



Metabolomics and modelling approaches for systems metabolic engineering

Jasmeet Kaur Khanijou^a, Hanna Kulyk^{b,g}, Cécilia Bergès^{b,g}, Leng Wei Khoo^a, Pnelope Ng^a,
Hock Chuan Yeo^c, Mohamed Helmy^{c,d}, Floriant Bellvert^{b,g}, Wee Chew^{a,**},
Kumar Selvarajoo^{a,c,e,f,*}

^a Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), Singapore 138673, Singapore

^b Toulouse Biotechnology Institute, Bio & Chemical Engineering (TBI), Université de Toulouse, CNRS 5504 - INRA 792 - INSA, France

^c Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore

^d Department of Computer Science, Lakehead University, ON, P7B 5E1, Canada

^e Synthetic Biology for Clinical and Technological Innovation (SynCTI), National University of Singapore (NUS), Singapore 117456, Singapore

^f School of Biological Sciences, Nanyang Technological University (NTU), Singapore 637551, Singapore

^g MetaToul-MetaboHUB, National Infrastructure of Metabolomics & Fluxomics (ANR-11-INBS-0010), France

ARTICLE INFO

Keywords:

Quantitative metabolomics
Dynamic metabolomics
Spatial metabolomics
Kinetic modelling
Constraint-based modelling
Machine learning

ABSTRACT

Metabolic engineering involves the manipulation of microbes to produce desirable compounds through genetic engineering or synthetic biology approaches. Metabolomics involves the quantitation of intracellular and extracellular metabolites, where mass spectrometry and nuclear magnetic resonance based analytical instrumentation are often used. Here, the experimental designs, sample preparations, metabolite quenching and extraction are essential to the quantitative metabolomics workflow. The resultant metabolomics data can then be used with computational modelling approaches, such as kinetic and constraint-based modelling, to better understand underlying mechanisms and bottlenecks in the synthesis of desired compounds, thereby accelerating research through systems metabolic engineering. Constraint-based models, such as genome scale models, have been used successfully to enhance the yield of desired compounds from engineered microbes, however, unlike kinetic or dynamic models, constraint-based models do not incorporate regulatory effects. Nevertheless, the lack of time-series metabolomic data generation has hindered the usefulness of dynamic models till today. In this review, we show that improvements in automation, dynamic real-time analysis and high throughput workflows can drive the generation of more quality data for dynamic models through time-series metabolomics data generation. Spatial metabolomics also has the potential to be used as a complementary approach to conventional metabolomics, as it provides information on the localization of metabolites. However, more effort must be undertaken to identify metabolites from spatial metabolomics data derived through imaging mass spectrometry, where machine learning approaches could prove useful. On the other hand, single-cell metabolomics has also seen rapid growth, where understanding cell-cell heterogeneity can provide more insights into efficient metabolic engineering of microbes. Moving forward, with potential improvements in automation, dynamic real-time analysis, high throughput workflows, and spatial metabolomics, more data can be produced and studied using machine learning algorithms, in conjunction with dynamic models, to generate qualitative and quantitative predictions to advance metabolic engineering efforts.

1. Introduction

Systems metabolic engineering drives the transformation of microorganisms into effective bio-factories that produce large amounts of target molecules for various industries such as food, biotechnology, and

pharmaceuticals (Choi et al., 2019). The integration of -omics data with computational tools used in systems biology defines systems metabolic engineering. This systems-level metabolic engineering integration allows the creation of new metabolic pathways and products and the rewiring of regulatory circuits (Lee et al., 2011), which in turn aids in

* Corresponding author. Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore.

** Corresponding author.

E-mail addresses: chew_wei@sifbi.a-star.edu.sg (W. Chew), kumar_selvarajoo@bii.a-star.edu.sg (K. Selvarajoo).

<https://doi.org/10.1016/j.mec.2022.e00209>

Received 20 June 2022; Received in revised form 13 October 2022; Accepted 14 October 2022

Available online 15 October 2022

2214-0301/© 2022 The Authors. Published by Elsevier B.V. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

improving strain development.

Metabolomics plays a crucial role amongst the -omics technologies. As the metabolome is downstream of the proteome, it is the endpoint of biological processes, hence, reflecting cell responses and phenotypical interplays from genetic and environmental perturbations (Griffin, 2006). Furthermore, metabolites also maintain intercellular signalling, energy balance and other cellular functions (Rubakhin et al., 2011; Amantonico et al., 2010). With improvements in the quality and coverage of metabolomics technologies, changes in metabolite levels due to complex biological interactions can be determined and can be integrated into models to determine the physiological state of a metabolic network (Töpfer et al., 2015; Volkova et al., 2020). Through systems-level metabolic engineering, with the focus on integrating metabolomics data with modelling approaches, the efficiency of microbial cell factories can be enhanced. Moreover, integrating metabolomics data together with fluxomics, transcriptomics, and proteomics into mathematical models seeks to promote rational approaches to improve strain and cell-line development (Roume et al., 2013; Vorreiter et al., 2016).

In this review, we focus on metabolomic approaches and the integration of modelling strategies. Insights on the progress and limitations of metabolomic data generation, dynamic and constraint-based modelling approaches will also be shared. Finally, the challenges faced and the potential of machine learning in systems metabolic engineering as a strategy to produce efficient bio-factories will be discussed.

2. Generation of metabolomics data

Metabolomics mainly involves the identification and quantitation of intracellular and extracellular small-molecule metabolites in a biological system. Metabolomics is important in acquiring an overview of a broad array of metabolites concurrently and can be executed in a high throughput manner (Hollywood et al., 2006; Wishart, 2007). Metabolites need to be extracted from cells, and prepared prior to instrumental analysis and data generation, to capture high fidelity time-dependent snapshots of metabolite profiles over the course of culture growth.

2.1. Sample preparation

A metabolomics workflow generally involves a few crucial steps, for instance, robust sample preparations with rapid quenching of metabolism, metabolite extraction, possible chemical derivatization, and instrumental analysis, processing of data with or without database information and additional statistical analysis and data interpretation.

Quick and efficient quenching protocols are essential for accurate quantitative snapshots of metabolic information as the metabolites are in a network that is continuously undergoing various biochemical reactions within a cell (da Luz et al., 2014). Microbial cells are usually quenched with solvent solutions at extreme pH or temperatures in attempt to rapidly arrest intracellular biochemical reactions (Doran et al., 2017). However, such solutions can damage the cell wall which confines the intracellular metabolites within the cell, causing the leakage of intracellular metabolites into the extracellular cell culture medium. These intracellular metabolites become diluted in the more voluminous extracellular medium, adding to significant variability in intracellular metabolite quantitation due to the presence of interfering extracellular metabolites and concentrated media components (Faijes et al., 2007). Fast filtration has also been used to separate microbial cells from cell culture medium prior to quenching through liquid nitrogen immersion or cold/hot solvent (Castaño-Cerezo et al., 2019; Wordofa et al., 2017; Kim et al., 2013; Jagtap et al., 2021). Even though this method overcomes the issue of intracellular metabolite leakage, the quenching of cell metabolism is not immediate due to the time required for the filtration process. Extracellular metabolites can be part of the initial cell culture medium composition and can be introduced through cell secretion, cell lyses or polymer degradation (Behrends et al., 2014).

The turnover rates of extracellular metabolites are much lower than that intracellularly due to the lack of enzymes or the low concentrations of enzymes in the cell culture medium (Mo et al., 2009). Variability of extracellular metabolites usually results due to the presence of living microbial cells in the cell culture medium. Therefore, cells need to be removed quickly from the culture medium to prevent severe changes in the concentrations of extracellular metabolites and quenching solutions can also be employed to prevent metabolite degradation (Patejko et al., 2017; Pinu and Villas-Boas, 2017).

Intracellular metabolites are confined within a cell envelope or membrane, thus extraction of metabolites from the intracellular compartment usually involves the use of extraction solvents (Kim et al., 2013), acids (Zhong et al., 2018) or alkalis (Prasanna et al., 2018). Biological metabolites have extensively varied physicochemical properties, which makes it rather challenging for quick, complete, and non-biased sampling of metabolites from biological systems and their subsequent analysis (Pinu et al., 2017). For instance, methanol extraction is better for the recovery of polar and mid-polar metabolites, as compared to non-polar metabolites (Villas-Bôas et al., 2005a). Thus, either a few different protocols for metabolite sampling have to be employed concurrently or one comprehensive set of processes is utilized.

Solid phase extraction (SPE) and solid phase micro-extraction (SPME) are clean-up strategies, which can be used to reduce sample complexity prior to instrumental analysis. In a recent study, SPE was utilized to enrich 12 metabolites from the glycolysis and pentose phosphate pathways of yeast cells whilst facilitating the concurrent removal of abundant organic acids and sugars (Si-Hung et al., 2019). The matrix reduction aided metabolite quantitation during instrumental analysis. In another study, high throughput SPME followed by instrumental analysis was utilized to identify low-abundance and short-lived metabolites from bacterial cells (Mousavi et al., 2019).

2.2. Mass spectrometry-based metabolomics

The coverage of metabolites can be improved with various analytical techniques. Coupling chromatography to mass spectrometry (MS) enables the identification and quantitation of metabolites through the reduction in sample complexity by enabling the separation of metabolites before MS detection, thus reducing matrix interferences that may occur during the ionization process. Untargeted and targeted metabolomics are two main approaches used to analyze metabolites. Untargeted metabolomics is hypothesis-generating and is executed to analyze all detectable metabolites in a sample (Schrimpe-Rutledge et al., 2016). Targeted metabolomics may involve the validation and absolute quantitation of a selected subset of metabolites, for instance metabolites in a specific pathway (Cai et al., 2015). Targeted metabolomics can also be utilized to complement an untargeted approach. In this section, different metabolomics platforms for generation of metabolomics data will be discussed.

In recent years, the coupling of liquid chromatography (LC) and MS has been increasingly popular due to its versatility, sensitivity, and reproducibility of chromatographic data and mass fragmentation patterns (Gika et al., 2014). These intra- and extracellular metabolites are more suited for liquid chromatographic separation, as most of them are polar and non-volatile (Gika et al., 2014). In a targeted analysis, the application of LC allows these metabolites to separate with specificity and accuracy in quantification. For untargeted analysis of most cellular metabolites, a comprehensive gradient separation (typically using 0.1% formic acid in water and acetonitrile) with a suitable reversed-phase column can be applied (Pyke et al., 2015). Metabolites that are more polar in nature can also be analyzed using typically a HILIC (hydrophilic interaction chromatography) column (Contrepolis et al., 2015). Applying ultra-high performance liquid chromatography (UHPLC/UPLC) methods can achieve good chromatographic resolution of peaks detected by the MS at shorter retention times (Cajka and Fiehn, 2016). Furthermore, the usage of ion mobility and high resolution MS have also

contributed to improving the identification of isomers, providing better insight into biological functions (Wu et al., 2020). Dynamic labelling of metabolites in yeast has been studied using ^{13}C -tracing and LC-MS analysis (Czajka et al., 2020). However, LC-MS can be prone to interferences from the sample matrix resulting in ion suppression which affects reproducibility and accuracy when determining trace metabolites in complex biological samples (Antignac et al., 2005). Ion suppression could suggest the occurrence of less volatile substances that can alter droplet formation or droplet evaporation efficiency, which affects the amount of gaseous charged ion that reaches the detector (Annesley, 2003). Ion suppression can be minimized or corrected by applying clean-up strategies to biological samples, chromatographic re-optimization, modifications to reagents utilized, and suitable calibration approaches such as the use of external calibration using matrix-matched samples, standard addition, internal standards, and dilution (Furey et al., 2013).

Numerous metabolites have carboxyl, hydroxyl, and/or amino functional groups, which are all polar. Therefore, chemical derivatization of extracted metabolites is usually executed prior to gas chromatography (GC) analysis to increase their volatility (Pawliszyn, 2012). However, this additional derivatization step can result in analyte loss and/or artifact formation (Kanani et al., 2008). Central carbon metabolites and amino acids are easily analyzed using GC-MS. Organic acids, sugar phosphates and alcohols, are usually derivatized in pyridine methoxyamine solution followed by trimethylsilylation (Phan and Blank, 2020). However, metabolites that are non-volatile or thermally unstable cannot be analyzed by the GC and requires LC instead. Several groups have also used enzymatic cleavage to convert non-volatile phosphorylated isoprenoid metabolites to their isoprenoid alcohols prior to GC-MS analysis (Huang et al., 2011). However, this approach may not result in high conversion efficiencies for the metabolites, and this further complicates accurate quantitation. For the quantitative analysis of such non-volatile phosphorylated isoprenoid metabolites, LC approaches would be beneficial. Furthermore, unlike LC approaches, the typical ionization method used in GC-MS is electron ionization (EI), which is less susceptible to ionization suppression. The fragmentation from EI ionization is useful for the identification of untargeted metabolites and reduces the demand for tandem MS (Papadimitropoulos et al., 2018). However, this results in metabolite fragments containing phosphate groups or a limited carbon backbone, which causes a loss of isotope labelling information in experiments involving tracers (Chu et al., 2015). For experiments involving tracers, a softer ionization such as chemical ionization (CI) can be utilized instead. In a recent study, negative chemical ionization was utilized to acquire ^{13}C -labelling information of sugar phosphates for ^{13}C metabolic flux analysis (Okahashi et al., 2019).

Capillary Electrophoresis (CE) coupled with MS has also been used in metabolomics (Barbas et al., 2011; Zhang et al., 2017). CE relies on the electrophoretic mobility of molecules arising from its low nanoscale flow for achieving compound separations. Such a difference in molecular separation from CE enables a complementary profiling pattern as compared to LC and GC approaches. However, even though CE has the highest separation efficiency CE-MS is the least robust and sensitive (Zhang et al., 2017).

Ion chromatography (IC) has conventionally been utilized for the analysis of both organic and inorganic ions in liquid samples (Weiß, 1987). Numerous groups have analyzed anionic metabolites such as carbohydrates, nucleotides, organic acids, and sugar phosphates from biological samples through IC coupled with MS (Petucci et al., 2016). IC-MS offers benefits over other profiling platforms with regards to its unique selectivity. The exceptional resolution of IC has enabled the distinction between numerous polar metabolites which are isobaric and isomeric. Furthermore, IC is adept in analyzing various glycolysis and the tricarboxylic acid cycle (TCA cycle) intermediates, which have a wide range of chemical properties. Capillary IC coupled to high resolution MS has also been used to detect low quantities of sugar

phosphates, organic acids, and co-factors, in cell cultures (Hirayama et al., 2020).

