first two principal components explain 54% of the variance within the dataset, with PC1 (30%) dominated by differences between the two media types, and PC2 (24%) by differences between induced and uninduced cultures. By using such an approach, comparison of PCA loadings against blank media allows the identification of bacterial metabolites that differ between sample classes.

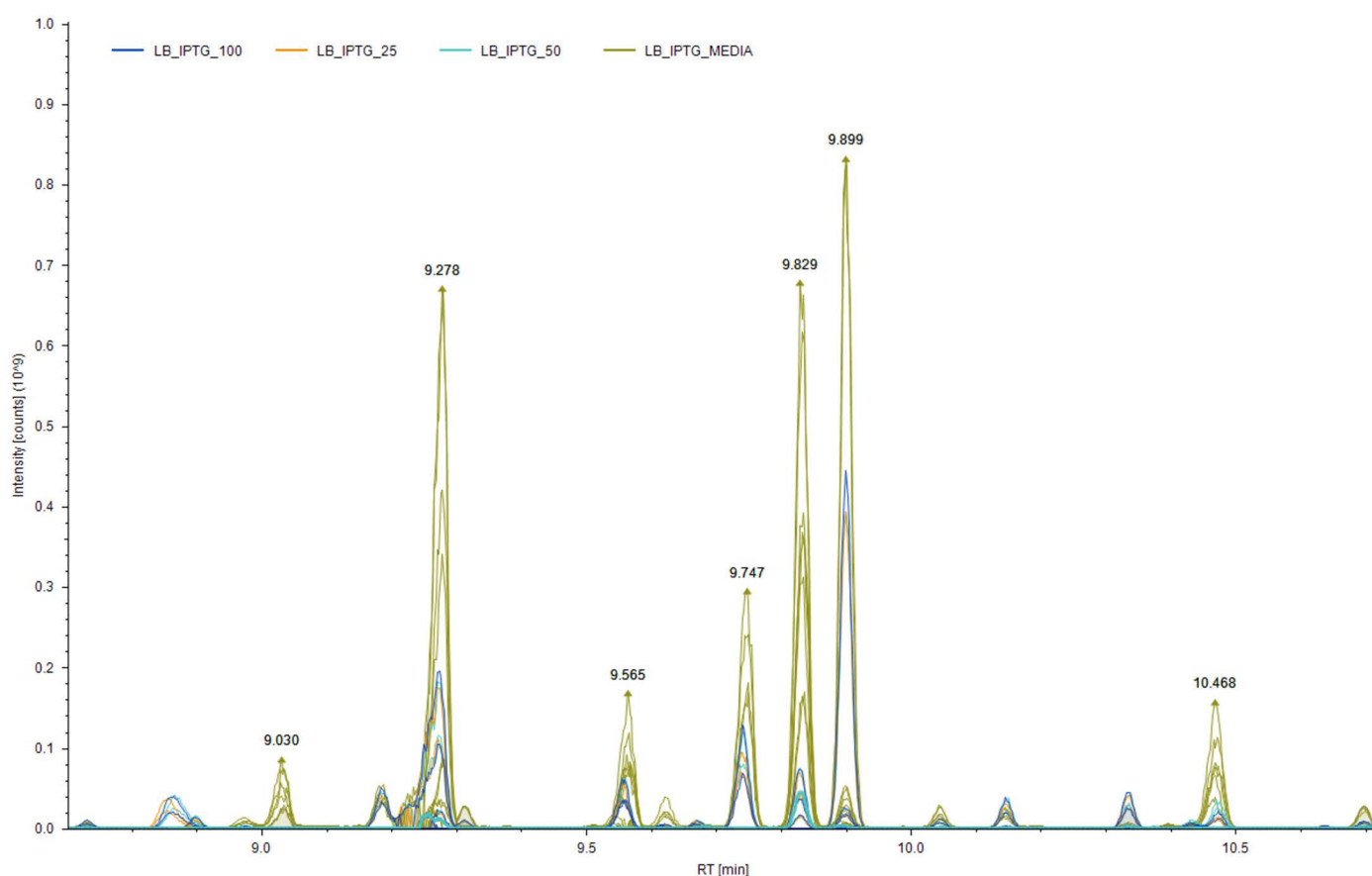


Figure 1. Example of retention time alignment in Compound Discoverer software for several peaks detected in *E. coli* DH5 α cultures induced with IPTG and grown in LB media.

