From Discovery to Process Control: Positioning DIA Proteomics in Biomanufacturing Pipelines

Matthew R. Russell [1*], Philip J. Brownridge [2], Joeseph Windo [1], Nigel S. Scrutton [1], Claire E. Eyers [2], Perdita Barran [3*]

[1]Future Biomanufacturing Research Hub, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Rd, Manchester, M1 7DN, UK.

[2]Centre for Proteome Research, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool, L69 7ZB, UK.

[3]Michael Barber Centre, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Rd, Manchester, M1 7DN, UK.

*Corresponding author(s). E-mail(s): matthew.r.russell@manchester.ac.uk; perdita.barran@manchester.ac.uk;

Abstract

Gaining control of existing biomanufacturing chassis organisms, such as *Escherichia coli K12*, and novel isolates, such as the salt tolerant *Halomonas bluephagenesis sp TD01* studied here may be facilitated by the investigation and monitoring of their metabolic and regulatry processes, particularly through proteomics. Here we consider the performance of a range of typically available proteomics platforms across a range of price points to map chasis organisms' metabolic pathways. A set of model bacterial samples was prepared from *E. coli* and *H. bluephagenesis sp. TD01* in 1:2 and 2:1 ratios and analysed using five LC-MS systems. Data from the timsTOF HT, Exploris 480, ZenoTOF 7600 and Select Series MRT were processed through DIANN and MSStats. Data from the Vion was processed through Skyline then MSstats. Of the 8,222 proteins identified across all samples analysed (4,401 proteins from *E. coli*; 3,821 from *Halomonas sp. TD01*), the TimsTOF and Exploris were able to achieve extensive proteome coverage quantifying 5.5k and 5k proteins respectively, with the ZenoTOF, Waters MRT and the legacy Waters Vion respectively quantifying 3.5k, 1.3k, and ~850 proteins at 1% FDR. Proteins comprising the pathways of chassis' organisms core metabolism critical to biomanufacturing were quantified with all instruments, demonstrating suitability of these platforms to explore their manipulation in the context of biomanufacturing.

Keywords: Mass Spectrometry, Proteomics, Quantitative Proteomics, data-independent acquisition, DIA, SWATH-MS, DIA-NN / DIANN, MSstats, LC-MS, Biomanufacturing, Engineering Biology, Halomonas Bluephagenesis TD01, Escherichia coli K12

Main

Proteomics is a valuable but underused tool in engineering biology and biomanufacturing. It can provide insight into established bacterial strains such as *E. coli*, and aid adoption of novel chassis organisms like *H. bluephagenesis*. Here we use both established and novel strain to demonstrate the utility of proteomics to biomanufacturing projects, by providing quantitative data on the proteins present in the host chassis that could be related to product titre in an engineered strain.

Biomanufacturing puts living organisms to work producing chemical products at industrial scale. It is an economically important industrial sector, projected to reach $3.9 trillion in revenue by 2030[1]. Such industrial production - deriving carbon from the biosphere, rather than from hydrocarbon feed stocks - is a critical technology enabling transition to a "net zero" economy.[2]. Proteomics, which provides a systems level overview of the protein machinery of biological systems, is a key technology to enhance understanding and design of biomanufacturing chassis organisms. However application of proteomics to biomanufacturing remains under explored compared to medicine[3]; or indeed to microbiology in general[4–6]; but has shown utility in biopharmaceutical production[7], including by us[8].

The limited adoption of proteomics by the biomanufacturing industry can be attributed to startup costs, and the risk that projects fail to positively impact the bottom line. Costs include capital investment in proteomics facilities and equipment, and staffing cost for skilled operators. Legitimate doubts about the payoff remain given the absence of cases studies complete with reference data, methodology and workflows that demonstrate the benefits of application of this technology to increase product yield and/or lower production costs[9].

Our assessment of proteomics instruments' performance for the analysis of biomanufacturing relevant samples alongside economic considerations demonstrates the applicability of the approach to solve problems in the sector. We envision a future in which early development of recombinant expression systems, or adoption of novel chassis, is supported by discovery protein-level data generated in collaboration with universities or contract research organizations that can provide access to the necessary expertise and high-end facilities. The data and know-how generated would then facilitate adoption of lower cost proteomics platforms, permitting the application of targeted approaches, at production sites where it would add value by supporting quality control and enabling problem solving[8]. The data then acquired would provide a reference dataset for others wishing to use these organisms in a biomanufacturing setting.

To demonstrate the approach, we first prepared mixtures of peptides from *E. coli* and *H. bluephagenesis* in ratios of 1:1, 1:2 and 2:1 in a similar approach to that adopted in multi-site comparison studies[10]. Aliquots of each were used to acquire liquid chromatography mass spectrometry data using data-independent acquisition (DIA) on a range of instruments across a scale of price tags and capabilities. The numbers of proteins identified, quantified, and the data's sensitivity to differential expression are reported of each instrument. Our intention was not to identify a "best" instrument, but rather to show the value that could be obtained irrespective of platform, as relevant to advance biomanufacturing, and to provide model datasets and robust methodology for such applications. As a worked example, we showcase how the Sciex ZenoTOF, ranking middle in terms of both cost and performance, is able to support bioengineering development in *H. bluephagenesis* for biofuel production. Specifically, we were able to demonstrate that replacement of glucose with a carboxylic acid feed stock - a low cost biproduct of biodiesel production – directs carbon flux towards potential butane precursors, characterising the metabolic adaptation of *H. bluephagenesis*.

Results

Instrument Performance

A model sample set comprising ratio mixtures of 1:2, 1:1 and 2:1 by total peptide mass, of *E. coli* and *Halomonas bluephagenesis sp TD01* whole cell lysates wass analysed on current high end mass spectrometry instrumentation: Bruker timsTOF HT; Thermo Exploris 480; and Sciex ZenoTOF; a modern bench top instrument focused on low cost routine measurement: Xevo MRT; and a legacy instrument developed for MPK and CRO work now available second hand: Vion.

The headline protein identification numbers for each of the conditions are tabulated Table 1 alongside approximate instrument list prices at time of publication. Bacteria, of the kind used for biomanufacturing, typically have genomes coding for between 500 and 5,000 proteins. The fasta protein sequences used here contain 4,401 and 3,821 proteins for *E. coli* and *H. bluephagenesis* respectively for a total of 8,222 proteins potentially detectable in the combined samples. Peptides from all of the 8,222 proteins were included in the spectral library generated by DIANN. The full complement of proteins coded for by a bacterial genome will not be expressed; for example proteins required only for spore formation are not likely to be expressed in fully active cells in liquid media. With this in mind, our quantification of 4,726 and 5,554 proteins respectively using either the Exploris 480 or the timsTOF HT instruments likely represent near complete coverage of the expressed genome of these two species under the sampling conditions. Fewer proteins (3,402) were quantified using the ZenoTOF, albeit with a shorter LC separation time that would permit approximately twice the sample throughput of the other workflows. While it is possible that higher coverage could have been achieved with this instrument using a longer LC gradient, we elected to use the manufacturer recommended methods than re-optimise conditions across the 5 platforms. The remaining two instruments permitted quantification of 969 and 862 proteins respectively for the Xevo MRT and the Vion respectively, equating to 17% and 15% of the numbers identified using the timsTOF. However, based on the proteins identified, these data are sufficient to provide valuable troubleshooting or process monitoring information in an industrial production application.