Ion mobility MS is an electrophoretic technology based on gas phase, where ions are differentiated in the gas phase and electric field depending on their charge, size, mass, and shape (Ren et al., 2018). Ion mobility MS enables the distinction between isobaric and isomeric metabolites, and can determine the conformation of metabolites (Paglia and Astarita, 2017). LC coupled with ion mobility MS has been used to distinguish isobaric disaccharides in biological samples (Pičmanová et al., 2022). Even though ion mobility MS has high throughput capabilities, aids in the identification of metabolites, and improves metabolome coverage, there are issues with lower sensitivity and repeatability between instruments (Zhang et al., 2018a).

2.3. Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is another popular analytical technique applied in metabolomics. NMR spectroscopy has the advantages of being rapid, highly reproducible with a relatively simple and non-destructive sample preparation protocol, and also allows the analysis of complex mixtures without the need for separation (Murovec et al., 2018). NMR spectroscopy can uniquely characterize the chemical structure of a molecule by determining nuclear chemical shifts and their location inside a magnetic field, where the ratio of peak intensity and proton number is directly proportionate to concentration of the metabolite (Cascante and Marin, 2008). The qualitative and quantitative data generated is especially useful for the exploration of the complex interactions of biological networks in a metabolomics approach. However, compared to MS, NMR has a lower sensitivity with a detection limit of 1–5 μM and also possesses the problem of overlapping signals generated by different metabolites in the spectrum (Zhang and Powers, 2012). This is a problem for 1D NMR which can be overcome by 2D NMR which can better characterize the signals. Despite these drawbacks, the quality of NMR data can always be improved through the use of probes that are cryogenically cooled to 4.5 K, resulting in a four-fold increase in sensitivity and two-dimensional NMR measurements can also be performed (Lankadurai et al., 2013). Stable-isotope-guided NMR has demonstrated the transformation patterns of various metabolites from numerous biota (Uchimiya et al., 2017). MS and NMR are complementary techniques which have provided fundamentally distinct knowledge in metabolomics as demonstrated by several metabolomics studies (Jousse et al., 2017; Meyer et al., 2013; Nematodzi et al., 2020). NMR has been also used to measure *in vivo* metabolites and fluxes (Bastawrous et al., 2018). Furthermore, more recently, NMR methods have been developed whereby the requirement for continuous metabolomics monitoring has been addressed. For instance, flow NMR has been used to study extracellular metabolites from organisms cultivated in liquid culture (Bastawrous et al., 2018).

2.4. Quantitative metabolomics

The quantitation of metabolites is challenging due to the various issues that arise during the quantitative workflow, which involves the quenching of metabolism and the extraction, identification, and quantitation of metabolites. The generation of such quantitative information from metabolic networks is important for model building. By understanding the underlying issues associated with quantitation in each step of the workflow, one can deploy strategies to eliminate or minimize quantitative pitfalls.

The aim of quenching and extraction is to create a metabolic extract which quantitatively reveals the metabolites present. During quenching, the metabolite levels could be perturbed in the harvesting process or there could be incomplete rapid termination of enzyme activity. As mentioned previously, fast filtration followed by rapid placement in quenching solvent is recommended. To compensate for analyte losses

during extraction, standards should be added to the original extraction solvent. Furthermore, as metabolite extracts have numerous compounds that can degrade or interconvert (Siegel et al., 2014), the time interval between sample preparation and analysis should be reduced.

The concentration of metabolites from extracts can be determined by comparison to internal standards, where intensity differences between ^{15}N or ^{13}C labelled standards and unlabelled metabolites can be measured using MS while a reference metabolite is usually used for measurements on NMR. Metabolite concentration can also be determined by external comparison to metabolite standards of a concentration range. As matrix effects can interfere with quantitation, external calibration curves are usually made through the addition of standards into the extracted biological samples. Although isotopically labelled internal standards are frequently used for quantitative metabolomics, they are commercially unavailable for several metabolites. To overcome this, microbial cells can be fed with a labelled nutrient such as ^{13}C -glucose and the labelled metabolites can be used as internal standards (Castaño-Cerezo et al., 2019). However, correction for incomplete metabolite labelling is vital for proper quantitation (Bennett et al., 2008).

There are various issues with the quantitation of metabolites using LC-MS. To transform LC-MS data into metabolite abundances, peaks must be matched to metabolites, where there is a reliance on retention time and mass. Although there is specificity from both the retention time and mass, peak mis-annotation persists as a key issue. Peak mis-annotation could arise due to isomers, interferences of similar molecular weight, and in-source degradation products. LC and tandem MS can be utilized to separate isomeric metabolites (Koley et al., 2022), and if this is insufficient for isomer separation, chemical derivatization can be used instead followed by analysis either on the LC or GC (Struys et al., 2004; Gibson et al., 1993). In the case of interferences between analytes of similar molecular weight, the use of high-resolution MS can aid in reducing such interferences. In-source fragmentation decreases the signal of the metabolite ion and results in fragments that have the same molecular formula as the molecular ion of another metabolite (Purwaha et al., 2014). The development of ion sources which reduce in-source fragmentation could overcome this problem (Lu et al., 2016). Metabolite extracts from different groups of samples should be run in a randomized manner for quantitation in LC-MS as the absolute signal intensity can drift due to factors such as alterations in ionization efficiency. Another important issue that affects metabolite quantitation in LC-MS is ion suppression. The occurrence of ion suppression and mitigation measures have been discussed in the previous section.

For quantitation using the GC-MS, an area of concern is the loss of derivatized trimethylsilyl groups. The GC-MS is robust for the quantitation of phosphates, hydroxy acids, and sugars as *O*-trimethylsilyl groups are stable. However, as *N*-trimethylsilyl groups are less stable, during the analysis of trimethylsilylated amino acids and amines, there could be varying ratios of partially and fully derivatized metabolites across samples. To improve the precision and accuracy in the quantitation of metabolites, a few derivatizations can be executed, resulting in more than one analytical run. For instance, trimethylsilylation can be used for the quantitation of sugars, hydroxy acids, and phosphates, while tertiary butyldimethylsilylation can be used for amino acids and amines (Niehaus et al., 2014).

For metabolite quantitation using NMR, calibration of signals from experimental data should be executed by using standard mixtures prepared at known concentrations. Quantitative errors in NMR could arise due to problems in resolving individual peaks, whereby peak fitting and multi-dimensional NMR may be used to improve measurement precision (Mauve et al., 2016). NMR is useful for the quantitation of flux through metabolic networks due to its ability in determining position-specific isotope labelling in metabolites (Reardon et al., 2016). However, there is a major issue with using such isotope-based analyses for quantitation, as the ^1H - ^{13}C signal is artificially deflated due to signals from ^1H - ^{13}C relaxing more rapidly than those from ^1H - ^{12}C (Lewis et al., 2010).

Although, this quantitative pitfall can be corrected with isotopomer standards, such standards required for this calibration are often costly or unavailable.

GC-MS and LC-MS are complementary techniques, which can be utilized for the quantitation of different metabolites. In this section, various quantitative pitfalls and mitigation measures have been highlighted to improve the quantitation of metabolites using such analytical techniques. Executing *in vivo* measurements using NMR could potentially alleviate the numerous ways in which metabolite concentrations can be modified during quenching and extraction.

2.5. Dynamic metabolomics

Dynamic metabolomics involves the generation of time-series data from intracellular and extracellular metabolites. Appropriate methodologies (as discussed in the previous few sections) are required to quench metabolism, extract, and quantify intracellular and extracellular metabolites (Villas-Bôas et al., 2005b), at various time-points to generate dynamic data (Fig. 1). In previous studies, time-series data has been generated from bacteria (Chassagnole et al., 2002) and yeast (Rizzi et al., 1997; Theobald et al., 1997; Sasidharan et al., 2012). This *in vivo* dynamic data enables the determination of enzyme kinetics and creation of kinetic or dynamic models which help to explain metabolite dynamics (Chassagnole et al., 2002; Rizzi et al., 1997; Nikerel et al., 2006; Link et al., 2014).

To gain information about the changes in the metabolic state of microbial cells through dynamic time-series data is experimentally demanding. Therefore, research studies need to consider the duration of the measurements, number of experimental replicates and the time resolution in generating such data. There are also issues with extraction biases in metabolomics experiments (Sitnikov et al., 2016) as well as analytical and biological variances (Tabatabaei Anaraki et al., 2018). Furthermore, due to the need for rapid extraction of intracellular metabolites from microbial cells, large-scale real time monitoring of metabolite abundance is lacking. Improvements in high throughput metabolomics and real-time metabolomics, which will be discussed in the next two sections, will aid in data generation that can further enhance our understanding of metabolic networks in microbial cells.

2.6. High throughput metabolomics

Robotics can be applied to increase coverage in high throughput applications of strain selection through the automation of sample preparation and metabolite extraction (Yizhak et al., 2010; Vavricka et al., 2020). The use of liquid handling technologies also allows for good reproducibility of samples extracted and is therefore vital for high throughput sample preparation (Liu et al., 2019a). Cultivation devices such as microtiter plates or micro-bioreactors have also been integrated to robotic cultivation platforms, enabling growth parameters for cell cultures to be fully automated and controlled (Long et al., 2014; Faust et al., 2014; Unthan et al., 2015; Heux et al., 2014; Bergès et al., 2021).

Microfluidic platforms allow for enhanced analysis speed, diminished sample volumes, and improved multiplexing capabilities (Wang et al., 2015). Microfluidic platforms can mix small volumes of reagents and solutions through integrated channels and reaction chambers. This makes such devices suitable for studying the growth of cells. Many microfluidic applications for cell culture, including isolation, lysing, concentration, cell sorting, single cell analysis, and metabolite monitoring have been reported (Huang et al., 2008).

Electrochemical microsensors or LC-MS can be used to monitor metabolites from lab on chip devices. Various miniaturized separation techniques such as microchip-LC, microchip-CE, and micro-SPE have also been used together with MS (Gao et al., 2013; Ohla and Belder, 2012). Fluorescence and electrochemical detection are frequently used with micro-chip CE for metabolite analysis (Huang et al., 2008; García-Pérez et al., 2008; Kraly et al., 2009), for instance, adenosine

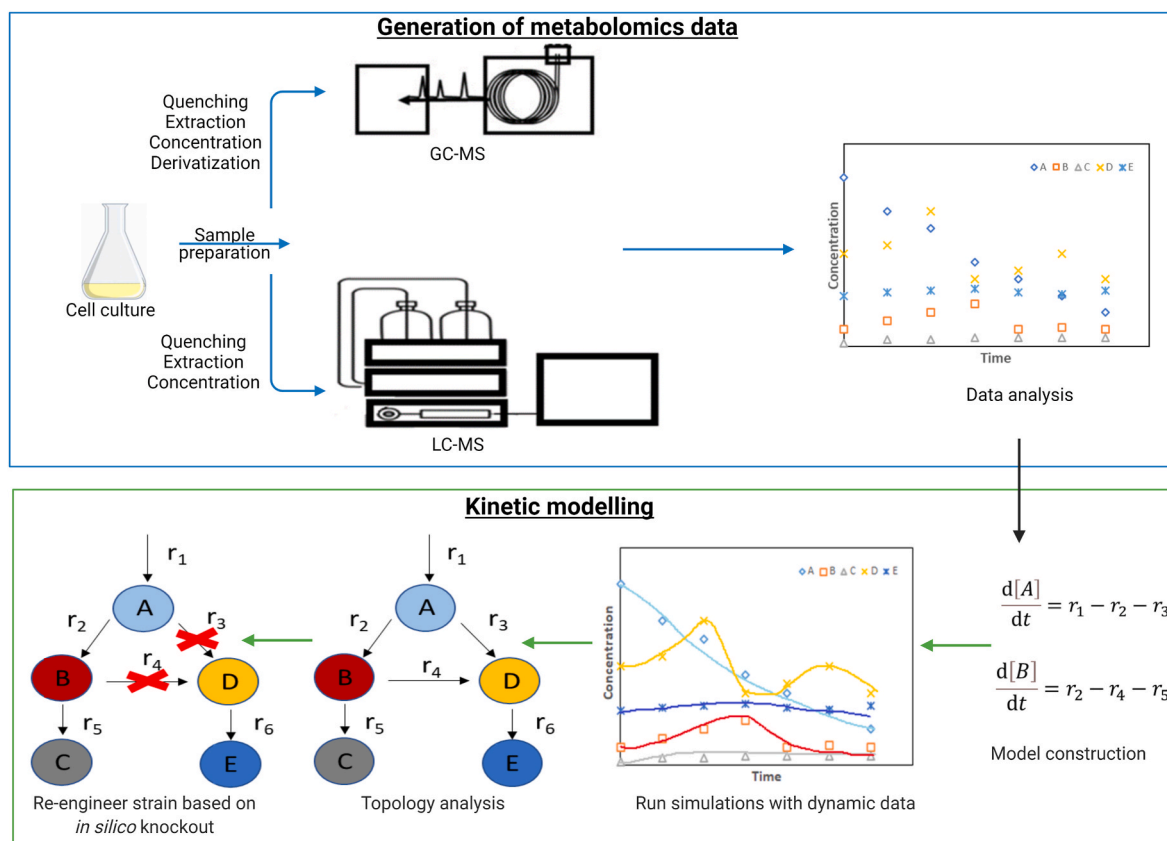


Fig. 1. Workflow of generating dynamic time-series data of metabolites followed by kinetic modelling of the network. The metabolic network topology is adjusted to fit the time-series metabolic data of metabolites A, B, C, D and E. *In silico* knockouts to increase the yield of desired metabolite C, for instance, can guide metabolic re-engineering of the microbial strain.

triphosphate in bacterial cells were lysed on a microchip-CE device prior to quantitation using bioluminescent detection (Liu et al., 2005). However, such devices are only useful in the detection of electroactive species and therefore limiting their potential for multiplexing, and the electrodes used in the fluidic system eventually degrade with time. Although LC-MS allows for the detection of many metabolites, aliquots need to be removed from the chip prior to analysis. As NMR spectroscopy is non-invasive, it is particularly suitable for the metabolomics observation of live systems, where quantitation of metabolites can be executed directly (Cox et al., 2019). Sensitivity becomes an issue though, due to the small sample volumes used in microfluidic systems. Microfluidics and micro-NMR have been combined to quantitatively investigate metabolism of cells in a single spheroid (Patra et al., 2021; Kalfe et al., 2015). An innovative platform utilizing microfluidics with LC-MS-NMR was also used for targeted and untargeted analysis of metabolites from a bacterial extract (Lin et al., 2008). The nanoSplitter enabled MS analysis, while the HPLC effluent was concentrated by microdroplet NMR, allowing for greater sample throughput efficiency than the traditional flow-injection methods.