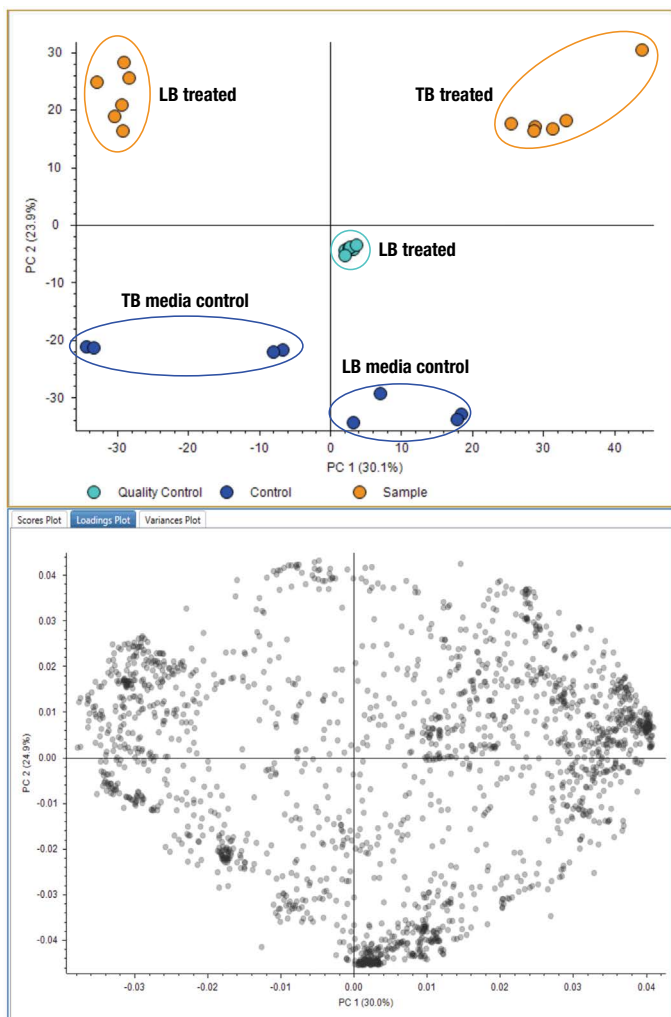


Figure 2. Centered and \log_2 -scaled Principal Component Analysis (PCA) scores plot (top) and loading plot (bottom). Data points within specific ellipses represent *E. coli* DH5 α growing in either TB or LB media (Blue - TB media control, LB media control) and after the inclusion of various levels of the pBbA1a-RFP (IPTG) plasmid, under different media conditions (Orange - TB treated, LB treated). QCs (n=8, pooled samples) were also analyzed to test instrument and method performance.

As complete group separation was noted within the PCA, a wholly unsupervised approach was adopted. The next three data processing steps were designed to select significant features that contributed to group differences,

in this case LB control vs. LB IPTG treated. An analysis of variance test (ANOVA) was performed alongside a subsequent multiple comparison Tukey Honest Significance Difference (T-HSD) test. This supplied an adjusted p -value of compound significance that was subsequently used as input, alongside associated compound fold change, into the volcano plot tool available in Compound Discoverer software. This tool plotted \log_2 fold change vs. $-\log_{10}$ p -value and identified compounds that were important in group discrimination and also had a suitable large fold change (Figure 3). Significant compounds (2045 ions corresponding to 212 compounds selected based on p -values < 0.05 and \log_2 fold change values > 1) were then selected and sent to TraceFinder software for attempted identification using spectral matches against libraries/databases.



Figure 3. Discriminatory analysis (Volcano-plot) generated for LB IPTG 100 (green) and LB control (red) samples showing the compounds that significantly contributed to group difference to the left and right sides. The x-axis represents the \log_2 of the fold change between the two sample groups, and the y-axis represents the $-\log_{10}$ of the adjusted ANOVA p -value. The top-ranking ions in each group are highlighted in blue.