Reproducibility of protein and peptide level quantification for each of the instruments across replicate injections is presented in Figure 1. At the protein level Figure 1 A, replication was more consistent across the timsTOF HT, Exploris 480 and ZenoTOF platforms, with 75-85% of proteins identified in at least one of the 5 replicate injections. These numbers decreased to ~52% for the Vion and the MRT. As expected, Vion reported the fewest quantified proteins. Nevertheless, even this coverage can provide actionable insight for troubleshooting and process monitoring For comparison our previously published survey of *H. bluephagenesis* expressed proteins and assissiated spectral library covered 1160 proteins from a fractionated sample analysed on a Waters XevoTOF[11], where the top three instruments here, the ZenoTOF, timsTOF and Exploris quantified 1758, 2928 and 2529 halomonas proteins respectively. So substantialy more of the proteome is covered here, than previosly.

Quantitative reproducibility at the peptide level Figure 1 B broadly aligned with the protein level data, with data from the more sensitive instruments (timsTOF HT and Exploris 480) indicating a higher number of peptides per protein being used for quantification. Confidence in quantified protein identification is reported in Figure 1 C. The rank order of instrument performance is consistent with the the first two plots. Protein identifications in all cases come close to saturation before exceeding protein discovery q-value of 0.01.

Volcano plots (Figure 1 D-H) and MA plots (Figure 1 I-M), allow relative comparison of the instrument overall performance in terms of sensitivity to differential expression and quantitative accuracy. Although the total numbers of proteins quantified differs (as previously discussed), the quantitative accuracy of the four TOF-based instruments (timsTOF HT, ZenoTOF, MRT and Vion) and the Orbitrap-based Exploris 480 were similar, yielding ratios of with median 1.76 +/- median absolute deviation of 0.3 compared with the ratio of dilution of 2. Interestingly, specified protein ratios wer quantitatively detected with higher statistical confidence for data generated with the Exploris 480, although this is unlikely to make a substantial difference to the insights gained for the purpose of biomanufacturing. MRT ratios were compressed relative to expected values, likely due to background noise. We are grateful to Waters for early access to this platform and understand that further optimisation is planned. Our results indicate that this bench-top system has potential for biotechnology proteomics with additional method development. Our findings show that this versatile bench top platform also has potential as an instrument for proteomics for biotechnology applications.

Direct comparison of the absolute abundance of the *E. coli* proteins from each instrument against each other, and the PAX database of absolute protein abundance (https://pax-db.org/species/511145)[12] is reported in Figure 2. The highest correlation was observed between the Exploris 480 and the ZenoTOF (0.757), with reduced consistency apparent between the Exploris 480 and timsTOF HT (0.710). Consistency of protein abundance was below 0.65 for all other instrument comparisons. More importantly perhaps, particularly in the context of systems modelling, absolute abundance correlated relatively well (>0.65) between the timsTOF HT, Exploris480 and ZenoTOF with data from PAX DB. It is clear from this comparison that quantification of low abundance proteins is compromised with both the Vion and MRT instruments, explaining the reduction on the overall numbers of proteins quantified. Additionally, the three higher coverage instruments (timsTOF, Exploris and ZenoTOF) overestimate the higher abundance proteins with respect to the PAX database values.

All of the raw data, processing scripts and processed data is available in the Pride repository as a community resource.

Biological Insight

An important part of understanding the behaviour of biomanufacturing chassis organisms is monitoring of key metabolic pathways. The percent protein coverage of key KEGG pathways in *E. coli* achieved with each instrument is reported in Figure 3 for pathways with 50 proteins or more associated with them. As expected from the depth of proteome coverage, the timsTOF HT and Exploris 480 consistently deliver the highest coverage of these pathways, this reducing for the ZenoTOF, MRT and the Vion. Coverage of these metabolic pathways is higher with the ZenoTOF than the headline protein-level data would suggest, perhaps because these pathways are generally of higher abundance due to their critical biological functions. The most important pathway for understanding carbon flux which is critical to biomanufacturing is *Carbon metabolism*, for which 43% coverage was achieved with the Vion, compared with 91% for the timsTOF HT. Obtaining quantitative data for almost half of this pathway is still useful and an example of how a lower cost/sensitivity platform such as the Vion, can provide sufficient information of importance to a biomanufacturing facility. A corollary is the relatively low (<10%) coverage of the *ABC transporters*, for which over 80% is covered by the timsTOF HT. Such transporters could be critical in a given biomanufacturing chassis for their role in the export of toxic product molecules to media; understanding, and optimizing their contribution will necessitate deeper proteome profiling[13–15].

Exemplar Application

Finally, as an example of the power of proteomics to provide insight into the biology of novel chassis organisms for biomanufacturing, we compared data from *H. bluephagenesis* cultured in either carbon source of glucose which enters core metabolism via the glycolytic pathway, or acetate which enters core metablosim via conjugation with Coenzyme A Figure 4. Using the ZenoTOF, we quantified 2,818 proteins from the 3,821 entries in the *H. bluephagenesis* uniparc proteome (UP000838464). Of these, 120 proteins (4.3% of those quantified) were statistically significantly differentially regulated, with 58 being elevated and 42 being at lower levels (considering those at an adjusted p-value <=0.01, of fold-change >2). The proteins names from the uniparc *H. bluephagenesis* sequence databases were mapped to KEGG gene names and so to KEGG KO numbers which indicate gene function. In the volcano plot in Figure 4 we highlight the locations of differentialy expressed proteins with KO numbers linked to KEGG pathways hbp00010 (Glycolysis / Gluconeogenesis) and hbp00620 (Pyruvate metabolism), the latter of which includes metabolism of acetate including its conversion to acetyl-CoA and so the the direction of carbon flux into the citrate cycle. A protein annotated with K01895 (acetyl-CoA synthetase) is up regulated in acetate fed culture over glucose fed culture. This is to be expected as the bacteria must componsate for it's loss of glucose by increasing carbon flux from acetate directly into the citrate cycle. Another set of proteins with KEGG annotations K01571, K00845, K00134, K00873, K01810 are all down regulated in acetate fed cultures with respcet to glucose fed cultures, are part of the glycolysis pathway. Again it makes sense that *H. bluephagenesis* in an environment with reduce glycose should diminsh expresion of proteins required to make use of glucose. These exemplar proteins belong to pathways that we might expect to be modulated under the respective feedstock conditions, provides confidence in the general utility of our quantitative proteomics data for such interrogations. Further research into how and why the levels of these proteins are altered when *H. bluephagenesis* is grown in different carbon sources will provide insight to how any novel chassis system under active development for biomanufacturing applications could be conrolled and adapted. Understading how such organisms respond to changing conditions will guide selection of culture conditions. In the spesific case of feed stocks, which are a major cost in biomanufacturing, understanding and mopdifying organisms capability to grow on various feed stocks may help transtions a biomanufactuing platform to a lower carbon source. Understaning the promotors driving differential expression of these proteins under differring feed stock conditions guide selection of promotors for use in production pathways, and even enable initiation of production by switching feed stock We are currently exploring genetic interventions to boost, or suppress, expression of proteins of interest identified in this way could could increase yeild of products in model bioreactors. This demonstrates how even a simple proteomics experiments can produce a lot of useful, and actionable, information to guide development for biomanufacturing production.

Discussion

Microorganims based biomanufacturing faces key challenges[16]. Better control of established organisms is required to improve yields and cut costs, while accelerating adoption of lesser-known species with valuable traits. Our investigations model these scenarios using proteins from established *E. coli* alongside recently isolated *Halomonas bluephagenesis sp TD01*[17–25].