Single cells can be individually analyzed through compartmentalization, allowing for the analysis of cell heterogeneity when culture conditions are altered. Isolation of single cells in microfluidics can be executed through cell confinement within microfluidic structures such as droplets, nano-liter wells and channels (Prakadan et al., 2017). Droplet microfluidics is one such platform that allows for the compartmentalization of single cells in cell culture. Typically for droplet microfluidics, pico-liter monodisperse aqueous droplets are created in an oil phase on a microfluidic device (Curran, 1998). In a recent study, metabolites from yeast were determined from microfluidic droplets using HPLC, where different incubation formats were found to affect cell metabolism (Bjork et al., 2015). The control of culture conditions is

important for the screening of single cells in droplets, as the production yields from cell factories are affected by the metabolic state of the cells. In another study, a microfluidic platform was developed to determine the consumption and secretion of metabolites, where a microfluidic droplet maker was utilized to trap yeast and bacterial cells and growth medium, enabling the identification of strains with the ability to over-produce or overconsume metabolites (Wang et al., 2014). This high-throughput approach could be utilized to identify mutations accountable for phenotypes which are technologically vital. There are some limitations with droplet microfluidics as assays used, such as enzymatic fluorescence, need to be compatible with the platform and metabolites can only be water soluble and immiscible in the oil phase for successful encapsulation (Wang et al., 2014). Furthermore, the number of metabolites that can be profiled using such single-cell platforms need to be increased, with sensitivity being the main bottleneck. Studying the metabolome of single individual cells is rather difficult due to the low quantities and large diversity of metabolites of the cellular metabolome and it is challenging to interface such samples with the MS (Zenobi, 2013). The handling of samples in single cell analysis requires a precise platform as minute losses in the usual cell population methods may result in large variations amongst individual cells (Liu et al., 2019b). Many individual cells are also required to be profiled concurrently to enhance the quality of data (Prakadan et al., 2017). Furthermore, the media of cell cultures contain various components which could result in matrix effect especially at the single cell level, requiring the further removal of matrix interference (Zhang et al., 2016).

Thus far, high-throughput metabolomics from cell cultures has not seen the coupling of microfluidic separation techniques with MS. There could be potential for high-throughput metabolomics based on previous work which shows the possibility of parallel sample preparations (Huft et al., 2013). Coupling of microfluidics to MS will aid in the development

of high-throughput analysis systems. Electrospray ionization (ESI) is frequently used for on-line microfluidic analysis due to its compatibility with low flow rates, while matrix assisted laser desorption/ionization (MALDI) has been used with microfluidic devices due to the possibility of automation and high-throughput sample preparation (Wang et al., 2015; Ibáñez Alfredo et al., 2013). Improvements in single cell microfluidics can also aid in understanding cell heterogeneity in cell cultures, as product yields from bio-factories are affected by the metabolic state of the cell and cells from mutant strains can be screened in a high-throughput manner to identify phenotypes which are important for enhanced product formation. Furthermore, large sample sizes can provide important insights into the dynamics of biological systems (Cannon et al., 2018; Sengupta et al., 2016). With advancements in microfluidics-MS, perhaps there would be improvements in throughput and sensitivity allowing for better metabolite coverage.

Another important consideration in the application of high-throughput metabolomics using microfluidics is the hydrodynamic, mass and heat transfer contexts that such apparatus operate under (Chhatre et al., 2011). Such environmental conditions directly impact the growth and viability of cells, and thus their metabolism. Cells in microfluidic devices are categorically growing in environments different from typical scales in laboratory flasks and bioreactors. Larger cultures at pilot (10^2 - 10^3 L) and commercial scales (10^3 - 10^5 L) are even more disparate in comparison (Flickinger, 2010; McConville and Kessler, 2010; Vivek et al., 2022). The metabolomics profiles and culture information acquired via microfluidics experimentation have to be understood within its respective environmental multi-physics considerations.

2.7. Real-time metabolomics

Real-time metabolomics allows for the simultaneous and high throughput analysis of metabolites from microbes without the need for sample preparation (Link et al., 2015). Real-time metabolome profiling has been executed in live bacteria and yeast through the direct injection of liquid cultures of single cells into a high-resolution MS which allowed for the automated monitoring of metabolites (Link et al., 2015). Through real-time profiling of metabolites, data modelling, and inhibitor experiments, the authors found switch-like feedback inhibition between starvation and growth in bacterial cells. In another study, live microbial colonies growing on agar were directly sampled through a surface sampling probe for ESI-MS which resulted in the immediate extraction and ionization of metabolites without the need for sample pre-treatment (Hsu et al., 2013). A microfluidic-based extraction system coupled to MS was also used for the continuous real-time analysis of metabolites from *Escherichia coli*, where cyclic patterns and metabolic trajectories could be predicted through the study of bacterial growth and stress (Heinemann et al., 2014). As the bacterial cells displayed predictable changes during stages of growth and stress, the authors suggested that future phenotypes could be implied through these trajectories. This approach could be useful when studying viable strains to produce targeted compounds. In another recent study, NMR was used to monitor central carbon and amino acid metabolism in microbe cells in real-time (Judge et al., 2019).

Such studies driving the generation of high-resolution real-time metabolite time-series data provide the real-time quantitation of changes in intracellular metabolites. This in turn could enhance conventional metabolomics assays for the prediction of phenotypes due to experimentally derived metabolic models (Link et al., 2014).

2.8. Spatial metabolomics

Spatial metabolomics involves the detection and interpretation of metabolites in the spatial context of cells and organisms (Petras et al., 2017), where the localization of metabolites is mainly determined through imaging MS (Doerr, 2018; Buchberger et al., 2018; Bodzon-Kulakowska and Suder, 2016). In conventional metabolomics, for

instance, the localization of metabolites within specific organelles is not known as extracted metabolites (intracellular and extracellular) are subjected to quantitation usually through the MS (Patti et al., 2012; Lu et al., 2017). Imaging MS allows for the mapping of MS data to two dimensions showing the chemical makeup of microbes (Watrous and Dorrestein, 2011) and can profile cells *in situ*, enabling the link of molecular information such as cell-cell interactions with spatial information. Sampling is executed systematically through the division of a sample into a grid of pixels (Alexandrov, 2020). Metabolites are then desorbed from every pixel in the grid through a laser, subsequently, a mass spectrum is produced displaying relative molecular intensities within the pixel (Fig. 2), imaging MS results in large datasets, with 10,000 to 1,000,000 spectra each relating to their own pixel (Nguyen et al., 2021). There are also other approaches in spatial metabolomics such as spectroscopy (Movasaghi et al., 2008; Langer et al., 2020) and molecular cartography (Protsyuk et al., 2018). As discussed in the previous section, MS is the most common technique for metabolomics (Alseekh et al., 2021) since it offers good sensitivity, specificity, and molecular coverage. This makes imaging MS the most preferred method for spatial metabolomics.

Metabolic imaging techniques can probe microbial metabolism allowing for better insight at the resolution of sub-cellular, single-cell and population level. Furthermore, there is also the possibility to determine changes in microbial metabolism in various host environments, and better understand dynamics and regulatory mechanisms of cell metabolism (Aldridge and Rhee, 2014). Regulation and cellular metabolism may be linked through high throughput imaging together with sensing techniques (Bellin et al., 2014). Traditional fluorescent imaging has also shown links between metabolism and the physiological state of bacteria (Parry et al., 2014). Metabolites or the general redox state of some metabolic processes are usually determined from metabolic imaging techniques. Such methods allow for time-lapse and measurements at single cell level while imaging MS sacrifices time-lapse and resolution at single cell level to determine metabolites at the systems level (Aldridge and Rhee, 2014). For instance, the multidimensional metabolic state of bacteria has been determined at single-cell and sub-cellular level using dynamic secondary ion mass spectrometry (SIMS) imaging MS (Watrous and Dorrestein, 2011), while MALDI-TOF (matrix assisted laser desorption/ionization-time of flight) and nano-DESI (desorption electrospray ionization) imaging MS have been utilized to study metabolites produced within microbial communities and the interactions between such communities (Traxler et al., 2013; Yang et al., 2009).

Many metabolites are structurally similar and related spatially as they are involved in similar metabolic networks. In a recent study, such chemical relations in networks were studied using MALDI imaging MS (Geier et al., 2020). Even though imaging MS provides insights to the spatial distribution of metabolites, to be fully convinced of the metabolite identification additional information such as tandem MS or molecular purification followed by structural elucidation with NMR would be beneficial (Nguyen et al., 2021). Imaging MS which is non-invasive can also be used to highlight cell heterogeneity over time. For instance, phenotypic heterogeneity was illustrated through the imaging of metabolite production, such as in clonal communities which were in the same microenvironment (Ackermann, 2015).

Even though spatial metabolomics using imaging MS allows for the localization of metabolites, it does suffer from some drawbacks compared to conventional metabolomics, where developments in the computational field could help to improve these gaps. Firstly, imaging mass spectra are highly complex as there is lack of analyte separation, such as separation through a chromatographic column, prior to MS analysis. This results in numerous peaks (which could potentially overlap) with background noise. The use of ion mobility spectrometry has the possibility to overcome this drawback as metabolites can be separated depending on their collisional cross section prior to imaging MS (Sans et al., 2018). Secondly, molecular identification possess a big

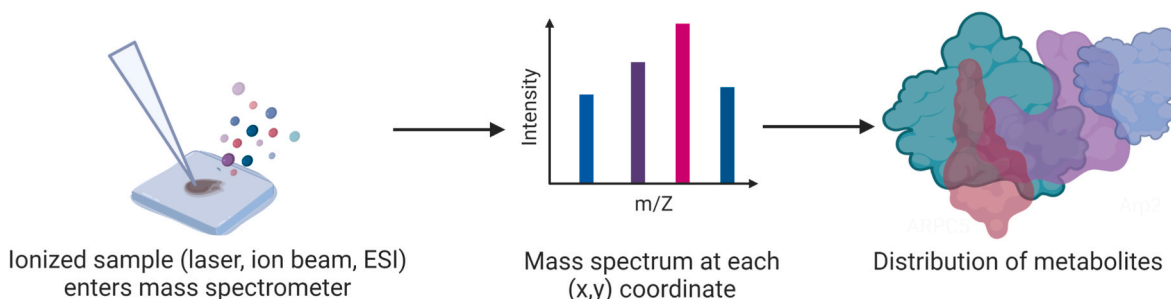


Fig. 2. Imaging MS to determine spatial localization of metabolites from a sample. In imaging MS, molecules are ionized from a sample spot, generating a mass spectrum. Image construction occurs through the mapping of intensities acquired from specific mass peaks at each spot across the sample.

challenge in spatial metabolomics involving imaging MS. For conventional metabolomics, tandem MS is often used, where analytes are fragmented and then measured in a mass spectrometer (Kim et al., 2018a). This MS/MS aids in the identification of molecules. Imaging MS does not use tandem MS, instead only mass-to-charge ratios are determined, with no fragmentation, which hampers molecular structure identification. Imaging MS is also currently unable to gather spatially resolved data-dependent tandem MS data. In imaging MS, the MS/MS spectra that can be gathered are only for a few selected ions for a few pixels (usually not using a true imaging mode) and are typically of poor quality due to the low quantity of metabolites present in one pixel (Alexandrov, 2020). Recently in attempts to improve metabolite identification from imaging MS data, an approach on metabolite to signal match score was used based on imaging MS data and theoretically derived properties of metabolites (Palmer et al., 2017; Alexandrov et al., 2019). Even though this approach outperformed mass-to-charge ratio matching, a large portion of imaging MS data was left unannotated as currently there is inadequate knowledge of signal formation of analytes and ionization pathways when using imaging MS.

Spatial single-cell metabolomics is an emerging area, where various imaging MS approaches have exhibited the capability to obtain data that can determine cell types and the related metabolic changes which occur with perturbations (Rappez et al., 2019; Gilmore et al., 2019; Do et al., 2017). Artificial intelligence and machine learning have been used in single-cell transcriptomics (Angerer et al., 2017; Hrovatin et al., 2022) and can potentially drive the analysis and interpretation of large spatial datasets obtained from single cells. As more insight is gained about metabolic networks in microbes, correlating metabolomics with spatial resolution through imaging MS and with single cell measurements through time will become increasingly important.

3. Modelling approaches

Constraint-based and dynamic-based modelling can be used to model the metabolomics data that has been generated experimentally. In doing so, researchers can gain distinct and complementary insights into intracellular metabolism (reactions) that is not obvious based on the concentrations of metabolites alone. Extracellular metabolite concentrations can be measured via analytical approaches (as discussed earlier) over various timepoints, and their rate of exchange with the modelled biological system can be evaluated as the change in their bioreactor concentration per unit time, further divided by the dry cell weight per unit bioreactor volume. The intracellular rate of substrate consumption and product formation (i.e., reaction) can be similarly evaluated. For constraint-based flux balance analysis of a stoichiometric metabolic model, the calculated rate of metabolite exchanges is used to constrain the model in deriving feasible flux solution. Whereas for kinetic modelling, the mathematical model for a reaction/exchange rate is fitted with both its assessed rate of reactions/exchanges as well as substrate concentrations to derive the model parameters. Fitted models are then simulated collectively to obtain the temporal profiles of metabolites,

which can be further tested in a perturbed condition, for instance, using a deleted enzyme.

Constraint-based modelling can be used in larger biological networks at the genome scale, as it just involves information on the stoichiometry of reactions in the metabolic network and mass balances for the metabolites under pseudo steady-state assumption (Oberhardt et al., 2009). Alternatively, dynamic-based modelling involves biochemical reactions which are described by ordinary or partial differential equations (Tomar and De, 2013). The application of dynamic models is, however, restricted to well characterized biochemical systems as many intracellular metabolite measurements are required experimentally for this modelling framework. Machine learning can be potentially used to overcome some of the problems faced by conventional constraint-based and dynamic-based models, as discussed in the following section.

3.1. Constraint-based modelling

Metabolic models have been used for metabolic engineering to better comprehend complex metabolic networks. Even though dynamic models have been used (Saa and Nielsen, 2017), such models are more suitable for small-scale modelling. For large-scale modelling approaches, for instance genome-scale models, it is rather difficult to integrate kinetic modelling as obtaining reliable kinetics and large data sets measured through experiments is challenging (Kim et al., 2018a). Hence, constraint-based modelling approaches have been used instead for large-scale modelling. Genome-scale models have been developed to determine metabolic activities and thus identify targets for gene manipulation (Park et al., 2009; Orth et al., 2010; Thiele and Palsson, 2010). This approach is useful in understanding slow growth rates and low product yield, that occurs after using a rational engineering approach at the systems level. Such stoichiometric constraint-based models have been used to direct metabolic engineering without requiring mechanistic detail (Kim et al., 2017). These models enable flux determination through optimizing the maximum rate of cell growth or product formation using mass balance equations with stoichiometric metabolic reactions occurring under steady state assumptions (Lee et al., 2012).