Compound identification using Orbitrap-GC HRAM metabolomics library

The overall goal of untargeted GC-MS metabolomics studies is to detect and annotate (identify) the metabolites responsible for group differences. This is usually accomplished by comparing the measured spectra against in-house standard databases or unit mass spectral libraries such as NIST or Wiley. Such annotations against an in-house library ascribe to the proposed minimum reporting standards of the community-led Metabolomic Standards Initiative at the highest level.⁴ Statistically significant features were sent to TraceFinder

software and identified using both NIST2017 and the Thermo Scientific Orbitrap GC-MS HRAM metabolomics library and retention time index derived from a C10–C19 alkane mix. This HRAM metabolomics library was created using pure metabolite standards analyzed on the Orbitrap-GC, and it contains ~850 unique metabolite spectra (each with retention time index, CAS numbers, and PubChem identifiers) acquired in EI using 70 eV and 60,000 resolution. An example of spectral matching is shown in Figure 4 for glycine trimethylsilyl ester (glycine 3TMS).

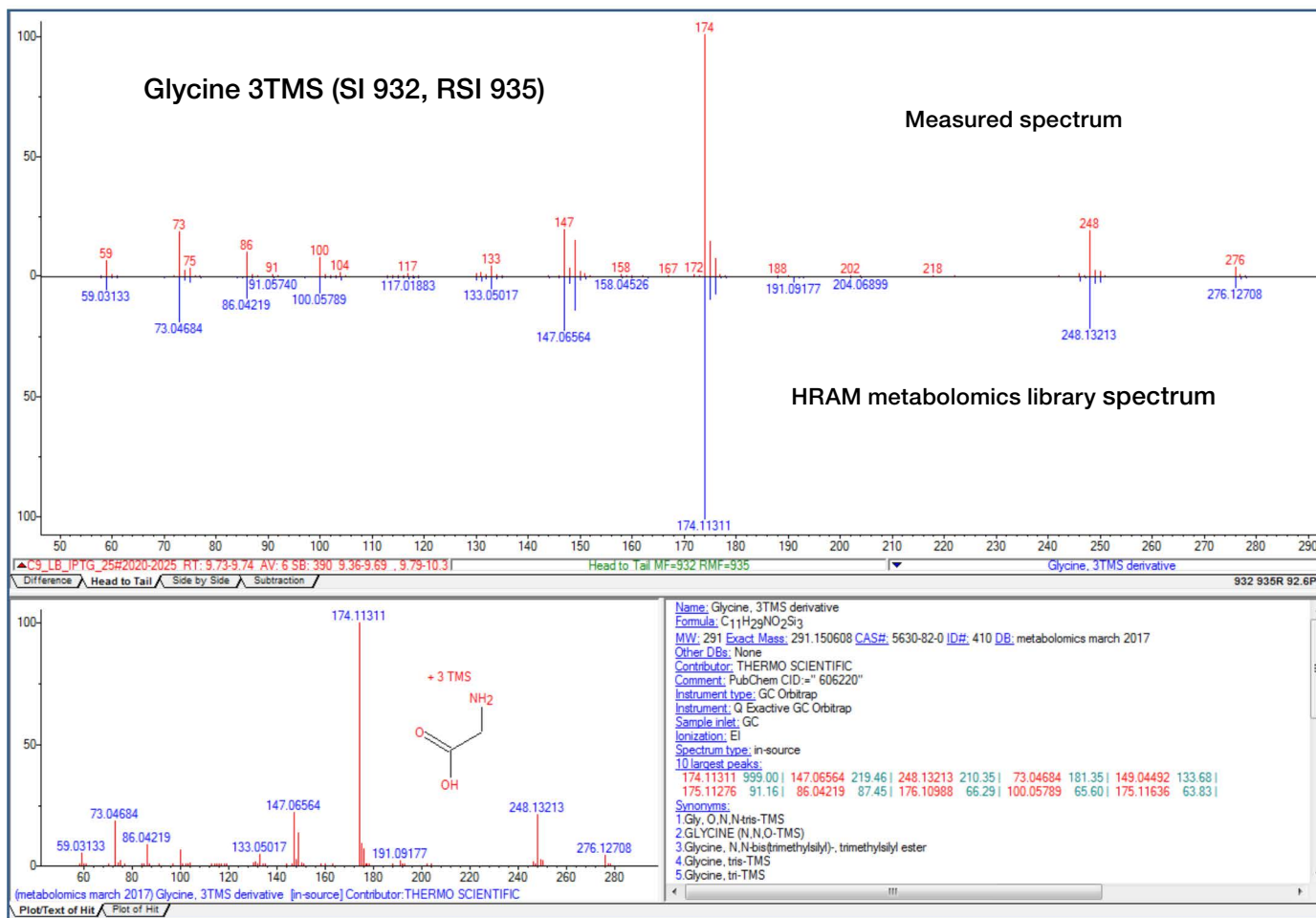


Figure 4. Glycine 3TMS identification using the Thermo Scientific Orbitrap GC-MS HRAM metabolomics library. Forward and reverse search indices in addition to accurate mass information add to the confidence in compound identification.

Compound annotation was achieved using a search index dot product value of >750, a total score⁵ of >80, and a maximum retention time index difference (Δ RI) of 100 (measured versus expected). An example of the TraceFinder software deconvolution browser showing GABA 3TMS identified based on the criteria stated above is shown in Figure 5.