E. coli is the dominate biomanufacturing organism by historical accident[26;27] and even with a knowledge base from over century of research on the model organism implementing novel engineered biology pathways in *E. coli* remains difficult[16]. While some alternative species see limited use (*Bacillus subtilis*[28], *Streptomyces*[29], *Pseudomonas putida*[30], *Corynebacterium glutamicum*[31]), nature

likely contains species with more favourable traits: toxicity resistance, efficient carbon utilisation, and ability to thrive in economical bulk production conditions[32;33]. *H. bluephagenesis* demonstrates such advantages; found in a salt lake, it thrives in high-salt, high-pH environments hostile to most bacteria, enabling cost reducing non-sterile cultivation.[25] The data we provide here will allow us and others to better exploit its potential.

For biomanufacturing to replace petroleum-based industry, it must produce commodity chemicals at competitive costs, which can only be achieved when there is understanding of the pathways than can, and should be, optimised for maximal yield. Effective proteomics support for biomanufacturing requires methods and platforms spanning development through production. Building on prior DIA benchmarking studies[10] , we thus tested five different MS platforms for their utility in this context. By analysing tryptic digests from *E. coli K-12 MG1655* and *H. bluephagenesis* in 1:2, 1:1, and 2:1 ratios, we were able to assess coverage, sensitivity, quantitative reproducibility and accuracy. We demonstrate here that affordable MS instruments that yield partial proteome coverage can still provide sufficient data to support investigation of proteome-level changes with a view to optimisation of economically viable production. While higher-cost instruments such as the timsTOF HT and Exploris 480 were able to quantify greater proportions of the two bacterial proteomes investigated here, lower-cost platforms such as the Xevo MRT and Vion could still be used effectively to monitor core metabolic pathways. Such platforms could therefore find a place in troubleshooting or process control, such as the FDA/EMA Quality by Design (QbD) regulatory framework for biopharmaceuticals[34–39], relating key metabolic pathways to bioreactor productivity and product quality. Alternatively, high end instruments that enable more in-depth and robust protein quantification are better suited for initial discovery studies as a pre-cursor to targeted, analyte focussed assays using selected/multiple/parallel reaction monitoring (SRM/MRM/PRM) 38,39, as we have demonstrated previously for over 1,800 proteins in S. cerevisise[40] and for biomarkers of health in CHO cell culture[8] , or the application of non-MS methods such as qPCR or multiplexed immunoassays[41].

We demonstrate that quantitative proteomics can provide crucial insights during production process development, particularly with novel organisms, and could further support production monitoring and troubleshooting. Exploiting the potential of proteomics and other systems biology tools to advance our understanding of living systems that are put to work for manufacturing.

Online Methods

Bacteria Strains and Cultures

Halomonas bluephagenesis sp TD01 is tolerant of high salt and pH conditions has been under continuous development for biomanufacturing since its isolation in Aydingkol Lake in Xinjiang, China in 2009[17–24]. The control *E. coli* (K12/MG1655) samples were processed identically as both a control and for comparison against literature data.

LB Broth Miller was prepared from powder (Formedium, 25 g L^{-1} comprising tryptone 10 g L^{-1}, yeast extract 5 g L^{-1} and sodium chloride 10 g L^{-1}, pH 7.0) and for *H. bluephagenesis* adjusted to 5% sodium chloride (additional 40 g L^{-1}) and pH adjusted with sodium hydroxide (5 M, 4 mL L^{-1}, to pH 9). Liquid media was sterilised by autoclave. For plates, agar (15 g L^{-1}) was added to the appropriate sterile low- or high-salt media described above and heated in a microwave (30% power, 20 min) then poured into plates and allowed to cool in a laminar flow hood. Plates were stored (5 °C) for up to four weeks prior to use.

Cells from a glycerol stock were aseptically inoculated onto LB-agar plates, incubated (37 °C, overnight) then transferred to cold storage (4 °C). Single bacterial colonies of *E. coli* or *H. bluephagenesis* were inoculated aseptically into liquid media (50 mL) in a 250 mL conical flask. Cultures were incubated (37 °C, 200 rpm) and OD600 monitored. When OD600 reached 0.3–0.6, in which growth was late log/early stationary phase, culture aliquots (450 µL) were recovered to pre-weighed 5 mL microfuge tubes on ice and pelleted by centrifugation (3200 g, 10 min, 4 °C). Supernatant was discarded and the wet mass of the pellet determined by difference with tubes empty mass. Pellets were frozen (−20 °C) for later lysis and digestion.

To determine the effect of carbon feed stocks glucose or acetate were prepared from *H. bluephagenesis TD1.0* (a T7-like RNA polymerase-integrated derivative of *H. bluephagenesis sp TD01*[42]) cultures grown in 100 mL MM63 supplemented with 5 gL^{-1} glucose or acetate and harvested in the exponential growth phase after OD600 exceeded 1. Cultures were pelleted by centrifugation (3200 g, 10 min, 4 °C), followed by three washes with phosphate buffered saline and repeat centrifugation. Pellets were frozen (−80 °C) for later lysis and digestion.

Pellet Recovery and Digestion

Pellets were simultaneously defrosted and resuspended in S-trap[43] lysis buffer (7% SDS, 70 mM TEAB, pH 7.55, 3 µL buffer per mg wet mass of pellet). DNA was fragmented by addition of benzonase (Sigma/Merk, 1 µL) and sample vortexed briefly. If required, a second addition of benzonase removed any remaining viscosity. Cell debris were removed by centrifugation (4200 g, 2 min). Samples were buffer exchanged by ultrafiltration (Amicon Ultra 0.5 mL 3 kDa cut-off, 4 washes) with further SDS-free lysis buffer (70 mM TEAB, pH 7.55) removing lipids, carbohydrates, nucleic acid fragments and salt.

Protein was digested with the S-trap[43] mini kit (Protifi) according to the manufacturer's instructions, the sample already being in the recommended SDS-lysis buffer (see above). Briefly, sample volume was taken to give 1–5 mg total protein, reduced with TCEP (60 °C, 30 min, 50 mM final concentration), alkylated with iodoacetamide (room temperature in the dark, 30 min, 40 mM final concentration). The sample was acidified with phosphoric acid (1/10 vol, 1.2% final concentration) and diluted in S-trap binding buffer (7 vols, 90% methanol, 100 mM TEAB). The sample was loaded onto S-trap in aliquots of 190 µL aliquots followed by centrifugation (4,000 g, 1 min). The S-trap was washed with S-trap binding buffer (4 × 150 µL) with centrifugation as for loading. Protein was digested by addition of trypsin (Promega, 20 µL, 0.5 µg μL^{-1}) in digestion buffer (50 mM TEAB) and incubated (47 °C, 1 hr) in a water bath.

Peptide Recovery Estimate

The yield of peptides was determined by the Pierce quantitative fluorometric peptide assay (PN:23290) according to the manufacturer's instructions. Fluorometry was performed on a Clariorstar-Plus plate reader (BMG Labtech) and data processed with Mars v3.40 R2 fitting a four-parameter curve.

High pH Fractionation for Vion Spectral Library

To enable analysis of HDMSe data obtained on the Vion, a spectral library was acquired from a set of fractionated samples. Peptides (50 µg, 300 µL 0.1% TFA) were loaded onto a high-pH C18 spin column (Pierce peptide desalting columns PN:89852) each column and centrifuged (3000 *g*, 2 min), washed with pure water (300 µL) and centrifuged as above. Peptide was eluted in fractions by stepwise addition of stepwise increasing acetonitrile concentration (5, 7.5, 10, 12.5, 15, 17.5, 20,

50% acetonitrile, 300 mM ammonium bicarbonate adjusted to pH 10 with ammonium hydroxide), collecting as separate fractions. Fractions were lyophilized by vacuum centrifugation. Peptide was assumed to be approximately equally split across fractions giving 6-7 µg per fraction.