Flux balance analysis (FBA) is commonly used in constraint-based modelling (Orth et al., 2010). Represented as a linear programming problem, the method maximises biomass production as a cellular objective, subject to flux capacity, stoichiometric, and thermodynamic constraints. There have been various variations to these FBA models. In steady state regulatory FBA, the model is linked to the regulation network at a transcriptional level (Shlomi et al., 2007). Transcriptional regulation has also been integrated in a probabilistic approach in a genome-scale model (Chandrasekaran and Price, 2010). In another approach, reaction thermodynamics have been considered (Henry et al., 2007), while genomics and flux-converging patterns have also been considered in FBA (Park et al., 2010). In another study flux balance analysis with an extra constraint of occupancy in the membrane was introduced (Zhuang et al., 2011).

Metabolic flux analysis (MFA) can be used to ascertain the internal flux distribution of the network by measuring uptake and secretion rates and isotopic labelling information in metabolites (Stephanopoulos et al., 1998). Frequently, ^{13}C -MFA is used to determine *in vivo* metabolic fluxes (Antoniewicz, 2015). The most likely flux distribution in the network is determined from the difference between experimentally measured and the predicted isotopomer distributions amongst metabolites. Through the comparison of *in silico* simulation with isotopologue data, one can report which of the predicted metabolic fluxes is true and the difference from *in vivo* metabolomic data. ^{13}C -MFA can also be utilized as an inverse problem to determine fluxes which results in the best match of fluxes which were experimentally measured (Fondi and Liò, 2015). For instance, through ^{13}C -MFA and *in silico* flux based analysis, the metabolism of xylose in yeast cells was studied, where futile pathways and the relation between xylose utilization and energy maintenance of cells were discovered (Feng and Zhao, 2013). Large scale ^{13}C -MFA has also been executed in cyanobacteria (Hendry et al., 2019) and in another study, ^{13}C -MFA was used in parallel labelling experiments with ^{13}C -labelled glucose tracers in bacterial cultures (Long and Antoniewicz, 2019). A novel approach used ^{13}C -MFA in bacterial co-cultures to establish simultaneous metabolic flux distributions in several species and to estimate metabolite exchanges between species (Gebreselassie and Antoniewicz, 2015). When comparing MFA and FBA, the metabolic network used in MFA only involves carbon transitions whilst flux balance analysis involves carbon and non-carbon internal and transport reactions. An extension of MFA to the genome-scale and the capability to establish *in vivo* intracellular distributions of non-carbon atoms will be rather intriguing (Ravikirthi et al., 2011).

While the assumption of metabolite steady state (i.e., intracellular concentration remains unchanged with time) in FBA is satisfied in continuous cell culture conditions and during the exponential growth phase of the batch culture, it does not hold true for the entire duration of the batch culture and in the bioreactor setups such as fed-batch condition. Given the computational intractability of kinetic modelling for genome-scale metabolic network, significant efforts have been devoted to extending the FBA approach for dynamic conditions by presuming and modelling steady-state metabolism sequentially in time. Such approaches include dynamic FBA (dFBA) (Mahadevan et al., 2002) and dynamic MFA (dMFA) (Leighty and Antoniewicz, 2011). Similarly to FBA, they evaluate the growth rate at each modelled time point under constraints (stoichiometric, metabolite exchanges, etc.), but innovatively use it to derive the exchange rates for the subsequent time point by integrating mass balance equations relating metabolite consumption and production (Yasemi and Jolicoeur, 2021). For simulation to proceed smoothly, the growth rate may be fitted to a continuously differentiable function, such as the Monod model. Discrete transcriptional regulatory effects have been initially introduced as additional flux constraints for specified reactions at respective time points (Covert et al., 2001). Subsequently, they have been more deeply integrated into the model by formulating the optimization problem nonlinearly (Mahadevan et al., 2002). While it has allowed for more dynamic regulatory information to be included, the increased computational complexity also resulted in poorer scalability with the size of the metabolic network. In addition, the inclusion of non-equilibrium thermodynamic constraints, such as those of enthalpy, entropy, free energy, and energy conservation, have improved flux predictions (Qian and Beard, 2005). Furthermore, combining dFBA with ODEs and Boolean logic (Covert et al., 2008) or via integration of extracellular cues (Min Lee et al., 2008) have resulted in similar or more accurate predictions, compared to kinetic models of the same size, while retaining their relative scalability. More recently, dFBA has been applied to predict batch and chemostat time courses in terms of the activity and composition of a microbiome, consisting of hundreds of microbial species, whose genome-scale models have access to a common metabolite pool (Popp and Centler, 2020). The method thus allows for the rationalization of meta-omics profiles and the compositional dynamics.

Genome-scale models have also been used to guide strain development. For instance, cubebol production in *Saccharomyces cerevisiae* was enhanced through the identification of gene deletion targets by minimizing metabolic adjustments (Asadollahi et al., 2009). This approach of using simulations to identify gene knockouts was also used to improve production of amino acids L-threonine and L-valine in *E. coli* (Lee et al., 2007; Park Jin et al., 2007; Park et al., 2011), succinate (Lee et al., 2005; Wang et al., 2006), and malate (Moon et al., 2008). By linking growth rate with the rate of product formation, gene knockout targets were identified, which resulted in the improved production of 1,4-butanediol (Yim et al., 2011). The genome-scale model was also used to guide strategies to enhance lycopene production in *E. coli* through identifying gene amplification targets by using flux scanning built on enforced objective flux (Choi et al., 2010). Thermodynamic analysis and flux balance analysis were used to improve taxadiene biosynthesis as the native non-mevalonate pathway in *E. coli* was found to be more energetically favourable for taxadiene production (Meng et al., 2011). In another approach integrating -omics, ^{13}C metabolic flux analysis has been used to identify reactions causing bottlenecks in lysine formation (Becker et al., 2011). Even though these studies highlight the usefulness of genome-scale metabolic models in improving the phenotype of the overproduction strains, there are other issues that cannot be resolved when using such models. For instance, in a study on enhancing lycopene production, gene targets identified for improved phenotype using the stoichiometric model were found to mainly affect regulatory genes, whereby deleterious effects were detected when the gene deletion targets determined by resequencing were combined (Alper et al., 2005). In another study on identifying amplification targets for putrescine production, the authors observed false-positive targets probably due to transcriptional and translational regulation that could not be accounted for (Park et al., 2012). In improving vanillin production, some incorrect predictions were made which led the authors to believe that using only stoichiometric models can result in unrealistic flux predictions when kinetic constraints or regulatory interactions are not accounted for (Pharkya et al., 2004). In a study on xylitol production, the stoichiometric model could not account for the inhibitory effects of metabolites (Chin et al., 2009), while in another study, authors found a novel pathway for central carbon metabolites for *E. coli* mutant strains by incorporating both metabolomics and fluxomics after discovering incorrect predictions made using the stoichiometric model (Nakahigashi et al., 2009). These studies show that while constraint-based models can improve strain design, such models cannot capture regulatory and kinetic interactions which can be altered to enhance target compound production in strains.

Even though constraint-based metabolic models have been beneficial in improving the yields of target compounds, there are some drawbacks associated with such models as they are based on steady-state assumptions. Furthermore, these constraint-based models cannot determine the associations between enzyme expression, metabolic flux, metabolite concentrations, and regulation which is feasible with dynamic models (Saa and Nielsen, 2017).

3.2. Dynamic-based modelling

Dynamic-based or kinetic models are particularly suitable when studying dynamic effects in systems as they are more predictive (Tummler and Klipp, 2018). For instance, kinetic modelling has been used to study interferences in regulatory networks. Using a kinetic model of glycolysis, over 100 alleged allosteric interactions were tested based on dynamic data, where allosteric interactions that reversibly regulate the changes between glycolysis and gluconeogenesis were identified (Link et al., 2013). In another study, variable allosteric terms were used in Michaelis-Menten rate laws to fit data of fluxes, metabolite, and enzyme concentrations on a reaction-by-reaction basis (Hackett Sean et al., 2016).

As kinetic models portray systems in terms of kinetic parameters,

'bottom-up' models of metabolism have been created where during model parameters determination either predicted or measured enzyme kinetic properties are utilized (Strutz et al., 2019). However, there could be unrealistic model behaviour when using *in vitro* enzyme properties due to the lack of consideration of regulatory effects (Heijnen and Verheijen, 2013) and data varying in substantial amounts from actual *in vivo* experiments (Selvarajoo et al., 2009; Teusink et al., 2000; Magnus et al., 2006). It is also challenging to create a dynamic model using a bottom-up approach due to uncertainties associated with the kinetic properties of enzymes and associated reactions (Andreozzi et al., 2016; Schaber et al., 2009). Furthermore, it is rather difficult to determine enzyme reaction kinetics for all pathway or network reactions, rendering this approach more appropriate for small-scale models. In one approach Michaelis-Menten rate law approximations, which utilizes kinetic data, was used to substitute detailed rate laws for well-characterized reactions (Du et al., 2016). In another approach, structural kinetic modelling, local linear models were used in the investigation and statistical analysis of each point in the parameter space allowing for biochemical interpretation without requiring explicit knowledge on the kinetic models with enzyme kinetics and rate equations (Steuer et al., 2006). The intricate and non-linear nature of associations between metabolites can also affect the effectiveness of optimization algorithms used to determine model parameters in bottom-up approaches (Cuperlovic-Culf, 2018; Srinivasan et al., 2015).

In a 'top-down' modelling approach, dynamic metabolomics data is utilized to deduce the flux rates, kinetics, or metabolite concentrations by establishing causation and correlation networks between metabolites (Cuperlovic-Culf, 2018). The cause-effect associations amongst metabolites in a network is typically created using metabolomics data collected in a time series whilst the correlation network utilizes statistical and mathematical tools to understand the likely association between metabolites and enzymes in a network (Srinivasan et al., 2015). The generation of *in vivo* time-series metabolite data is crucial for determining enzyme kinetics intracellularly (Chassagnole et al., 2002; Nikerel et al., 2006). In order to generate this data, appropriate methodology needs to be applied for quenching metabolism, extracting and quantifying the intracellular metabolites (Villas-Bôas et al., 2005b). With this data in hand, kinetic parameters in dynamic models can then be deduced. Dynamic models of metabolism have been constructed for central metabolites using mechanistic rate equations and time-series metabolite data in *E. Coli* (Chassagnole et al., 2002) and in *S. cerevisiae* (Rizzi et al., 1997; Theobald et al., 1997). Time-series data of metabolites allow for the development of kinetic models which describe metabolite dynamics (Link et al., 2014). Fig. 1 explains kinetic modelling of the generated dynamic time-series data of metabolites. In another approach to reduce model complexity, linlog kinetics was applied to artificially generated time-series metabolite data resulting in a kinetic model with estimated parameters (Kresnowati et al., 2005). Linlog kinetics was also applied with dynamic metabolite data from bacteria to model concentrations and fluxes in the valine/leucine pathway (Magnus et al., 2006). Furthermore, convenience kinetics has also been used as an approximative kinetic rate format to reduce model complexity (Liebmeister and Klipp, 2006).

As with the bottom-up approach, the top-down approach also utilizes optimization algorithms to estimate model parameters, which has its limits due to possible non-linear associations between metabolites and the diverse range of kinetic parameters and concentrations (Cuperlovic-Culf, 2018). However, top-down approach has been successfully utilized with little parameter sensitivity in the analysis of cellular pathways using simple mass-action kinetic models or linear response reactions (Selvarajoo et al., 2009; Selvarajoo, 2011).

3.3. Machine learning

The limitations of current modelling approaches and the accumulation of vast amounts of multi-omics data raises the need to involve

innovative data analytic and mining methods, such as artificial intelligence (AI). AI is a branch of computer science that develops computer programs capable of performing tasks that usually require human intelligence. The AI programs follow pre-determined rules or search for pattern in the data so that it can make independent decisions (Helmy et al., 2020). Several research fields in the biomedical sciences are taking advantage of AI such as in accelerating the drug discovery through developing advanced analytics for large-scale data. Although AI is heavily utilized in the fields of genomics, signalling pathways prediction and analysis, protein-DNA binding, protein-protein interaction prediction and cancer diagnosis amongst several other applications (Alipanahi et al., 2015; Hui et al., 2013; Poplin et al., 2018), AI is not similarly employed in the systems metabolic engineering field. Although the idea of using AI is relatively old (Zelezniak et al., 2018), this promising technology is still under-utilized.

Machine learning (ML) is a branch of artificial intelligence concerned with the usage of computational algorithms to perform specified tasks through the process of (1) identifying useful patterns from some data (data mining), based on which, (2) inferences and then decisions are made on the course of actions, with minimal human intervention. This field has significantly and continuously improved in the last few years and proved very useful in analyzing large-scale data generated using analytical and experimental platforms. Recent reports show ML applications in the discovery of food identity markers (Erban et al., 2019), identifying weight loss biomarkers (Dias-Audibert et al., 2020) and farm animal metabolism (Ghaffari et al., 2019). In systems metabolic engineering, the integration of ML and systems biology has several promising applications including pathway discovery (Cuperlovic-Culf, 2018; Quest et al., 2010), identifying essential enzymes (Plaimas et al., 2008; Nandi et al., 2017), genome annotation (Yip et al., 2013), protein modelling (Billings et al., 2019; Yang et al., 2020) and engineering bio-economy strains (Costello and Martin, 2018). Applications of ML in the field have been reviewed in several recent articles (Choi et al., 2019; Helmy et al., 2020; Volk et al., 2020).

ML methods can be classified into two main categories, as illustrated in Fig. 3. The first is supervised learning, whereby the objective is to derive a model for predicting the output variables-of-interest based on some input variable values by learning from instances of input-output pairs, whereas for the second category, unsupervised learning is aimed at the exploration of variations and relationships among data variables. Supervised learning can be further divided into classification and regression methods, to facilitate their understanding and choice of application. Classification is used, when the output(s) to be inferred is/are categorical, i.e. discrete and few in numbers, whereas regression is more relevant for continuous outputs. However, some methods, such as artificial neural network and support vector machine, can be used for both classification and regression. Unsupervised learning can be further classified into either clustering algorithms for grouping data points (samples) based on inherent similarities, or association algorithms for uncovering hidden trends or rules among the variables of data points.