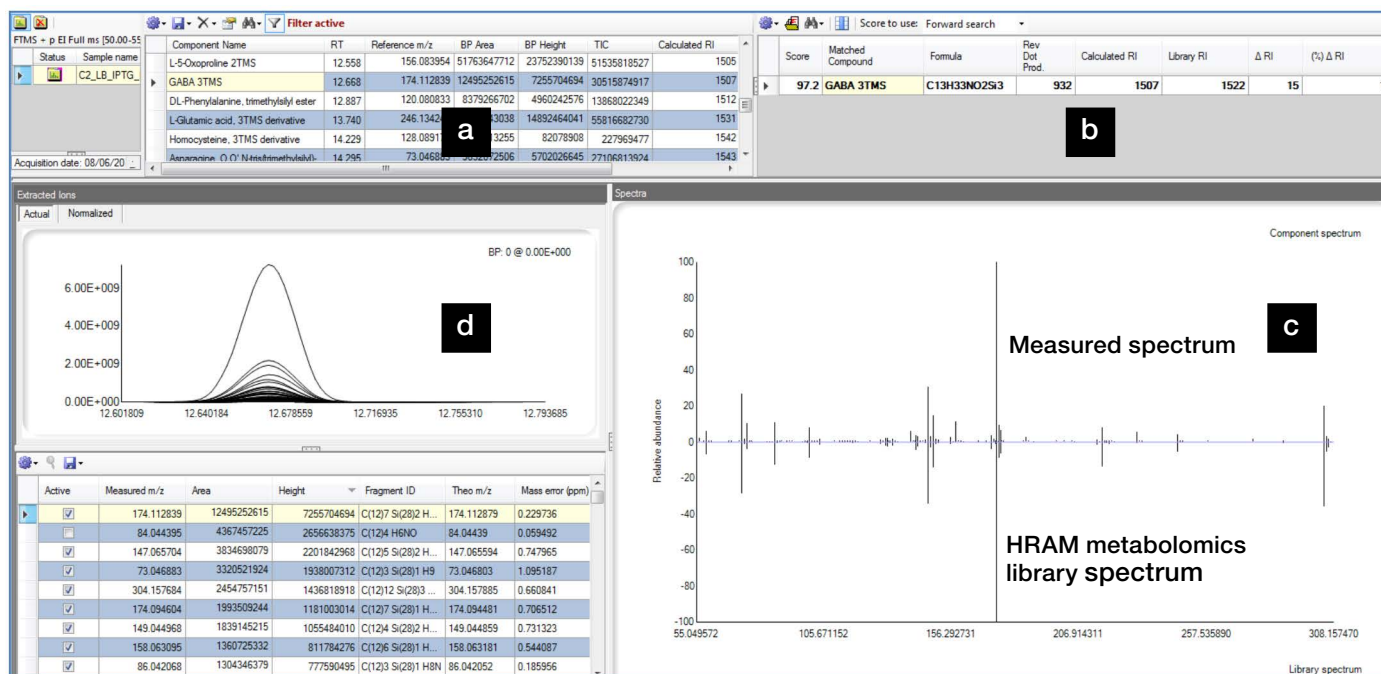


Figure 5. Example of metabolite identification in the TraceFinder software deconvolution browser showing a list of compounds (a), identified based on a total (average) score and retention index information (b), across the retention time aligned media samples (c), spectral match against the HRAM Orbitrap metabolomics library (c) as well as the deconvoluted spectrum (d) for GABA 3TMS are shown.

Following this process, 39 significant metabolites were confidently identified from the ANOVA and volcano plot analysis, and their corresponding peak area fold-changes in each measured sample were calculated (Table 3). Fold-changes of each metabolite identified at each IPTG concentration were all within one-fold unit of change of each other. This indicates that minimal changes in the same metabolite were observed upon use of higher amounts of IPTG. Therefore, at a metabolite level at least, conservative levels of IPTG should be encouraged to reduce cost

and environmental impact. Optimization of IPTG levels in experimental design are a logical and important step, and this approach has already been proven successful in the production of terpenes² and proteins.^{6,7} Yet, this is not a fully accepted practice as researchers regularly use an excess amount of IPTG to save time. But with IPTG currently costing upward of \$400 per 100 g, developing a precise approach to inducer usage becomes even more important when methods are transferred to the industrial scale.

Table 3. Table of fold change of detected metabolites that significantly contributed to group differences in the LB IPTG 25, LB IPTG 50, and LB IPTG 100 groups. Green to red color gradients indicate the fold change of associated metabolite upon comparison to blank media. All comparisons are made between LB media control and each of the IPTG-treated samples (e.g., putrescine is upregulated 5-fold in LB IPTG 25 sample as compared to LB media control). The color and intensity of the boxes is used to represent changes of fold change. In the example below, red represents up-regulated metabolites, and green represents down-regulated metabolites.