Fixed Ratio Samples

The fluorometric estimate of peptide recovery from both the *E. coli* and *H. bluephagenesis* whole cell lysates allowed the concentrations of both to be equalised at 0.2 µg µL. Volumes of *E. coli* and *H. bluephagenesis* whole cell lysates (150:150, 200:100, and 100:200 µL) were combined to give samples of 300 µL, 60 µg total peptide in species specific ratios of 1:1, 1:2 and 2:1 in a similar approach to that adopted in multi-site comparison studies[10]. Each sample was divided into aliquots (10 × 30 µL, 6 µg each) and lyophilized alongside aliquots of pure single species peptide (30 µL, 6 µg each).

Sample Preparation for LC-MS

The full sample set now comprised eight fractions each of *E. coli* and *H. bluephagenesis* peptides to generate a spectral library on the Vion; combined ratios of whole cell lysate digests at 1:1, 1:2 and 2:1 to assess instrument performance and a pure species specific sample for each species, each sample of which aliquoted at of 6 µg lyophilized peptide. For LC-MS analysis lyophilised peptides were resuspended in loading buffer (3% v/v acetonitrile, 0.1% v/v formic acid, 24 µL, 0.25 µg μL^{-1} final peptide concentration).

The fractionated samples were injected onto the Vion, in order to create an instrument specific spectral library. Data was acquired in triplicate for 1:1, 2:1, 1:2 and single species samples on each of the instruments in the study.

LC-MS Waters Vion

Liquid Chromatography

Peptide were injected onto a Waters Acquity M-Class with a nanoEase M/Z peptide CSH 130 Å pore size, 1.7 µm particle size 150 µm ID, 150 mm in length in a column oven at 50 °C at starting conditions of 97% mobile phases A (10.1% formic acid) and 3% mobile phase B (acetonitrile, 0.1% formic acid) at 2 µL min^{-1} for three minutes. Peptides were eluted with an analytical gradient of 3-30% B over 3-65 minutes, followed by a steeper gradient to 45% B by 78 minutes. The column was then washed with 95% B for 5 minutes and equilibrated to starting conditions.

Vion HDMSe

LC-eluent entered the Vion IMS QTOF (Waters) instrument via the Waters nanospray source in positive polarity and sensitivity analyser mode. Nebulising gas flow to the emitter was supplied directly from the gas line supplying the instrument through a control valve, rather that by the instrument itself. Data was acquired in HDMSe mode *m/z* range 120-2000 with 1 s scan times. Low energy scan collision energy was 7V, high energy collision energy ramped from 22-28 V. Voltages and gas flows were as follows Capillary: 4 µV, Sample cone voltage 40 V, Source offset voltage 80 V, Cone gas: 50 L h^{-1}; Purge gas: 350 L h^{-1}.

LC-MS Sciex ZenoTOF

Liquid Chromatography

Peptide were injected onto a Waters Acquity M-Class with identical setup and method to that described for the Vion above.

ZenoTOF SWATH-MS

LC-eluent entered the ZenoTOF 7600 (Sciex) mass spectrometer through a Sciex 1-50uL Micro/Microcal optiflow source with a 1-10 µL emitter installed. The source temperature was 225°C and gasses were as follows, curtain gas 35; CAD gas 7; ion source gas 1 20; and ion source gas 2 60 PSI The SWATH-MS method comprised an MS-TOF scan (0.1 s accumulation time; *m/z* range 400–2050) followed by 100 Zeno-trap SWATH scans (0.018 s accumulation time; *m/z* range 100-2000) of variable widths with stepped collision energy (see supplementary data for tabulated values). Data was acquired for minutes 8-35 post injection in a 50 minute method, the total scan time was 2.394 s and 1253 complete scans were acquired through the acquisition.

Exploris 480

Liquid Chromatography

Peptide were loaded onto Evotips using the manufacturers protocol and analysed by the Evosep One LC system (Evosep) using the inbuilt "15 samples per day" method. The operation of the EvoSep is different to that of the other LC systems in the study, but for comparison the analytical gradient approximates to 100% mobile phase A at 220 nL min^{-1} to 35% B over 88 min through a Thermo Scientific ES906 Easyspray column with built in emitter (C18 150 m x 150 mm 2 m column).

Exploris DIA

Data was acquired on an Exploris 480 mass spectrometer (Thermo Scientific) in DIA profile mode. The method comprised 42 DIA fragment window scans with a full scan every 14 scans, for a total of 45 scans per duty cycle. Full scans were aquired in the *m/z* range 345-900 at 120,000 resolution with 300% normalised AGC target and automatically selected max IT. Data-independent fragments were acquired in 42 13 *m/z* fragmentation windows selected with "window placement optimisation" without overlap; over an acquisition range 350-896 *m/z*; with 15 000 resolution; NCE at 32%; and AGC target set to 1000% A full scan was acquired after every 14 DIA scans, so it took 3 full scans (with each CV) to fragment the ions in all 42 DIA windows.

timsTOF HT

Liquid Chromatography

Peptides were injected onto a NanoElute (Bruker Daltonics) ultra-high-pressure Nano-flow chromatography with a trapping column (Thermo Trap Cartridge, PepMap100, C18, 300 m x 5 mm) and analytical column (PepSep, twenty-five Series 150 µm, 1.5 µm column). Peptides were eluted with an analytical gradient of 2-35% B over 90 minutes at a flow rate of 0.5 L/min, then washed at 95% B before re-equilibration to starting conditions.

TimsTOF DIA

Data was acquired on a TimsTOF HT mass spectrometer (Bruker Daltonics) with a nano-electrospray ion source (CaptiveSpray, Bruker Daltonics) in dia-PASEF mode. The dia-PASEF method was defined with a *m/z* range 400-1201 and a mobility range of 1/K0 = 0.6-1.6 Vs/cm2 using equal ion accumulation and ramp times of 100 ms in the dual TIMS analyzer. The collision energy was lowered as a function of increasing ion mobility from 59 eV at 1/K0 = 1.4 Vs/cm^2 to 20 eV at 1/K0 = 0.6 Vs/cm^2. The dia-PASEF window scheme consisted of 32 isolation windows of 26 *m/z* width (cycle time: 1.8 s) with 1 Da mass overlap and 1 ion mobility window.

MRT

Liquid Chromatography

Peptide samples were loaded onto Evotips using the manufacturers protocol and analysed by the Evosep One LC system (Evosep) using the inbuilt "15 samples per day" method in an identical manner to the Exploris 480 method described above.

MRT MS^{e}

The Xevo MRT mass spectrometer was operated in positive electrospray ionisation (ESI) mode. The capillary voltage was 2.8 kV, cone voltage was 30 V and source temperature was set at 100 °C. Data were acquired over *m/z* range 50-2000 with a scan rate of 1 Hz. All mass spectral data were acquired in continuum mode using MS^{E} to obtain fragmentation data simultaneously[44]. Function one (low energy) data were collected using a constant collision energy of 6 meV whilst the second (high energy) function consisted of a collision energy ramp of 25 to 55 meV. For mass accuracy, leucine enkephalin (*m/z* = 556.27658) (100 mpg µL^{-1}, 50:50 acetonitrile:water, 0.1 % formic acid) was acquired every 5 min. The lock mass was delivered to the reference sprayer of the MS source using the Waters Reagent Delivery System with a flow rate of 1 µL min^{-1} .