Chosen judiciously, ML methods are well suited for addressing specific challenges associated with high throughput multi-omics studies (Mirza et al., 2019). ML can also be utilized for processing MS spectra during analysis of metabolites, where supervised machine learning methods can aid in picking peaks, missing data imputation, and normalization (Liebal et al., 2020). Omics datasets generally have many more variables than samples and are thus saddled with the so-called 'curse of dimensionality'. The latter refers to the overfitting of explanatory model by too many variables, whereby the model appears predictive on training data, but are otherwise not generalizable to new observations. It is also challenging to identify meaningful variables from the large pool. In this regard, ML approaches can be applied to reduce the feature size by projecting the high dimensional data to a lower dimensional space (Hira and Gillies, 2015; Dekermanjian et al., 2022; Faquih et al., 2020). Some examples of such ML methods are non-negative matrix factorization (NMF), principal component analysis

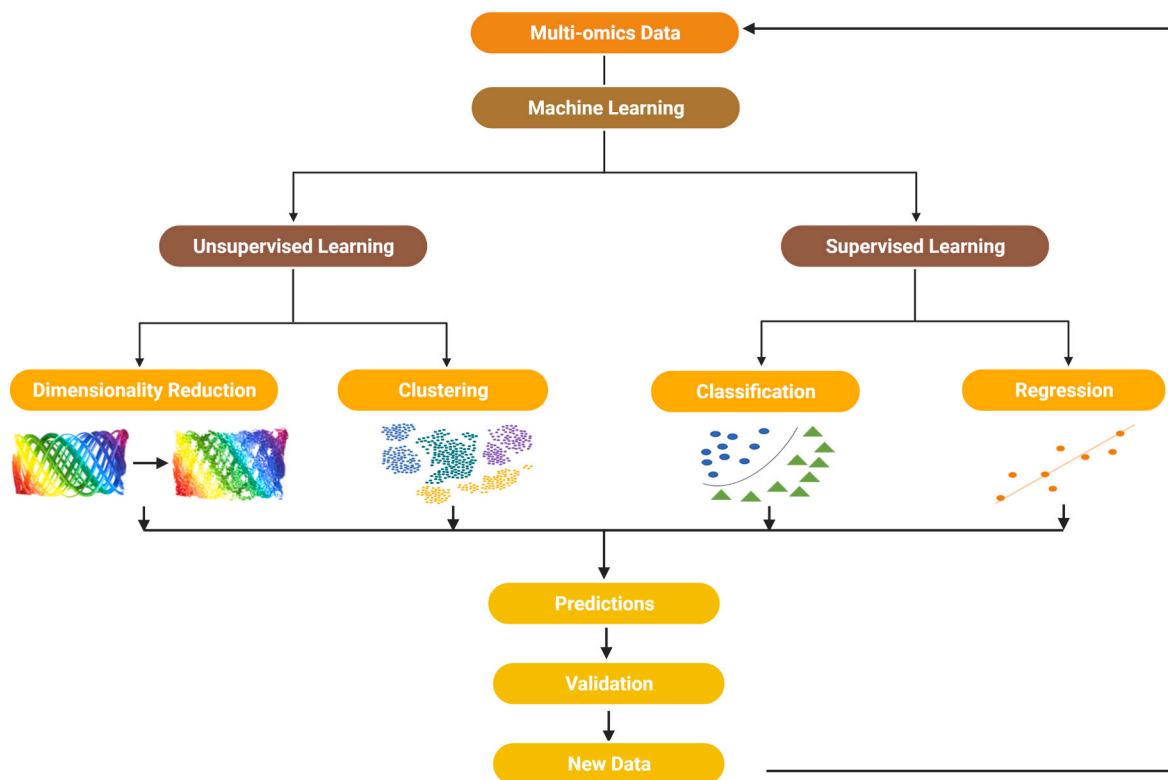


Fig. 3. Machine learning models to analyze and interpret multi-omics data. Validation of predictions is required after the application of the appropriate machine learning approach. The new data generated can be used to improve the learned model, enhance prediction performance, and aid in designing new experiments for data generation based on biological hypotheses.

(PCA), t-distributed stochastic neighbour embedding (t-SNE), and autoencoder (a type of ANN), with the last two especially suited for analysing non-linear data. To give some examples of their applications, PCA has been generalized as multi-omics factor analysis (MOFA) (Argelaguet et al., 2018), while NMF-based approaches, such as ‘intNMF’ (Chalise and Fridley, 2017), and ‘integrative-NMF’ (iNMF) (Yang and Michailidis, 2016) have been developed for identifying meaningful sample sub-groups. In addition, t-SNE is used for both clustering multi-omics samples (Lake et al., 2018), and the integration of single-cell transcriptomic datasets (Butler et al., 2018), whereas autoencoder has been used for identifying biological states that are predictive of phenotype (Ding et al., 2018), as well as for integrating domain knowledge (Ma and Zhang, 2019). ML can also aid in imputation of missing data for multi-omics datasets. Missing data is common with omics dataset for various reasons (Misra et al., 2018). Besides being replaced by the mean or median of other samples, missing values can be imputed with superior performance using regression, k-nearest neighbours (KNN) method (Kim et al., 2016), and singular value decomposition (SVD) (Linderman et al., 2022). Correlations have also been used to this end (Lin et al., 2016; Mias et al., 2016). In addition, maximum likelihood estimates (Allison, 2012) are widely used for estimating parameter values in linear models. Also, random forest (Shah et al., 2014; Kokla et al., 2019), and deep learning (Gondara and Wang, 2017) based approaches have been found to be highly effective for non-linear data.

Even though ML requires abundant high-quality data, this may not be sufficient as appropriate experimental design is essential to capitalize on ML (Carbonell et al., 2019). Thus, experimental designs need to consider various factors affecting the response for the generation of high-quality training data for ML algorithms. Complementarily, machine learning can also be utilized to design the next set of experiments to enhance the quality of experimental data, reduce errors in approximations and to test hypotheses.

It is important to recognise that the data-driven approach of ML and such related computational methods are categorically different from the constraint-based or dynamic-based approaches described above. Each of the three modelling approaches has its advantages and epistemological limitations, therefore one has to appropriately choose suitable numerical approach and algorithms for unraveling metabolomics information from corresponding experimental data. Both the constraint-based and dynamic-based methods formulate their numerical modelling from some theoretical basis, may it be from metabolic network, reaction thermodynamics or cell growth kinetics considerations, and could explain metabolomics findings against such. One of the common numerical issue faced, especially with large models having many parameters to optimize, is overfitting. It is particularly pronounced if such model equations translates into sparse matrices that are ill-conditioned. For ML and its like, patterns within the experimental data are discernible through chosen applied algorithms. Depending on the intrinsic variations within the data, and in concert with the algorithm(s) utilized, numerical conclusions are inferred. Such inferences, largely arise from either statistical correlations or pattern dependencies in the data, thus their causal or theoretical link to metabolic network, reaction thermodynamics or cell growth kinetics has to be further investigated and corroborated.

3.4. Machine learning in systems metabolic engineering

3.4.1. Model reconstruction

ML has greatly assisted the reconstruction of genome-scale metabolic models. For example, unsupervised learning methods have been commonly used to infer metabolic networks from omics datasets, sequence information, and enzyme-enzyme associations (Vert et al., 2007; Yamanishi et al., 2005). In another study, Naïve Bayes, decision trees, and logistic regression were found to perform as well (Dale et al., 2010) as the widely-used PathoLogic algorithm (Karp et al., 2011) in predicting metabolic pathways from genomic sequences. Importantly,

these methods provide statistical confidence for predictions that was not available using PathoLogic. In addition, of interest to enzyme constraint-based modelling, enzyme turnover numbers were predicted from biochemical, structural, and network data, besides assay condition, by using an ensemble model comprising of random forest, deep neural network, and linear elastic net (Heckmann et al., 2018).

3.4.2. Machine learning-based modelling

The development of novel modelling approaches have also benefited from the application of ML methods. In a study on the kinetic modelling of an *E. coli* system producing 1,4-butanediol, decision tree algorithm was used to determine enzyme parameters that worked within narrow ranges of values. The identified enzymes were presumed, and later found, to strongly influence the production of the desired metabolite, which is consistent with previous studies (Andreozzi et al., 2016). In another innovative study, an automated ensemble approach and genetic algorithm were used to predict pathway dynamics from proteomics and metabolomics time-series data (Costello and Martin, 2018). Importantly, the approach is distinct from conventional kinetic modelling, in that it does not assume interactions, and instead, implicitly chooses the most predictive ones. In another application, a deep neural network model was structurally designed to imitate the biological hierarchy of *S. cerevisiae* based on gene ontologies, thus allowing it to predict the phenotype outputs and in-between molecular mechanisms from genotype inputs and their perturbations (Ma et al., 2018).

3.4.3. Augmentation of modelling tools

ML has also been used for facilitating and improving flux predictions, as well as enhancing the utility of metabolic modelling tools (Zampieri et al., 2019). For example, mechanistic model-assisted ML has been particularly useful for overcoming the black-box limitation of ML on one hand, and the lack of straightforwardness for mechanistic modelling by itself to suggest effective leads for hypothesis generation. In this regard, a white-box ML approach has been developed by combining mechanistic modelling with a two-stage Elastic net regularization, which results in its effective prediction of the modes of action (Yang et al., 2019a).

3.4.4. Pathway/strain design

AI/ML are also used for designing both *de novo* and retrosynthetic pathways (Lee et al., 2019; Lin et al., 2019). A semi-supervised Gaussian process was also utilized to compute the probability of enzyme candidates catalysing a specified reaction. If no enzyme is found, similar approaches may be used to identify the required protein sequences and structures for catalysing the reaction (AlQuraishi, 2019; Yang et al., 2019b). Alternatively, the techniques can be used to guide the necessary directed evolution of proteins (Biswas et al., 2018). For example, linear regression, neural network, and ensemble methods had been used to generate a nitric oxide dioxygenase protein, which is capable of synthesizing both enantiomers of the product (Wu et al., 2019). Other than starting from an end-product perspective, researchers may also be interested in determining the possibilities of metabolite pairs transforming into each other, and the pathway required to do so. In this regard, an ANN-based approach has been developed (Tongman et al., 2017). Further reviews of ML applications for *de novo* pathway design are available elsewhere (Jeffryes et al., 2018; Kotera and Goto, 2016). To assist existing computational strain design strategy, an ensemble of machine learners were trained to predict production titres based on both process conditions (e.g., substrate levels and bioreactor type) and central metabolic fluxes inferred from the FBA of genome-scale models (Czajka et al., 2021). The ML model was tested on related strains, which verified its utility for transfer learning to similar hosts.

3.4.5. Pathway engineering

In a proteomics study of the mevalonate pathway, principal component analysis (PCA) was used to identify the over-expressions of specific enzymes for improving the production titres of two terpenes

(Alonso-Gutierrez et al., 2015). The recommendations were then validated by a production increase of 40%. In two other studies, linear (Lee et al., 2013) and quadratic (Xu et al., 2017) regression models were used for predicting the required gene expressions of a five-genes pathway, so as to enhance the production of violacein in *S. cerevisiae* and *E. coli* cultures, respectively. With the findings, partial-least-square regression (PLSR) (Jervis et al., 2019) and artificial neural network (ANN) (Meng et al., 2013) were then utilized for optimizing the required promoter strengths and other regulatory elements. In a separate study, the combinatorial space for the optimal combination of vector design parameters was reduced using linear regression (Carbonell et al., 2018). Similarly, AI/ML methods have been used for the design of CRISPR/Cas tools that can be deployed for gene manipulation in metabolic engineering. For example, the computational platform, 'DeepCRISPR', uses deep unsupervised representation learning to predict the efficacy of targeted knockout and off-target sites of single-guide RNAs (sgRNA) (Chuai et al., 2018), whereas 'sgRNA Scorer' (Chari et al., 2017) and 'DeepCpf1' (Kim et al., 2018b) employ support vector machine (SVM) classifier and convolutional neural network, respectively, for the same purpose. More recently, an ensemble ML approach has also been developed to optimize inputs, such as biomolecules, promoter constructs, and nutrient levels, for synthetic biology and metabolic engineering applications (Radivojević et al., 2020).

3.4.6. Bioprocess

ML has been directly applied to optimize bioprocessing conditions. For example, least-squares regression was used to find the best-performing carbon source, induction temperature, and inoculation ratio (Jones et al., 2016) for an *E. coli* co-culture, resulting in a 970-fold increase in flavan-3-ol titre over a monoculture. In another two studies with similar objectives, artificial neural network (ANN) was used to optimize bioreactor temperature, pH, and oxygen provision for xylitol production (Pappu and Gummadi, 2016), and also to screen for media components that would maximize the microbial degradation of polycyclic aromatic hydrocarbons (Gosai et al., 2018). ANN was further used in conjunction with a genetic algorithm to determine the best nutrient composition for producing potential anti-cancer compounds in *S. cerevisiae* (Zheng et al., 2017). In a different study, the focus is on predicting the required concentration of *S. cerevisiae* in a bioreactor, by simply using substrate flow rate as input (Masampally et al., 2018). To be able to do so, three sequential models, using Gaussian process regression based on Bayes' rules, were applied to infer important intermediate variables, such as the gas hold-up, and the concentrations of biomass and dissolved oxygen, which could not be measured in real time. In another broadly useful study (Li et al., 2019), support vector regression (SVR) was found to best infer the optimal growth temperature of microorganisms, with the parameter being newly obtained for 1438 prokaryotes and eukaryotes.

3.5. Integrating modelling with the Design-Build-Test-Learn cycle

The Design-Build-Test-Learn cycle has been used in systems metabolic engineering. In this approach, a microorganism strain is first chosen either based on its natural ability to produce the desired product or through genetic engineering techniques (Zhang et al., 2018b). Usually, there are many factors, such as regulatory feedback mechanisms, complex translation control mechanisms and competing metabolic pathways, that result which cause suboptimal production of target products (Selvarajoo, 2021). Next, metabolomics data can be collected and used to understand metabolic fluxes and build predictive models. Computational models which predict the behaviour of biological systems can be experimentally verified and can iteratively guide the next set of experiments (Helmy et al., 2020). ML models generally result in faster and enhanced performance for pathway predictions as compared to parametric and non-parametric computational models (Quest et al., 2010). ML models can also be mapped between proteomics and

metabolomics using training datasets followed by experimental verification and validation of predictions based on test data (Costello and Martin, 2018).

After computational modelling of omics datasets using machine learning approaches, strains can be re-engineered according to new predictions based on further perturbations or identified gene targets which would improve product yield. For instance, in a study integrating metabolomics, proteomics, genome-scale models, and metabolic engineering of bacterial strains, several genes and metabolic reactions were identified which resulted in re-engineering of strains which produced higher product yields (Brunk et al., 2016). The optimized strains obtained using the iterative Design-Build-Test-Learn cycle can then be tested at various bioreactor scales for industrial purposes.