| RT | Putative ID | m/z Base Peak | Average Score | LB IPTG 25 | LB IPTG 50 | LB IPTG 100 |
|------|---|---------------|---------------|------------|------------|-------------|
| 5.15 | propylene glycol 2TMS | 73.04691 | 97.3 | -2.3 | -2.0 | -2.3 |
| 6.06 | (R)-3-hydroxybutyric acid 2TMS | 147.0657 | 84.1 | 3.2 | 3.6 | 2.5 |
| 7.88 | D-isoleucine, N-acetyl TMS | 86.09647 | 84.6 | -0.8 | -0.7 | -0.6 |
| 9.22 | 2-[Bis(trimethylsilyl)amino]ethyl bis(trimethylsilyl) phosphate | 299.0715 | 98.9 | 1.5 | -1.4 | -1.7 |
| 9.25 | Leucine 2TMS | 158.1358 | 98.9 | 3.5 | 3.1 | 3.4 |
| 9.55 | L-allo-isoleucine 2TMS | 158.1358 | 96.5 | 6.9 | 7.0 | 7.1 |
| 9.62 | Proline 2TMS | 142.10467 | 98.7 | -2.1 | -1.9 | -1.7 |
| 9.75 | glycine 3TMS | 174.11307 | 99.5 | -0.7 | 0.3 | 0.3 |
| 9.82 | succinic acid 2TMS | 147.0657 | 99.4 | -1.7 | -1.8 | -2.3 |
| 10 | glyceric acid 3TMS | 147.0657 | 99.1 | -0.4 | -0.3 | -0.3 |
| 10.1 | uracil 2TMS | 241.08238 | 95.5 | 1.4 | 1.6 | 1.6 |
| 10.3 | fumaric acid 2TMS | 245.06578 | 96 | 5.2 | 5.3 | 5.2 |
| 10.5 | serine 3TMS | 204.12357 | 98.4 | -6.2 | -6.0 | -6.4 |
| 11.2 | L-methionine 3TMS | 56.0496 | 94.6 | -1.3 | -1.0 | -1.2 |
| 11.8 | cadaverine 4TMS | 174.11308 | 92.6 | -2.8 | -3.1 | -3.1 |
| 11.9 | malonic acid 2TMS | 147.06569 | 94.6 | -1.1 | -0.7 | -0.8 |
| 12.5 | putrescine 4TMS | 174.11307 | 81 | 5.0 | 5.2 | 5.0 |
| 12.6 | pyroglutamic acid 2TMS | 156.08385 | 98.6 | -2.1 | -1.3 | -1.8 |
| 12.7 | GABA 3TMS | 174.11292 | 93.4 | -4.7 | -3.8 | -3.3 |
| 12.9 | Phenylalanine TMS | 120.08079 | 91.5 | -0.8 | -0.6 | -0.6 |
| 13.7 | glutamic acid 3TMS | 246.13425 | 98.7 | -1.6 | -1.2 | -1.0 |
| 16 | L-arginine 3TMS | 157.11508 | 96.7 | -4.1 | -3.5 | -3.7 |
| 16.5 | sorbose methoxyamine isomer 1 | 217.1075 | 97.6 | -2.4 | -2.3 | -2.2 |
| 16.6 | sorbose methoxyamine isomer 2 | 217.1075 | 96.5 | -1.4 | -1.1 | -1.2 |
| 16.8 | d-Mannose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)- | 147.06561 | 96.9 | -2.0 | -1.8 | -1.7 |
| 17 | glucose 5TMS isomer 2 | 73.04685 | 97.5 | -2.8 | -2.5 | -1.4 |
| 17 | L-lysine 4TMS | 156.12053 | 98.4 | -1.4 | -1.2 | -1.1 |