Data Processing

The complete data processing pipelines were scripted for command line tools and both the script and an explanatory pdf including both code and documentation are included in supplementary information. Brief descriptions of each processing pipeline are included below.

Spectral Library Generation for Skyline

The spectral library for Waters Vion used in the Skyline[45] analysis was generated from fractionate sample data aquired on the Vion as described above. Data was exported from the Vion instrument in uep format and converted to Waters raw format. The "raw" files were processed through the Waters Protein Lynx Global Server tools as described in the supplementary information producing a "final_fragment.csv" file for each injection. These files were then combined with a bespoke R script, and the resulting table processed through MAYU[46] from the trans-proteomics-pipeline to estimate and control the false discovery rate. The combined final_fragment.csv file was then imported into Skyline as a spectral library.

Vion Data Processed through Skyline

A skyline document was set up with the spectral library generated on the Vion as described above. Vion data for the 1:1, 1:2 and 2:1 ratio samples was imported into Skyline, and an MSstats results table exported.

timsTOF; Exploris; ZenoTOF and MRT DIANN processing

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[47] partner repository with the dataset identifier PXD071869[48]. Briefly for each complete data set of triplicate injections of each ratio 1:1, 1:2 and 2:1 by total peptide mass, of *E. coli* and *H. bluephagenesis*from the timsTOF HT; Exploris 480; ZenoTOF 7600 and SELECT SERIES MRT instruments comprising were processed with DIANN v2.3.0. The MRT data was denoised and centroided using OpenMS tools prior to processing with DIANN, all other data was processed directly from proprietary raw data format. The fasta protein sequence database is included in the data deposition and was a concatenation of the combined reference proteome for *E. coli K12* (UP000000625) and only available proteome for *H. bluephagenesis sp. TD01* (UP000838464, now available through EBI UniParc) downloaded from uniprot (https://www.uniprot.org/). DIANN was run with a q-value threshold of 1%, a precursor m/z range 400-2000, fragment m/z range 100-2000, precursor charges 2-4, up to 1 missed cleavage, up to two variable modifications of oxidised methionine, constant carbamidomethylation of cysteine, and match between runs enabled. The resulting "report.parquet" files were imported into R, prepared for processing using MSstats with the "DIANNtoMSstatsFormat" function and then processed with MSstats which provided absolute protein estimates per sample on a log10 scale, and ratio estimates between samples. The resulting quantification tables were used to produce the figures reported in the Results and Discussion section.

Statistical Analysis

All data analysis was performed in R version 4.5.2 (2025-10-31 ucrt), data sets were quantified and differential expression between samples calculated using MSStats[49]. Scripts enabling reproduction of the analysis are included in supplementary information and with the data deposition at PRIDE[47]supporting this publication.

Data Availability

Further data referred to in this manuscript are found in the supplementary information and supporting dataset. Mass spectrometry data is available via ProteomeXchange with identifier PXD071869[48].

Acknowledgements

This work was supported by the Future Biomanufacturing Research Hub (grant EP/S01778X/1), funded by the Engineering and Physical Sciences Research Council (EPSRC) and Biotechnology and Biological Sciences Research Council (BBSRC) as part of UK Research and Innovation. Instrumentation was funded by BBSRC (grantgrants BB/W019892/1 and BB/T018127/1). We acknowledge the support of staff in the Mass Spectrometry and Separations Facility of the Faculty of Science and Engineering at the University of Manchester and in the Centre for Proteome Research (CPR) at the University of Liverpool. Finally, we thank Waters for their continued support to the Michael Barber Centre for Collaborative Mass Spectrometry and specifically Lee Gethings and Dale Shepherd for their help in acquiring data on the MRT.

References

[1] Biotechnology Market Size & Outlook (2025). URL https://www.grandviewresearch.com/horizon/outlook/biotechnology-market-size/global.

[2] Juan-Manuel Valverde, Juan-Manuel Valverde, Carmen Avilés-Palacios & Avilés-Palacios, C. Circular Economy as a Catalyst for Progress towards the Sustainable Development Goals: A Positive Relationship between Two Self-Sufficient Variables. *Sustainability* **13**, 12652 (2021).

[3] Liang, J., Tian, J., Zhang, H., Li, H. & Chen, L. Proteomics: An In-Depth Review on Recent Technical Advances and Their Applications in Biomedicine. *Medicinal Research Reviews* **45**, 1021–1044 (2025). URL https://onlinelibrary.wiley.com/doi/10.1002/med.22098.

[4] Abele, M. *et al.* Proteomic Diversity in Bacteria: Insights and Implications for Bacterial Identification. *Molecular & Cellular Proteomics* **24** (2025). URL https://www.mcponline.org/article/S1535-9476(25)00015-5/abstract.

[5] Monteiro, R. *et al.* The Secretome landscape of Escherichia coli O157:H7: Deciphering the cell-surface, outer membrane vesicle and extracellular subproteomes. *Journal of Proteomics* **232**, 104025 (2021). URL https://www.sciencedirect.com/science/article/pii/S1874391920303936.

[6] Zubair, M. *et al.* Proteomics approaches: A review regarding an importance of proteome analyses in understanding the pathogens and diseases. *Frontiers in Veterinary Science* **9**, 1079359 (2022). URL https://www.frontiersin.org/articles/10.3389/fvets.2022.1079359/full.

[7] Meissner, F., Geddes-McAlister, J., Mann, M. & Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nature Reviews Drug Discovery* 1–18 (2022). URL https://www.nature.com/articles/s41573-022-00409-3.

[8] Eldrid, C. *et al.* From Discovery to Delivery: A Rapid and Targeted Proteomics Workflow for Monitoring Chinese Hamster Ovary Biomanufacturing. *Molecular & Cellular Proteomics* **24**, 101011 (2025). URL https://linkinghub.elsevier.com/retrieve/pii/S1535947625001100.

[9] Gurdo, N., Volke, D. C. & Nikel, P. I. Merging automation and fundamental discovery into the design-build-test-learn cycle of nontraditional microbes. *Trends in Biotechnology* **40**, 1148–1159 (2022).

[10] Navarro, P. *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. *Nature Biotechnology* **34**, 1130–1136 (2016). URL https://www.nature.com/articles/nbt.3685.

[11] Russell, M. *et al.* Baseline proteomics characterisation of the emerging host biomanufacturing organism Halomonas bluephagenesis. *Scientific Data* **9** (2022). URL https://doi.org/10.1038%2Fs41597-022-01610-0.

[12] Huang, Q., Szklarczyk, D., Wang, M., Simonovic, M. & von Mering, C. PaxDb 5.0: Curated Protein Quantification Data Suggests Adaptive Proteome Changes in Yeasts. *Molecular & Cellular Proteomics* **22**, 100640 (2023). URL https://www.sciencedirect.com/science/article/pii/S1535947623001512.

[13] Lv, X., Xue, H., Qin, L. & Li, C. Transporter Engineering in Microbial Cell Factory Boosts Biomanufacturing Capacity. *BioDesign Research* **2022**, 9871087 (2022). URL https://www.sciencedirect.com/science/article/pii/S2693125724000761.

[14] Dunlop, M. J. *et al.* Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology* **7** (2011).