4. Conclusion: future challenges and directions

There are various challenges faced in the development of metabolomics approaches for metabolic systems engineering. These include metabolite coverage, automation, and throughput. Increasing the coverage of metabolites can compensate the restricted throughput of LC-MS methods whereas the use of robotic liquid handler platforms and microfluidics allow for high throughput analysis of metabolites as discussed in the previous section. However, microfluidic approaches still lack the sensitivity to quantitate metabolites of low amounts. Further developments in this area would drive the production of large datasets of high-quality data that could be coupled to ML algorithms which can enhance the understanding of biological networks to improve the production of a phenotype during strain development. Transfer learning, an area of ML, could be used to make inferences and predictions on a biological system based on observations made in other biological systems (Camacho et al., 2018). This would reduce the need in generating data from many different systems as various biological systems share similar characteristics. However, the challenge does arise on how best to apply the knowledge learnt in a particular system to a novel system for which data is limited.

ML has also been utilized to execute automated metabolite identification in conventional metabolomics (da Silva et al., 2015; Monge et al., 2019). Such ML approaches can possibly be applied to aid the current bottleneck of spatial metabolite identification in imaging MS. The use of ion mobility spectrometry (IMS) as a separation technique in imaging MS is another possibility of improving metabolite identification. ML has been used successfully to predict the collisional cross section of metabolites separated in IMS (Zhou et al., 2016) and therefore ML can potentially be applied to IMS coupled to imaging MS to drive metabolite identification. Furthermore, ML approaches have also been used in pathway engineering of microbes. For instance, an ensemble ML approach has been developed to optimize inputs, such as biomolecules, promoter constructs, and nutrient levels, for synthetic biology and metabolic engineering applications (Radivojević et al., 2020). AI/ML are also used for designing both *de novo* and retrosynthetic pathways (Lee et al., 2019; Lin et al., 2019).

The prospect of combining real-time metabolomics data together with the spatial distribution of metabolites to better understand the complexity of biology networks is intriguing. Further advancements in single-cell metabolomics could also enhance our understanding of cell-cell heterogeneity and the complexity of their interactions, which would be beneficial in future systems metabolic engineering applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the Intra-create Thematic Grant “Cities” (grant number: NRF2019-THE001-0007).

References

- Ackermann, M., 2015. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* 13 (8), 497–508.
- Aldridge, B.B., Rhee, K.Y., 2014. Microbial metabolomics: innovation, application, insight. *Curr. Opin. Microbiol.* 19, 90–96.
- Alexandrov, T., 2020. Spatial metabolomics and imaging mass spectrometry in the age of artificial intelligence. *Annual review of biomedical data science* 3, 61–87.
- Alexandrov, T., Ovchinnikova, K., Palmer, A., Kovalev, V., Tarasov, A., Stuart, L., Nigmatjanov, R., Fay, D., 2019. METASPACE: A Community-Populated Knowledge Base of Spatial Metabolomes in Health and Disease.
- Alipanahi, B., Delong, A., Weirauch, M.T., Frey, B.J., 2015. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33 (8), 831–838.
- Allison, P.D., 2012. In *Handling Missing Data by Maximum Likelihood*.
- Alonso-Gutierrez, J., Kim, E.-M., Batth, T.S., Cho, N., Hu, Q., Chan, L.J.G., Petzold, C.J., Hillson, N.J., Adams, P.D., Keasling, J.D., Garcia Martin, H., Lee, T.S., 2015. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metab. Eng.* 28, 123–133.
- Alper, H., Miyaoku, K., Stephanopoulos, G., 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.* 23 (5), 612–616.
- AlQuraishi, M., 2019. End-to-End differentiable learning of protein structure. *Cell systems* 8 (4), 292–301 e3.
- Alesekh, S., Aharoni, A., Brotman, Y., Contrepois, K., D’Auria, J., Ewald, J., J, C.E., Fraser, P.D., Giavalisco, P., Hall, R.D., Heinemann, M., Link, H., Luo, J., Neumann, S., Nielsen, J., Perez de Souza, L., Saito, K., Sauer, U., Schroeder, F.C., Schuster, S., Siuzdak, G., Skirycz, A., Sumner, L.W., Snyder, M.P., Tang, H., Tohge, T., Wang, Y., Wen, W., Wu, S., Xu, G., Zamboni, N., Fernie, A.R., 2021. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat. Methods* 18 (7), 747–756.
- Amantonico, A., Urban, P.L., Zenobi, R., 2010. Analytical techniques for single-cell metabolomics: state of the art and trends. *Anal. Bioanal. Chem.* 398 (6), 2493–2504.
- Andreozzi, S., Miskovic, L., Hatzimanikatis, V., 2016. iSCHRUNK – in silico approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metab. Eng.* 33, 158–168.
- Angerer, P., Simon, L., Tritschler, S., Wolf, F.A., Fischer, D., Theis, F.J., 2017. Single cells make big data: new challenges and opportunities in transcriptomics. *Curr. Opin. Struct. Biol.* 4, 85–91.
- Annesley, T.M., 2003. Ion suppression in mass spectrometry. *Clin. Chem.* 49 (7), 1041–1044.
- Antignac, J.-P., de Wasch, K., Monteau, F., De Brabander, H., Andre, F., Le Bizec, B., 2005. The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. *Anal. Chim. Acta* 529 (1), 129–136.
- Antoniewicz, M.R., 2015. Methods and advances in metabolic flux analysis: a mini-review. *J. Ind. Microbiol. Biotechnol.* 42 (3), 317–325.
- Argelaguet, R., Velten, B., Arnold, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., Stegle, O., 2018. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* 14 (6), e8124.
- Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. *Metab. Eng.* 11 (6), 328–334.
- Barbas, C., Moraes, E.P., Villaseñor, A., 2011. Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples. *J. Pharmaceut. Biomed. Anal.* 55 (4), 823–831.
- Bastawrous, M., Jenne, A., Tabatabaei Anaraki, M., Simpson, A.J., 2018. In-vivo NMR spectroscopy: a powerful and complementary tool for understanding environmental toxicity. *Metabolites* 8 (2).
- Becker, J., Zelder, O., Häfner, S., Schröder, H., Wittmann, C., 2011. From zero to hero—design-based systems metabolic engineering of *Corynebacterium glutamicum* for l-lysine production. *Metab. Eng.* 13 (2), 159–168.
- Behrends, V., Williams, H.D., Bundy, J.G., 2014. Metabolic footprinting: extracellular metabolomic analysis. *Methods Mol. Biol.* 1149, 281–292.
- Bellin, D.L., Sakhtah, H., Rosenstein, J.K., Levine, P.M., Thimot, J., Emmett, K., Dietrich, L.E.P., Shepard, K.L., 2014. Integrated circuit-based electrochemical sensor for spatially resolved detection of redox-active metabolites in biofilms. *Nat. Commun.* 5 (1), 3256.
- Bennett, B.D., Yuan, J., Kimball, E.H., Rabinowitz, J.D., 2008. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat. Protoc.* 3 (8), 1299–1311.
- Berges, C., Cahoreau, E., Millard, P., Enjalbert, B., Dinclaux, M., Heuillet, M., Kulyk, H., Gales, L., Butin, N., Chazalviel, M., Palama, T., Guionnet, M., Sokol, S., Peyriga, L.,

- Bellvert, F., Heux, S., Portais, J.-C., 2021. Exploring the glucose fluxotype of the *E. coli* y-ome using high-resolution fluxomics. *Metabolites* 11 (5).
- Billings, W.M., Hedelius, B., Millicam, T., Wingate, D., Corte, D.D., 2019. ProSPR: democratized implementation of alphafold protein distance prediction network. [bioRxiv, 330273](https://doi.org/10.1101/330273).
- Biswas, S., Kuznetsov, G., Ogden, P.J., Conway, N.J., Adams, R.P., Church, G.M., 2018. Toward machine-guided design of proteins. *bioRxiv*, 337154.
- Bjork, S.M., Sjostrom, S.L., Andersson-Svahn, H., Joensson, H.N., 2015. **Metabolite Profiling of Microfluidic Cell Culture Conditions for Droplet Based Screening** *Biomicrofluidics* [Online], 044128. <https://doi.org/10.1063/1.4929520>. PubMed.
- Bodzon-Kulakowska, A., Suder, P., 2016. Imaging mass spectrometry: instrumentation, applications, and combination with other visualization techniques. *Mass Spectrom. Rev.* 35 (1), 147–169.
- Brunk, E., George, Kevin W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., Petzold, Christopher J., McCloskey, D., Monk, J., Yang, L., O'Brien, Edward J., Bath, Tanveer S., Martin, Hector G., Feist, A., Adams, Paul D., Keasling, Jay D., Palsson, Bernhard O., Lee, Taek S., 2016. Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow. *Cell systems* 2 (5), 335–346.
- Buchberger, A.R., DeLaney, K., Johnson, J., Li, L., 2018. Mass spectrometry imaging: a review of emerging advancements and future insights. *Anal. Chem.* 90 (1), 240–265.
- Butler, A., Hoffman, P., Smibert, P., Papalex, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36 (5), 411–420.
- Cai, Y., Weng, K., Guo, Y., Peng, J., Zhu, Z.-J., 2015. An integrated targeted metabolomic platform for high-throughput metabolite profiling and automated data processing. *Metabolomics* 11 (6), 1575–1586.
- Cajka, T., Fiehn, O., 2016. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal. Chem.* 88 (1), 524–545.
- Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., Collins, J.J., 2018. Next-generation machine learning for biological networks. *Cell* 173 (7), 1581–1592.
- Cannon, W.R., Zucker, J.D., Baxter, D.J., Kumar, N., Baker, S.E., Hurley, J.M., Dunlap, J. C., 2018. Prediction of metabolite concentrations, rate constants and post-translational regulation using maximum entropy-based simulations with application to central metabolism of *Neurospora crassa*. *Processes* 6 (6), 63.
- Carbonell, P., Jervis, A.J., Robinson, C.J., Yan, C., Dunstan, M., Swainston, N., Vinaixa, M., Hollywood, K.A., Currin, A., Rattray, N.J.W., Taylor, S., Spiess, R., Sung, R., Williams, A.R., Fellows, D., Stanford, N.J., Mulherin, P., Le Feuvre, R., Barran, P., Goodacre, R., Turner, N.J., Goble, C., Chen, G.G., Kell, D.B., Micklefield, J., Breitling, R., Takano, E., Faulon, J.L., Scrutton, N.S., 2018. An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Communications biology* 1, 66.
- Carbonell, P., Radivojevic, T., García Martín, H., 2019. Opportunities at the intersection of synthetic biology, machine learning, and automation. *ACS Synth. Biol.* 8 (7), 1474–1477.
- Cascante, M., Marin, S., 2008. Metabolomics and fluxomics approaches. *Essays Biochem.* 45, 67–81.
- Castano-Cerezo, S., Kulyk-Barbier, H., Millard, P., Portais, J.C., Heux, S., Truan, G., Bellvert, F., 2019. Functional analysis of isoprenoid precursors biosynthesis by quantitative metabolomics and isotopologue profiling. *Metabolomics* 15 (9), 115.
- Chalise, P., Fridley, B.L., 2017. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One* 12 (5), e0176278.
- Chandrasekaran, S., Price, N.D., 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* 107 (41), 17845–17850.
- Chari, R., Yeo, N.C., Chavez, A., Church, G.M., 2017. sgRNA scorer 2.0: a species-independent model to predict CRISPR/cas9 activity. *ACS Synth. Biol.* 6 (5), 902–904.
- Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* 79 (1), 53–73.
- Chhatre, S., Titchener-Hooker, N.J., 2011. 2.65 - micro-biochemical engineering: using small-scale devices to predict industry-scale downstream performance. In: Moo-Young, M. (Ed.), *Comprehensive Biotechnology*, second ed. Academic Press, Burlington, pp. 891–903.
- Chin, J.W., Khankal, R., Monroe, C.A., Maranas, C.D., Cirino, P.C., 2009. Analysis of NADPH supply during xylitol production by engineered *Escherichia coli*. *Biotechnol. Bioeng.* 102 (1), 209–220.
- Choi, H.S., Lee, S.Y., Kim, T.Y., Woo, H.M., 2010. In silico identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* 76 (10), 3097–3105.
- Choi, K.R., Jang, W.D., Yang, D., Cho, J.S., Park, D., Lee, S.Y., 2019. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol.* 37 (8), 817–837.
- Chu, D.B., Troyer, C., Mairinger, T., Ortmayr, K., Neubauer, S., Koellensperger, G., Hann, S., 2015. Isotopologue analysis of sugar phosphates in yeast cell extracts by gas chromatography chemical ionization time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* 407 (10), 2865–2875.
- Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B., Gu, F., Qu, S., Huang, D., Wei, J., Liu, Q., 2018. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol.* 19 (1), 80.
- Contrepois, K., Jiang, L., Snyder, M., 2015. Optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by combining hydrophilic interaction (HILIC) and reverse-phase liquid chromatography (RPLC)-Mass spectrometry. *Mol. Cell. Proteomics*: MCP 14 (6), 1684–1695.
- Costello, Z., Martin, H.G., 2018. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ systems biology and applications* 4, 19.
- Covert, M.W., Schilling, C.H., Palsson, B., 2001. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213 (1), 73–88.
- Covert, M.W., Xiao, N., Chen, T.J., Karr, J.R., 2008. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics* 24 (18), 2044–2050.
- Cox, N., Kuemmerle, R., Millard, P., Cahoreau, E., François, J.-M., Parrou, J.-L., Lippens, G., 2019. Integrated pH measurement during reaction monitoring with dual-reception 1H–31P NMR spectroscopy. *Anal. Chem.* 91 (6), 3959–3963.
- Cuperlovic-Culf, M., 2018. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8 (1).
- Curran, D.P., 1998. Strategy-level separations in organic synthesis: from planning to practice. *Angew. Chem. Int. Ed.* 37 (9), 1174–1196.
- Czajka, J.J., Kambhampati, S., Tang, Y.J., Wang, Y., Allen, D.K., 2020. Application of stable isotope tracing to elucidate metabolic dynamics during *Yarrowia lipolytica* α -ionone fermentation. *iScience* 23 (2), 100854.
- Czajka, J.J., Oyetunde, T., Tang, Y.J., 2021. Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction. *Metab. Eng.* 67, 227–236.
- da Luz, J.A., Hans, E., Zeng, A.-P., 2014. Automated fast filtration and on-filter quenching improve the intracellular metabolite analysis of microorganisms. *Eng. Life Sci.* 14 (2), 135–142.
- da Silva, R.R., Dorrestein, P.C., Quinn, R.A., 2015. Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. U.S.A.* 112 (41), 12549–12550.
- Dale, J.M., Popescu, L., Karp, P.D., 2010. Machine learning methods for metabolic pathway prediction. *BMC Bioinf.* 11, 15–15.
- Dekermanjian, J.P., Shaddox, E., Nandy, D., Ghosh, D., Kechris, K., 2022. Mechanism-aware imputation: a two-step approach in handling missing values in metabolomics. *BMC Bioinf.* 23 (1), 179.
- Dias-Audibert, F.L., Navarro, L.C., de Oliveira, D.N., Delafiori, J., Melo, C., Guerreiro, T. M., Rosa, F.T., Petenuci, D.L., Watanabe, M.A.E., Velloso, L.A., Rocha, A.R., Catharino, R.R., 2020. Combining machine learning and metabolomics to identify weight gain biomarkers. *Front. Bioeng. Biotechnol.* 8, 6.
- Ding, M.Q., Chen, L., Cooper, G.F., Young, J.D., Lu, X., 2018. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. *Mol. Cancer Res.* 16 (2), 269–278.
- Do, T.D., Comi, T.J., Dunham, S.J., Rubakhin, S.S., Swedler, J.V., 2017. Single cell profiling using ionic liquid matrix-enhanced secondary ion mass spectrometry for neuronal cell type differentiation. *Anal. Chem.* 89 (5), 3078–3086.
- Doerr, A., 2018. Mass spectrometry imaging takes off. *Nat. Methods* 15 (1), 32–32.
- Doran, M.L., Myktyczuk, N., Bieniek, A., Methé, A., Merritt, T.J.S., 2017. Evaluation of quenching and extraction procedures for performing metabolomics in *Acidithiobacillus ferrooxidans*. *Metabolomics* 13 (12), 156.
- Du, B., Zielinski, D.C., Kavvas, E.S., Dräger, A., Tan, J., Zhang, Z., Ruggiero, K.E., Arzumanyan, G.A., Palsson, B.O., 2016. Evaluation of rate law approximations in bottom-up kinetic models of metabolism. *BMC Syst. Biol.* 10 (1), 40.
- Erbani, A., Fehrlé, I., Martínez-Seidel, F., Brigante, F., Más, A.L., Baroni, V., Wunderlin, D., Kopka, J., 2019. Discovery of food identity markers by metabolomics and machine learning technology. *Sci. Rep.* 9 (1), 9697.
- Fajjes, M., Mars, A.E., Smid, E.J., 2007. Comparison of quenching and extraction methodologies for metabolome analysis of *Lactobacillus plantarum*. *Microb. Cell Factories* 6, 27.
- Fauihi, T., van Smeden, M., Luo, J., le Cessie, S., Kastenmüller, G., Krumsiek, J., Noordam, R., van Heemst, D., Rosendaal, F.R., van Hylckama Vlieg, A., Willems van Dijk, M., Mook-Kanamori, D.O., 2020. A workflow for missing values imputation of untargeted metabolomics data. *Metabolites* 10 (12).
- Faust, G., Janzen, N.H., Bendig, C., Römer, L., Kaufmann, K., Weuster-Botz, D., 2014. Feeding strategies enhance high cell density cultivation and protein expression in milliliter scale bioreactors. *Biotechnol. J.* 9 (10), 1293–1303.
- Feng, X., Zhao, H., 2013. Investigating xylose metabolism in recombinant *Saccharomyces cerevisiae* via ¹³C metabolic flux analysis. *Microb. Cell Factories* 12 (1), 114.
- Flickinger, M.C., 2010. *Encyclopedia of Industrial Biotechnology, Bioprocess, Bioseparation, and Cell Technology*, vol. 3. Wiley, Hoboken, N.J, p. 3.
- Fondi, M., Liò, P., 2015. Multi-omics and metabolic modelling pipelines: challenges and tools for systems microbiology. *Microbiol. Res.* 171, 52–64.
- Furey, A., Moriarty, M., Bane, V., Kinsella, B., Lebane, M., 2013. Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta* 115, 104–122.
- Gao, D., Liu, H., Jiang, Y., Lin, J.-M., 2013. Recent advances in microfluidics combined with mass spectrometry: technologies and applications. *Lab Chip* 13 (17), 3309–3322.
- García-Pérez, I., Vallejo, M., García, A., Legido-Quigley, C., Barbas, C., 2008. Metabolic fingerprinting with capillary electrophoresis. *J. Chromatogr. A* 1204 (2), 130–139.
- Gebreselassie, N.A., Antoniewicz, M.R., 2015. ¹³C-metabolic flux analysis of co-cultures: a novel approach. *Metab. Eng.* 31, 132–139.
- Geier, B., Sogin, E.M., Michellod, D., Janda, M., Kompauer, M., Spengler, B., Dubilier, N., Liebecke, M., 2020. Spatial metabolomics of in situ host-microbe interactions at the micrometer scale. *Nat. Biol.* 5 (3), 498–510.
- Ghaffari, M.H., Jahanbekam, A., Sadri, H., Schuh, K., Dusel, G., Prehn, C., Adamski, J., Koch, C., Sauerwein, H., 2019. Metabolomics meets machine learning: longitudinal metabolite profiling in serum of normal versus overconditioned cows and pathway analysis. *J. Dairy Sci.* 102 (12), 11561–11585.
- Gibson, K.M., Ten Brink, H.J., Schor, D.S.M., Kok, R.M., Bootsma, A.H., Hoffmann, G.F., Jakobs, C., 1993. Stable-isotope dilution analysis of D- and L-2-hydroxyglutaric acid: application to the detection and prenatal diagnosis of D- and L-2-hydroxyglutaric acidemias. *Pediatr. Res.* 34 (3), 277–280.