| 17.2 | tyrosine 3TMS | 218.1029 | 98.7 | -4.3 | -3.8 | -3.9 |
| 17.3 | ribose 4TMS | 217.10722 | 84.8 | -1.4 | -1.2 | -1.4 |
| 17.6 | mannose 5TMS isomer 1 | 204.09943 | 99 | -2.3 | -2.0 | -1.9 |
| 18.6 | myoinositol 6 TMS | 217.10748 | 99 | -0.5 | -0.2 | -0.3 |
| 18.9 | guanine 3TMS | 352.1441 | 95 | 1.8 | 2.0 | 1.8 |
| 19.8 | L-tryptophan 3TMS | 202.10474 | 97.3 | -1.7 | -2.0 | -1.9 |
| 20.1 | sorbose isomer 1 | 73.04684 | 90.6 | -3.9 | -3.8 | -4.2 |
| 21.7 | uridine 2TMS | 217.10728 | 84 | -3.7 | -3.5 | -4.2 |
| 23.2 | adenosine 4TMS | 230.11531 | 92.8 | -4.8 | -5.3 | -5.8 |
| 24 | trehalose 8TMS | 361.16836 | 97.4 | -9.0 | -8.8 | -8.2 |
| 24.1 | guanoside 3TMS | 324.13055 | 91.2 | -1.8 | -2.1 | -2.1 |
| 24.5 | melibiose 3TMS isomer 1 | 204.09943 | 92.8 | 3.9 | 2.8 | 3.3 |

Assuming that any fold-changes by a factor larger than 2 represent real differential abundance, rather than noise within the system, the results indicate that increased levels of IPTG promoted a depletion of sugars (trehalose, sorbose, glucose, and mannose) and decreased the levels of several amino acids (serine, arginine, proline, and tyrosine). These changes potentially reflect plasmid induction and activation by IPTG and, once switched on, production of the red fluorescent protein. Sugar metabolism is needed to provide the energy for protein biosynthesis, while the reduction in amino acids highlights the demand for raw material. Interestingly, GABA (γ -amino

butyric acid) showed a marked four-fold reduction that could indicate a drive by the bacteria to use this as an alternative energy source. GABA can enter the TCA cycle by being converted to succinic acid via the GABA shunt pathway. Counter-intuitively, succinic acid itself also displayed a two-fold reduction upon plasmid induction, but fumaric acid, the metabolite following on from succinic acid within the TCA cycle, displayed a 5-fold increase. Reduction of succinic acid post-IPTG induction has also been observed within the W3110 *E. coli* strain, in which similar fundamental nutritional shifts were observed during protein production.⁶