[15] Wang, M., Deng, W.-W., Zhang, Z.-Z. & Yu, O. Engineering an ABC Transporter for Enhancing Resistance to Caffeine in Saccharomyces cerevisiae. *Journal of Agricultural and Food Chemistry* **64**, 7973–7978 (2016). URL https://doi.org/10.1021/acs.jafc.6b03980.

[16] Pandey, K., Pandey, M., Kumar, V., Aggarwal, U. & Singhal, B. Bioprocessing 4.0 in biomanufacturing: Paving the way for sustainable bioeconomy. *Systems Microbiology and Biomanufacturing* **4**, 407–424 (2024). URL https://doi.org/10.1007/s43393-023-00206-y.

[17] Tan, D., Xue, Y.-S., Aibaidula, G. & Chen, G.-Q. Unsterile and continuous production of polyhydroxybutyrate by Halomonas TD01. *Bioresource Technology* **102**, 8130–8136 (2011). URL http://www.sciencedirect.com/science/article/pii/S096085241100753X.

[18] Tan, D., Wu, Q., Chen, J.-C. & Chen, G.-Q. Engineering Halomonas TD01 for the low-cost production of polyhydroxyalkanoates. *Metabolic Engineering* **26**, 34–47 (2014). URL https://www.sciencedirect.com/science/article/pii/S1096717614001153.

[19] Tao, W., Lv, L. & Chen, G.-Q. Engineering Halomonas species TD01 for enhanced polyhydroxyalkanoates synthesis via CRISPRi. *Microbial Cell Factories* **16**, 48 (2017). URL https://doi.org/10.1186/s12934-017-0655-3.

[20] Zhao, H. *et al.* Novel T7-like expression systems used for Halomonas. *Metabolic Engineering* **39**, 128–140 (2017). URL https://www.sciencedirect.com/science/article/pii/S1096717616301367.

[21] Xiao-Ran, J., Jin, Y., Xiangbin, C. & Guo-Qiang, C. Halomonas and Pathway Engineering for Bioplastics Production. *Methods in Enzymology* **608**, 309–328 (2018).

[22] Shen, R., Ning, Z.-Y., Lan, Y.-X., Chen, J.-C. & Chen, G.-Q. Manipulation of polyhydroxyalkanoate granular sizes in Halomonas bluephagenesis. *Metabolic Engineering* **54**, 117–126 (2019). URL https://www.sciencedirect.com/science/article/pii/S1096717619300990.

[23] Bekker-Jensen, D. B. *et al.* A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Molecular & Cellular Proteomics* **19**, 716–729 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S1535947620350271.

[24] Jiang, X.-R., Yan, X., Yu, L.-P., Liu, X.-Y. & Chen, G.-Q. Hyperproduction of 3-hydroxypropionate by Halomonas bluephagenesis. *Nature Communications* **12**, 1513 (2021). URL https://www.nature.com/articles/s41467-021-21632-3.

[25] Zhang, J. *et al.* Nonsterile microbial production of chemicals based on Halomonas spp. *Current Opinion in Biotechnology* **85**, 103064 (2024). URL https://www.sciencedirect.com/science/article/pii/S095816692300174X.

[26] Ruiz, N. & Silhavy, T. J. How Escherichia coli Became the Flagship Bacterium of Molecular Biology. *Journal of Bacteriology* **204**, e0023022 (2022).

[27] Blount, Z. D. The unexhausted potential of E. coli. *eLife* **4**, e05826 (2015). URL https://elifesciences.org/articles/05826.

[28] Gu, Y. *et al.* Advances and prospects of Bacillus subtilis cellular factories: From rational design to industrial applications. *Metabolic Engineering* **50**, 109–121 (2018).

[29] Spasic, J., Mandic, M., Djokic, L. & Nikodinovic-Runic, J. Streptomyces spp. in the biocatalysis toolbox. *Applied Microbiology and Biotechnology* **102**, 3513–3536 (2018).

[30] Nikel, P. I. & de Lorenzo, V. Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metabolic Engineering* **50**, 142–155 (2018).

[31] Wendisch, V. F. *et al.* The flexible feedstock concept in Industrial Biotechnology: Metabolic engineering of Escherichia coli, Corynebacterium glutamicum, Pseudomonas, Bacillus and yeast strains for access to alternative carbon sources. *Journal of Biotechnology* **234**, 139–157 (2016).

[32] Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: A perspective review from classical to non-traditional microorganisms. *Microbial Biotechnology* **12**, 98–124 (2018). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302729/.

[33] Yu, L.-P., Wu, F.-Q. & Chen, G.-Q. Next-Generation Industrial Biotechnology-Transforming the Current Industrial Biotechnology into Competitive Processes. *Biotechnology Journal* **14**, 1800437 (2019). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/biot.201800437.

[34] Holm, P., Allesø, M., Bryder, M. C. & Holm, R. in *Q8(R2): Pharmaceutical Development* 1 edn, (eds Teasdale, A., Elder, D. & Nims, R. W.) *ICH Quality Guidelines* 535–577 (Wiley, 2017). URL https://onlinelibrary.wiley.com/doi/10.1002/9781118971147.ch20.

[35] Daniela M. Zagalo *et al.* Quality by design (QbD) approach in marketing authorization procedures of Non-Biological Complex Drugs: A critical evaluation. *European Journal of Pharmaceutics and Biopharmaceutics* **178**, 1–24 (2022).

[36] Beg, S., Rahman, M. & Panda, S. Pharmaceutical QbD: Omnipresence in the product development lifecycle. *European Pharmaceutical Review* **22**, 58–64 (2017).

[37] Zhang, L. & Mao, S. Application of quality by design in the current drug development. *Asian Journal of Pharmaceutical Sciences* **12**, 1–8 (2017). URL https://www.sciencedirect.com/science/article/pii/S1818087616300575.

[38] *Introduction to Quality by Design (QbD): Fundamentals, Principles, and Applications*, 1–17 (Elsevier, 2019). URL https://linkinghub.elsevier.com/retrieve/pii/B9780128157992000010.

[39] JG Duarte, MG Duarte, AP Piedade & F Mascarenhas-Melo. Rethinking Pharmaceutical Industry with Quality by Design: Application in Research, Development, Manufacturing, and Quality Assurance. *The AAPS journal* (2025).

[40] Lawless, C. *et al.* Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring*. *Molecular & Cellular Proteomics* **15**, 1309–1322 (2016). URL https://www.sciencedirect.com/science/article/pii/S1535947620336239.

[41] Vashist, S. K. in *Multiplex Immunoassays* (eds Vashist, S. K. & Luong, J. H.) *Point-of-Care Technologies Enabling Next-Generation Healthcare Monitoring and Management* 177–196 (Springer International Publishing, Cham, 2019). URL https://doi.org/10.1007/978-3-030-11416-9_7.

[42] Su, Q. *et al.* Engineering a mevalonate pathway in Halomonas bluephagenesis for the production of lycopene. *Frontiers in Microbiology* **13**, 1100745 (2023). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9885113/.

[43] HaileMariam, M. *et al.* S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *Journal of Proteome Research* **17**, 2917–2924 (2018). URL https://pubs.acs.org/doi/10.1021/acs.jproteome.8b00505.

[44] Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Molecular & Cellular Proteomics* **5**, 144–156 (2006).

[45] MacLean, B. *et al.* Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844992/.

[46] Reiter, L. *et al.* Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. *Molecular & Cellular Proteomics* **8**, 2405–2417 (2009). URL https://www.mcponline.org/article/S1535-9476(20)33929-3/abstract.