- Gika, H.G., Theodoridis, G.A., Plumb, R.S., Wilson, I.D., 2014. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J. Pharmaceut. Biomed. Anal.* 87, 12–25.
- Gilmore, L.S., Heiles, S., Pieterse, C.L., 2019. Metabolic imaging at the single-cell scale: recent advances in mass spectrometry imaging. *Annu. Rev. Anal. Chem.* 12 (1), 201–224.
- Gondara, L., Wang, K., 2017. Recovering loss to followup information using denoising autoencoders. In: *IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 1936–1945, 2017.
- Gosai, H.B., Sachaniya, B.K., Dudhagara, D.R., Panseriya, H.Z., Dave, B.P., 2018. Bioengineering for multiple PAHs degradation using process centric and data centric approaches. *Chemometr. Intell. Lab. Syst.* 179, 99–108.
- Griffin, J.L., 2006. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361 (1465), 147–161.
- Hackett Sean, R., Zanotelli Vito, R.T., Xu, W., Goya, J., Park Junyoung, O., Perlman David, H., Gibney Patrick, A., Botstein, D., Storey John, D., Rabinowitz Joshua, D., 2016. Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 354 (6311), aaf2786.
- Heckmann, D., Lloyd, C.J., Mih, N., Ha, Y., Zielski, D.C., Haiman, Z.B., Desouki, A.A., Lercher, M.J., Palsson, B.O., 2018. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat. Commun.* 9 (1), 5252.
- Heijnen, J.J., Verheijen, P.J.T., 2013. Parameter identification of in vivo kinetic models: limitations and challenges. *Biotechnol. J.* 8 (7), 768–775.
- Heinemann, J., Noon, B., Mohigmi, M.J., Mazurie, A., Dickensheets, D.L., Bothner, B., 2014. Real-time digitization of metabolomics patterns from a living system using mass spectrometry. *J. Am. Soc. Mass Spectrom.* 25 (10), 1755–1762.
- Helmy, M., Smith, D., Selvarajoo, K., 2020. Systems biology approaches integrated with artificial intelligence for optimized metabolic engineering. *Metabolic. Eng. Commun.* 11, e00149.
- Hendry, J.I., Gopalakrishnan, S., Ungerer, J., Pakrasi, H.B., Tang, Y.J., Maranas, C.D., 2019. Genome-scale fluxome of *Synechococcus elongatus* UTEX 2973 using transient (13)C-labeling data. *Plant Physiol.* 179 (2), 761–769.
- Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. *Biophys. J.* 92 (5), 1792–1805.
- Heux, S., Poinot, J., Massou, S., Sokol, S., Portais, J.-C., 2014. A novel platform for automated high-throughput fluxome profiling of metabolic variants. *Metab. Eng.* 25, 8–19.
- Hira, Z.M., Gillies, D.F., 2015. A review of feature selection and feature extraction methods applied on microarray data. *Adv. Bioinformatic.* 2015, 198363.
- Hirayama, A., Tabata, S., Kudo, R., Hasebe, M., Suzuki, K., Tomita, M., Soga, T., 2020. The use of a double coaxial electrospray ionization sprayer improves the peak resolutions of anionic metabolites in capillary ion chromatography-mass spectrometry. *J. Chromatogr. A* 1619, 460914.
- Hollywood, K., Brison, D.R., Goodacre, R., 2006. Metabolomics: current technologies and future trends. *Proteomics* 6 (17), 4716–4723.
- Hrovatin, K., Fischer, D.S., Theis, F.J., 2022. Toward modeling metabolic state from single-cell transcriptomics. *Mol. Metabol.* 57, 101396.
- Hsu, C.C., ElNaggar, M.S., Peng, Y., Fang, J., Sanchez, L.M., Mascuch, S.J., Møller, K.A., Alazeh, E.K., Pikula, J., Quinn, R.A., Zeng, Y., Wolfe, B.E., Dutton, R.J., Gerwick, L., Zhang, L., Liu, X., Månsson, M., Dorrestein, P.C., 2013. Real-time metabolomics on living microorganisms using ambient electrospray ionization flow-probe. *Anal. Chem.* 85 (15), 7014–7018.
- Huang, W.-H., Ai, F., Wang, Z.-L., Cheng, J.-K., 2008. Recent advances in single-cell analysis using capillary electrophoresis and microfluidic devices. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* 866 (1–2), 104–122.
- Huang, B., Zeng, H., Dong, L., Li, Y., Sun, L., Zhu, Z., Chai, Y., Chen, W., 2011. Metabolite target analysis of isoprenoid pathway in *Saccharomyces cerevisiae* in response to genetic modification by GC-SIM-MS coupled with chemometrics. *Metabolomics* 7 (1), 134–146.
- Huft, J., Haynes, C.A., Hansen, C.L., 2013. Microfluidic integration of parallel solid-phase liquid chromatography. *Anal. Chem.* 85 (5), 2999–3005.
- Hui, S., Xing, X., Bader, G.D., 2013. Predicting PDZ domain mediated protein interactions from structure. *BMC Bioinf.* 14, 27.
- Ibáñez Alfredo, J., Fagerer Stephan, R., Schmidt Anna, M., Urban Pawel, L., Jefimovs, K., Geiger, P., Dechant, R., Heinemann, M., Zenobi, R., 2013. Mass spectrometry-based metabolomics of single yeast cells. *Proc. Natl. Acad. Sci. USA* 110 (22), 8790–8794.
- Jagtap, S.S., Bedekar, A.A., Singh, V., Jin, Y.-S., Rao, C.V., 2021. Metabolic engineering of the oleaginous yeast *Yarrowia lipolytica* PO1f for production of erythritol from glycerol. *Biotechnol. Biofuels* 14 (1), 188–188.
- Jeffries, J.G., Seaver, S.M.D., Faria, J.P., Henry, C.S., 2018. A pathway for every product? Tools to discover and design plant metabolism. *Plant Sci.* 273, 61–70.
- Jervis, A.J., Carbonell, P., Taylor, S., Sung, R., Dunstan, M.S., Robinson, C.J., Breitling, R., Takano, E., Scrutton, N.S., 2019. SelProm: a queryable and predictive expression vector selection tool for *Escherichia coli*. *ACS Synth. Biol.* 8 (7), 1478–1483.
- Jones, J.A., Vernacchio, V.R., Sinkoe, A.L., Collins, S.M., Ibrahim, M.H.A., Lachance, D. M., Hahn, J., Koffas, M.A.G., 2016. Experimental and computational optimization of an *Escherichia coli* co-culture for the efficient production of flavonoids. *Metab. Eng.* 35, 55–63.
- Jousse, C., Dalle, C., Canet, I., Lagrée, M., Traïkia, M., Lyan, B., Mendes, C., Sancelme, M., Amato, P., Delort, A.-M., 2017. Metabolomic study of the response to cold shock in a strain of *Pseudomonas syringae* isolated from cloud water. *Metabolomics* 14 (1), 11.
- Judge, M.T., Wu, Y., Tayyari, F., Hattori, A., Glushka, J., Ito, T., Arnold, J., Edison, A.S., 2019. Continuous in vivo metabolism by NMR. *Front. Mol. Biosci.* 6, 26.
- Kalfe, A., Telfah, A., Lambert, J., Hergenröder, R., 2015. Looking into living cell systems: planar waveguide microfluidic NMR detector for in vitro metabolomics of tumor spheroids. *Anal. Chem.* 87 (14), 7402–7410.
- Kanani, H., Chrysanthopoulos, P.K., Klapa, M.I., 2008. Standardizing GC-MS metabolomics. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* 871 (2), 191–201.
- Karp, P.D., Latendresse, M., Caspi, R., 2011. The pathway tools pathway prediction algorithm. *Standard Genomic. Sci.* 5 (3), 424–429.
- Kim, S., Lee, D.Y., Wohlgemuth, G., Park, H.S., Fiehn, O., Kim, K.H., 2013. Evaluation and optimization of metabolome sample preparation methods for *Saccharomyces cerevisiae*. *Anal. Chem.* 85 (4), 2169–2176.
- Kim, M., Rai, N., Zorraquino, V., Tagkopoulos, I., 2016. Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nat. Commun.* 7 (1), 13090.
- Kim, W.J., Kim, H.U., Lee, S.Y., 2017. Current state and applications of microbial genome-scale metabolic models. *Curr. Opin. Struct. Biol.* 2, 10–18.
- Kim, O.D., Rocha, M., Maia, P., 2018a. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. *Front. Microbiol.* 9.
- Kim, H.K., Min, S., Song, M., Jung, S., Choi, J.W., Kim, Y., Lee, S., Yoon, S., Kim, H., 2018b. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat. Biotechnol.* 36 (3), 239–241.
- Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J., Hanhineva, K., 2019. Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study. *BMC Bioinf.* 20 (1), 492.
- Koley, S., Chu, K.L., Gill, S.S., Allen, D.K., 2022. An efficient LC-MS method for isomer separation and detection of sugars, phosphorylated sugars, and organic acids. *J. Exp. Bot.* 73 (9), 2938–2952.
- Kotera, M., Goto, S., 2016. Metabolic pathway reconstruction strategies for central metabolism and natural product biosynthesis. *Biophys. Physicobiol.* 13, 195–205.
- Kraly, J.R., Holcomb, R.E., Guan, Q., Henry, C.S., 2009. Review: microfluidic applications in metabolomics and metabolic profiling. *Anal. Chim. Acta* 653 (1), 23–35.
- Kresnowati, M.T.A.P., van Winden, W.A., Heijnen, J.J., 2005. Determination of elasticities, concentration and flux control coefficients from transient metabolite data using linlog kinetics. *Metab. Eng.* 7 (2), 142–153.
- Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., Zhang, K., 2018. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36 (1), 70–80.
- Langer, J., Jimenez de Aberasturi, D., Aizpurua, J., Alvarez-Puebla, R.A., Auguie, B., Baumberg, J.J., Bazan, G.C., Bell, S.E.J., Boisen, A., Brolo, A.G., Choo, J., Cialla-May, D., Deckert, V., Fabris, L., Faulds, K., García de Abajo, F.J., Goodacre, R., Graham, D., Haes, A.J., Haynes, C.L., Huck, C., Itoh, T., Käll, M., Kneipp, J., Kotov, N.A., Kuang, H., Le Ru, E.C., Lee, H.K., Li, J.F., Ling, X.Y., Maier, S.A., Mayerhöfer, T., Moskovits, M., Murakoshi, K., Nam, J.M., Nie, S., Ozaki, Y., Pastoriza-Santos, I., Perez-Juste, J., Popp, J., Pucci, A., Reich, S., Ren, B., Schatz, G. C., Shegai, T., Schlücker, S., Tay, L.L., Thomas, K.G., Tian, Z.Q., Van Duyn, R.P., Vo-Dinh, T., Wang, Y., Willets, K.A., Xu, C., Xu, H., Xu, Y., Yamamoto, Y.S., Zhao, B., Liz-Marzán, L.M., 2020. Present and future of surface-enhanced Raman scattering. *ACS Nano* 14 (1), 28–117.
- Lankadurai, B.P., Nagato, E.G., Simpson, M.J., 2013. Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environ. Rev.* 21 (3), 180–205.
- Lee, S.J., Lee, D.-Y., Kim, T.Y., Kim, B.H., Lee, J., Lee, S.Y., 2005. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. *Appl. Environ. Microbiol.* 71 (12), 7880–7887.
- Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U., Lee, S.Y., 2007. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol. Syst. Biol.* 3, 149.
- Lee, J.W., Kim, T.Y., Jang, Y.-S., Choi, S., Lee, S.Y., 2011. Systems metabolic engineering for chemicals and materials. *Trends Biotechnol.* 29 (8), 370–378.
- Lee, J.W., Na, D., Park, J.M., Lee, J., Choi, S., Lee, S.Y., 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* 8 (6), 536–546.
- Lee, M.E., Aswani, A., Han, A.S., Tomlin, C.J., Dueber, J.E., 2013. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* 41 (22), 10668–10678.
- Lee, S.Y., Kim, H.U., Chae, T.U., Cho, J.S., Kim, J.W., Shin, J.H., Kim, D.I., Ko, Y.-S., Jang, W.D., Jang, Y.-S., 2019. A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* 2 (1), 18–33.
- Leighty, R.W., Antoniewicz, M.R., 2011. Dynamic metabolic flux analysis (DMFA): a framework for determining fluxes at metabolic non-steady state. *Metab. Eng.* 13 (6), 745–755.
- Lewis, I.A., Karsten, R.H., Norton, M.E., Tonelli, M., Westler, W.M., Markley, J.L., 2010. NMR method for measuring carbon-13 isotopic enrichment of metabolites in complex solutions. *Anal. Chem.* 82 (11), 4558–4563.
- Li, G., Rabe, K.S., Nielsen, J., Engqvist, M.K.M., 2019. Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. *ACS Synth. Biol.* 8 (6), 1411–1420.
- Liebal, U.W., Phan, A.N.T., Sudhakar, M., Raman, K., Blank, L.M., 2020. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites* 10 (6).
- Liebermeister, W., Klipp, E., 2006. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.* 3 (1), 41.