Such bottlenecks within the energy cycle could be caused by a lack of capacity of the fumarate hydratase enzyme that converts fumaric acid to malic acid. Identifying such potential rate-limiting steps is vital input for subsequent design-build-test cycles to generate more efficient pathways. 3-hydroxybutyric acid also shows a high fold-change that points to issues with flow within the TCA cycle. In *E. coli*, this ketone body can be converted to acetoacetyl-CoA, which is eventually modified to acetyl-CoA, a vital metabolite in lipid metabolism energy production within glycolysis. Alongside this, increased levels of putrescine were observed in the IPTG-treated samples as compared to LB media control (Table 3). This observation of putrescine in spent media from an *E. coli* recombinant expression system mirrors results observed from the BL21(DE3) strain.⁷ Putrescine is a polyamine that is produced by the breakdown of amino acids from protein sources in living and dead cells. An up-regulation of the metabolite potentially indicates an increase in the need to derive amino acids for RFP production by degradation of other proteins. Compared to transporter activation, recycling of chemicals in this way is an energy-efficient route to gain essential raw materials.

Conclusions

The data obtained in this study clearly show significant changes in metabolism arising from IPTG induction and subsequent protein production on the metabolic fingerprints of *E. coli* cells. However, it is important to note that due to the proof-of-principle nature of this research, it is difficult to determine if induction or protein production and subsequent extracellular transportation caused the associated metabolic changes observed. A larger secondary experiment is required to fully validate and confirm the biological findings.

- The untargeted analysis leads to the identification of 39 significant metabolites that showed statistically significant differences between the exo-metabolome of *E. coli* grown in LB media and exposed to various levels of IPTG and the corresponding control.
- Metabolites identification was simplified by using a dedicated HRAM metabolomics library retention time index information, significantly increasing the confidence in compound identification, one of the most critical steps in metabolomics.
- Compound Discoverer software and TraceFinder software streamline data interrogation, proving qualitative and quantitative information and increase the confidence in the results.

Taken together, using the Q Exactive GC Orbitrap GC-MS/MS system operated in full-scan mode at high resolving power allows confident metabolite detection and identification, and eases the ability to discriminate between various biological sample groups.

References

1. Carbonell, P., et al., SYNBIOCHEM—a SynBio foundry for the biosynthesis and sustainable production of fine and speciality chemicals. *Biochemical Society Transactions*, **2016**, 44(3), 675–677.
2. Toogood, H.S., et al., Enzymatic Menthol Production: One-Pot Approach Using Engineered *Escherichia coli*. *ACS Synthetic Biology*, **2015**, 4(10), 1112–1123.
3. Muhamadali, H., et al., Metabolomics investigation of recombinant mTNF α production in *Streptomyces lividans*. *Microbial Cell Factories*, **2015**, 14(1), 157.
4. Sumner, L. W., et al., Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, **2007**, 3(3), 211–221.
5. *Application of GC Orbitrap mass spectrometry for untargeted metabolomics of pathogenic microorganisms*. Thermo Fisher Scientific Application Note 10532, **2016**.
6. Carneiro, S., et al., Metabolic footprint analysis of recombinant *Escherichia coli* strains during fed-batch fermentations. *Molecular Biosystems*, **2011**, 7(3), 899–910.
7. Muhamadali, H., et al., Metabolomic analysis of riboswitch containing *E. coli* recombinant expression system. *Molecular BioSystems*, **2016**, 12(2), 350–361.

Find out more at thermofisher.com/OrbitrapGCMS

ThermoFisher
SCIENTIFIC