[47] Perez-Riverol, Y. *et al.* The PRIDE database at 20 years: 2025 update. *Nucleic Acids Research* **53**, D543–D553 (2025). URL https://doi.org/10.1093/nar/gkae1011.

[48] Multi-instrument DIA proteomics of Escherichia coli and Halomonas bluephagenesis TD01 ratio mixtures for biomanufacturing-relevant instrument comparison PXD071869 (2020). URL https://doi.org/10.6019/PXD071869.

[49] Choi, M. Msstats: an r package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30** (2014).

Tables

Table 1: Table of headline protein and peptide identification count at 1% FDR for each instrument along side an approximate guide price for the instrument.

Instrument	No. Proteins IDs	No. Peptide IDs	Instrument List Price
zenoTOF	3402	28163	£510k
timsTOF	5554	71738	£900k
Exploris	4726	53836	£490k
MRT	969	5248	£390k
Vion	862	3376	£200k

Figures

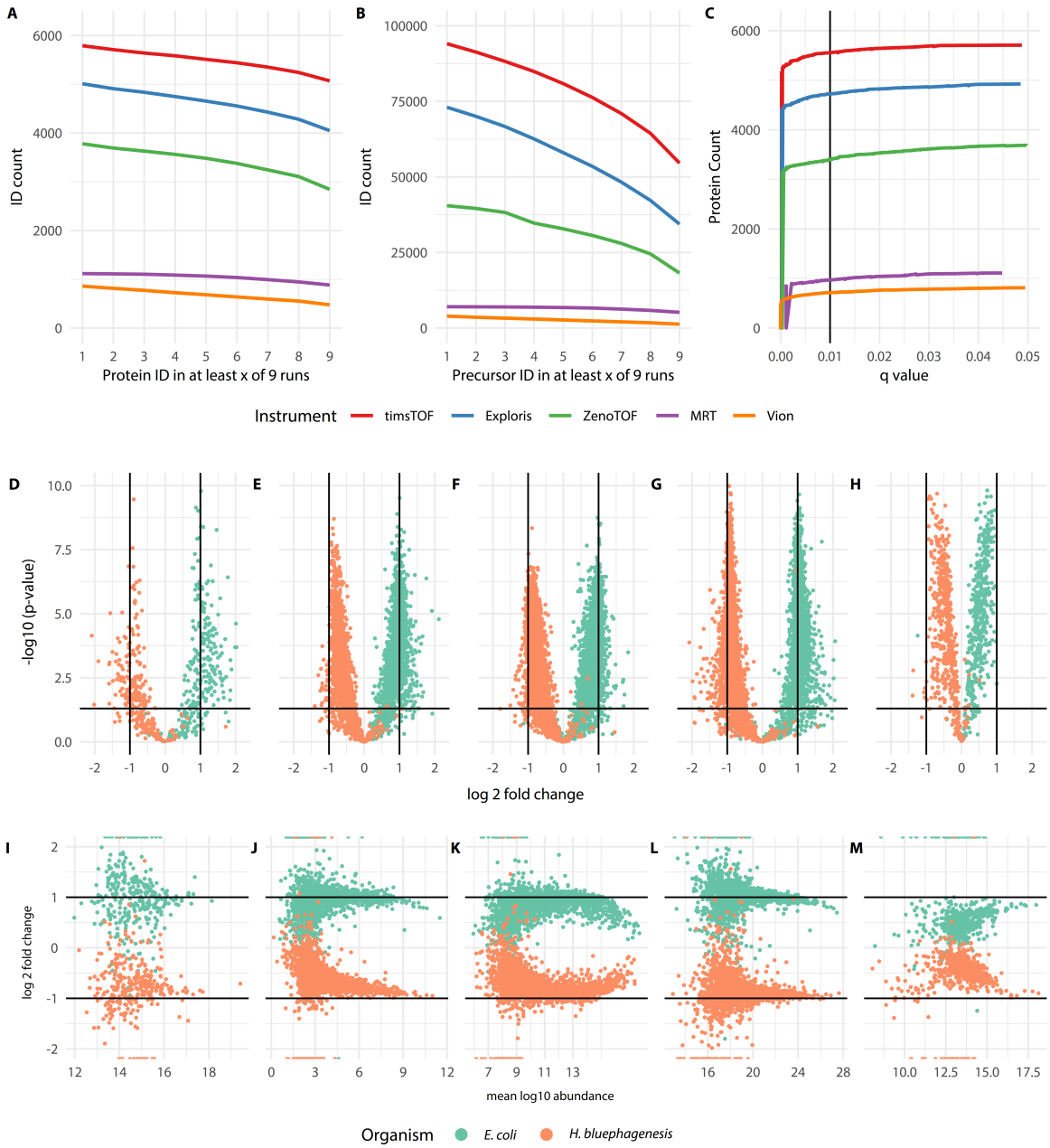

Figure 1: Consistency of protein (A) and peptide (B) identification across 9 injections. Protein identification q-value against number of proteins identified showing protein identification saturates around q-value 0.01, roughly a FDR of 1%. Volcano plots (D-H) and MA plots (I-M), both from left to right Vion, ZenoTOF, timsTOF, Exploris and MRT respectively, allow relative comparison of the instrument overall performance in terms of sensitivity to differential expression.

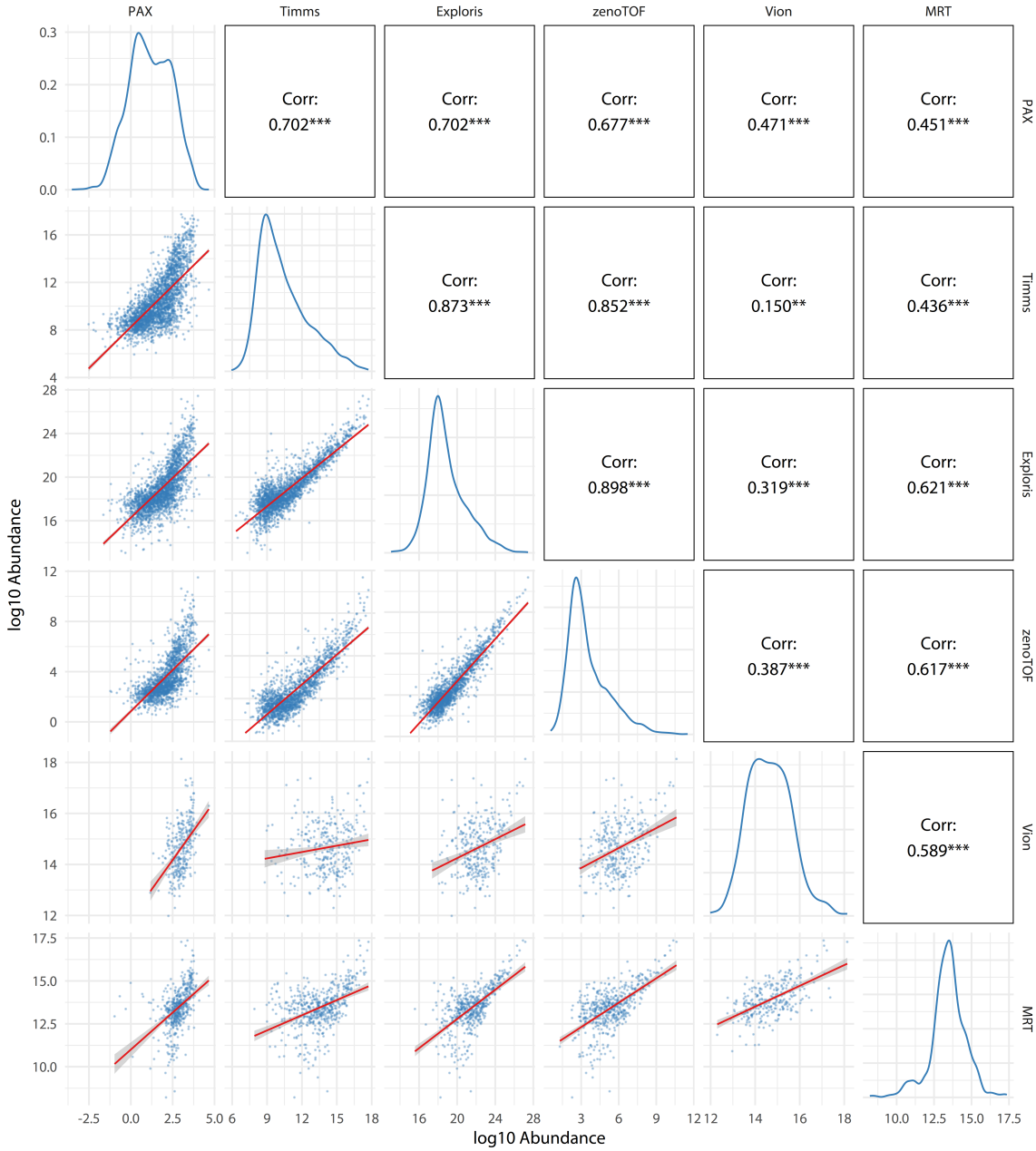

Figure 2: Plot of log10 protein abundance estimates from MSstats analyiss of *E. coli* data from each of the instruments and the PAX data base.

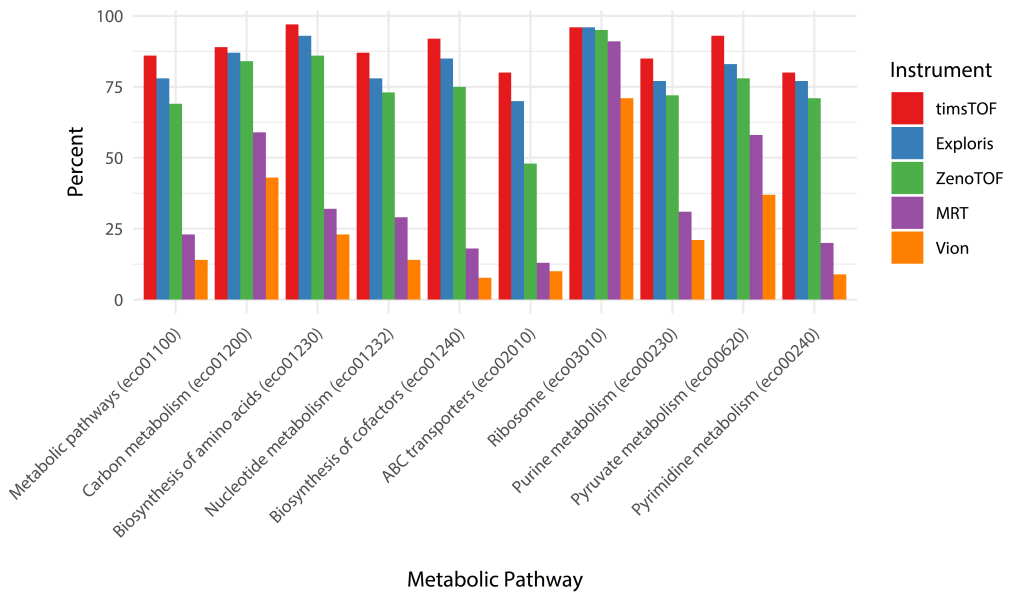

Figure 3: For the *E. coli* data the percentage coverage of proteins annotated in KEGG as belonging to each of the annotated pathways for each of the instruments in the study.

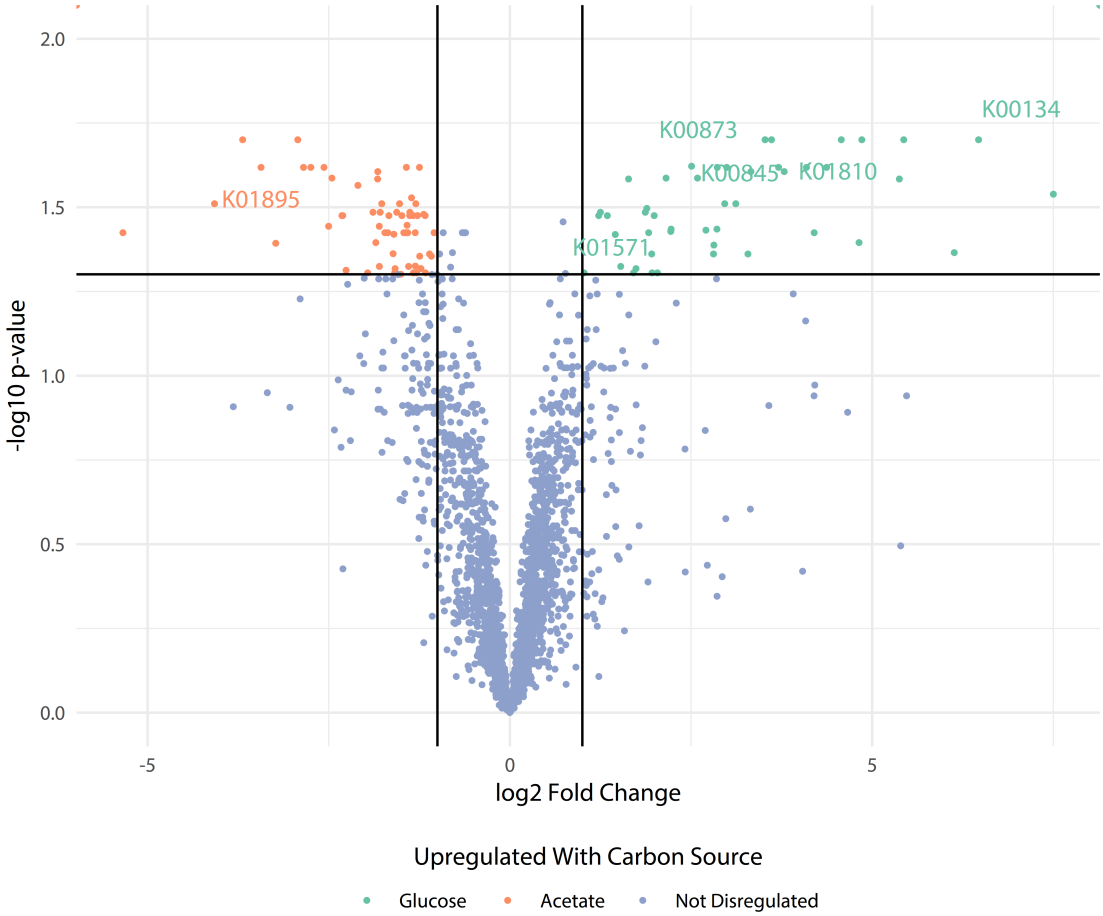

Figure 4: Volcano plot showing protein disregulation in *H.bluephagenesis* grown with carbon sources either glucose or acetate from protemics data aquired on Sciex ZenoTOF. Proteins with statistically significant (adjusted p-value <=0.01) disregultion greater than 2-fold between conditions are highlighted, and disregulated proteins assossiated with glycolysis or acetate metabolism are annotated with uniprot accesion numbers.