- Lin, Y., Schiavo, S., Orjala, J., Vouros, P., Kautz, R., 2008. Microscale LC-MS-NMR platform applied to the identification of active cyanobacterial metabolites. *Anal. Chem.* 80 (21), 8045–8054.
- Lin, D., Zhang, J., Li, J., Xu, C., Deng, H.-W., Wang, Y.-P., 2016. An integrative imputation method based on multi-omics datasets. *BMC Bioinf.* 17 (1), 247.
- Lin, G.-M., Warden-Rothman, R., Voigt, C.A., 2019. Retrosynthetic design of metabolic pathways to chemicals not found in nature. *Curr. Opin. Struct. Biol.* 14, 82–107.
- Linderman, G.C., Zhao, J., Roulis, M., Bielecki, P., Flavell, R.A., Nadler, B., Kluger, Y., 2022. Zero-preserving imputation of single-cell RNA-seq data. *Nat. Commun.* 13 (1), 192.
- Link, H., Kochanowski, K., Sauer, U., 2013. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat. Biotechnol.* 31 (4), 357–361.
- Link, H., Christodoulou, D., Sauer, U., 2014. Advancing metabolic models with kinetic information. *Curr. Opin. Biotechnol.* 29, 8–14.
- Link, H., Fuhrer, T., Gerosa, L., Zamboni, N., Sauer, U., 2015. Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nat. Methods* 12 (11), 1091–1097.
- Liu, B.F., Ozaki, M., Hisamoto, H., Luo, Q., Utsumi, Y., Hattori, T., Terabe, S., 2005. Microfluidic chip toward cellular ATP and ATP-conjugated metabolic analysis with bioluminescence detection. *Anal. Chem.* 77 (2), 573–578.
- Liu, X., Zhou, L., Shi, X., Xu, G., 2019a. New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. *TrAC, Trends Anal. Chem.* 121, 115665.
- Liu, Y., Chen, X., Zhang, Y., Liu, J., 2019b. Advancing single-cell proteomics and metabolomics with microfluidic technologies. *Analyst* 144 (3), 846–858.
- Long, C.P., Antoniewicz, M.R., 2019. High-resolution (13)C metabolic flux analysis. *Nat. Protoc.* 14 (10), 2856–2877.
- Long, Q., Liu, X., Yang, Y., Li, L., Harvey, L., McNeil, B., Bai, Z., 2014. The development and application of high throughput cultivation technology in bioprocess development. *J. Biotechnol.* 192, 323–338.
- Lu, W., Su, X., Klein, M., Lewis, I., Fiehn, O., Rabinowitz, J., 2016. Metabolite measurement: pitfalls to avoid and practices to follow. *Annu. Rev. Biochem.* 86, 277–304.
- Lu, W., Su, X., Klein, M.S., Lewis, I.A., Fiehn, O., Rabinowitz, J.D., 2017. Metabolite measurement: pitfalls to avoid and practices to follow. *Annu. Rev. Biochem.* 86, 277–304.
- Ma, T., Zhang, A., 2019. Integrate multi-omics data with biological interaction networks using Multi-view Factorization AutoEncoder (MAE). *BMC Genom.* 20 (11), 944.
- Ma, J., Yu, M.K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., Ideker, T., 2018. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* 15 (4), 290–298.
- Magnus, J.B., Hollwedel, D., Oldiges, M., Takors, R., 2006. Monitoring and modeling of the reaction dynamics in the valine/leucine synthesis pathway in *Corynebacterium glutamicum*. *Biotechnol. Prog.* 22 (4), 1071–1083.
- Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. 3rd, Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83 (3), 1331–1340.
- Masampally, V.S., Pareek, A., Runkana, V., 2018. Cascade Gaussian process regression framework for biomass prediction in a fed-batch reactor. In: *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 128–135, 18–21 Nov. 2018; 2018.
- Mauve, C., Khelifi, S., Gilard, F., Mouille, G., Farjon, J., 2016. Sensitive, highly resolved, and quantitative 1H–13C NMR data in one go for tracking metabolites in vegetal extracts. *Chem. Commun.* 52 (36), 6142–6145.
- McConville, F.X., Kessler, S.B., 2010. Scale-up of mixing processes: a primer. *Chem. Eng. Pharma. Indust.* 249–267.
- Meng, H., Wang, Y., Hua, Q., Zhang, S., Wang, X., 2011. In silico analysis and experimental improvement of taxadiene heterologous biosynthesis in *Escherichia coli*. *Biotechnol. Bioeng.* 16 (2), 205–215.
- Meng, H., Wang, J., Xiong, Z., Xu, F., Zhao, G., Wang, Y., 2013. Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. *PLoS One* 8 (4), e60288.
- Meyer, H., Weidmann, H., Lalk, M., 2013. Methodological approaches to help unravel the intracellular metabolome of *Bacillus subtilis*. *Microb. Cell Factories* 12, 69–69.
- Mias, G.I., Yusufaly, T., Roushangar, R., Brooks, L.R.K., Singh, V.V., Christou, C., 2016. MathOmics: an integrative platform for dynamic omics. *Sci. Rep.* 6 (1), 37237.
- Min Lee, J., Gianchandani, E.P., Eddy, J.A., Papin, J.A., 2008. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* 4 (5), e1000086.
- Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N.C., Ping, P., 2019. Machine learning and integrative analysis of biomedical big data. *Genes* 10 (2).
- Misra, B.B., Langefeld, C.D., Olivier, M., Cox, L.A., 2018. Integrated omics: tools, advances, and future approaches. *J. Mol. Endocrinol.* Jul 13:JME-18-0055.
- Mo, M.L., Palsson, B.O., Herrgård, M.J., 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* 3, 37.
- Monge, M.E., Dodds, J.N., Baker, E.S., Edison, A.S., Fernández, F.M., 2019. Challenges in identifying the dark molecules of life. *Annu. Rev. Anal. Chem.* 12 (1), 177–199.
- Moon, S.Y., Hong, S.H., Kim, T.Y., Lee, S.Y., 2008. Metabolic engineering of *Escherichia coli* for the production of malic acid. *Biochem. Eng. J.* 40 (2), 312–320.
- Mousavi, F., Bojko, B., Pawliszyn, J., 2019. High-throughput solid-phase microextraction-liquid chromatography-mass spectrometry for microbial untargeted metabolomics. *Methods Mol. Biol.* 1859, 133–152.
- Movasaghi, Z., Rehman, S., ur Rehman, D.I., 2008. Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* 43 (2), 134–179.
- Murovec, B., Makuc, D., Kolbl Repinc, S., Prevorsek, Z., Zavec, D., Sket, R., Pečnik, K., Plavec, J., Stres, B., 2018. 1H NMR metabolomics of microbial metabolites in the four MW agricultural biogas plant reactors: a case study of inhibition mirroring the acute rumen acidosis symptoms. *J. Environ. Manag.* 222, 428–435.
- Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H., Tomita, M., 2009. Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol. Syst. Biol.* 5, 306.
- Nandi, S., Subramanian, A., Sarkar, R.R., 2017. An integrative machine learning strategy for improved prediction of essential genes in *Escherichia coli* metabolism using flux-coupled features. *Mol. Biosyst.* 13 (8), 1584–1596.
- Nemadodzi, L.E., Vervoort, J., Prinsloo, G., 2020. NMR-based metabolomic analysis and microbial composition of soil supporting burkea africana growth. *Metabolites* 10 (10), 402.
- Nguyen, D.D., Saharuka, V., Kovalev, V., Stuart, L., Del Prete, M., Lubowiecka, K., De Mot, R., Venturi, V., Alexandrov, T., 2021. Facilitating imaging mass spectrometry of microbial specialized metabolites with METASPACE. *Metabolites* 11 (8).
- Niehaus, T.D., Nguyen, T.N.D., Gidda, S.K., ElBadawi-Sidhu, M., Lambrecht, J.A., McCarty, D.R., Downs, D.M., Cooper, A.J.L., Fiehn, O., Mullen, R.T., Hanson, A.D., 2014. Arabidopsis and maize RidA proteins preempt reactive enamine/imine damage to branched-chain amino acid biosynthesis in plastids. *Plant Cell* 26 (7), 3010–3022.
- Nikere, I.E., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 2006. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. *BMC Bioinf.* 7, 540–540.
- Oberhardt, M.A., Palsson, B., Papin, J.A., 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5, 320.
- Ohla, S., Belder, D., 2012. Chip-based separation devices coupled to mass spectrometry. *Curr. Opin. Chem. Biol.* 16 (3), 453–459.
- Okahashi, N., Maeda, K., Kawana, S., Iida, J., Shimizu, H., Matsuda, F., 2019. Sugar phosphate analysis with baseline separation and soft ionization by gas chromatography-negative chemical ionization-mass spectrometry improves flux estimation of bidirectional reactions in cancer cells. *Metab. Eng.* 51, 43–49.
- Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? *Nat. Biotechnol.* 28 (3), 245–248.
- Paglia, G., Astarita, G., 2017. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* 12 (4), 797–813.
- Palmer, A., Phapale, P., Chernyavsky, I., Lavigne, R., Fay, D., Tarasov, A., Kovalev, V., Fuchser, J., Nikolenko, S., Pineau, C., Becker, M., Alexandrov, T., 2017. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* 14 (1), 57–60.
- Papadimitropoulos, M.P., Vasilopoulou, C.G., Maga-Nteve, C., Klapa, M.I., 2018. Untargeted GC-MS metabolomics. *Methods Mol. Biol.* 1738, 133–147.
- Pappu, J.S.M., Gummadi, S.N., 2016. Modeling and simulation of xylitol production in bioreactor by *Debaryomyces nepalensis* NCYC 3413 using unstructured and artificial neural network models. *Bioresour. Technol.* 220, 490–499.
- Park, J.M., Kim, T.Y., Lee, S.Y., 2009. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnol. Adv.* 27 (6), 979–988.
- Park, J.M., Kim, T.Y., Lee, S.Y., 2010. Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses. *Proc. Natl. Acad. Sci. USA* 107 (33), 14931–14936.
- Park, J.H., Kim, T.Y., Lee, K.H., Lee, S.Y., 2011. Fed-batch culture of *Escherichia coli* for L-valine production based on in silico flux response analysis. *Biotechnol. Bioeng.* 108 (4), 934–946.
- Park, J.M., Park, H.M., Kim, W.J., Kim, H.U., Kim, T.Y., Lee, S.Y., 2012. Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst. Biol.* 6, 106.
- Park Jin, H., Lee Kwang, H., Kim Tae, Y., Lee Sang, Y., 2007. Metabolic engineering of *Escherichia coli* for the production of l-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc. Natl. Acad. Sci. USA* 104 (19), 7797–7802.
- Parry, Bradley R., Surovtsev, Ivan V., Cabeen, Matthew T., O'Hern, Corey S., Dufresne, Eric R., Jacobs-Wagner, C., 2014. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* 156 (1), 183–194.
- Patejko, M., Jacyna, J., Markuszewski, M.J., 2017. Sample preparation procedures utilized in microbial metabolomics: an overview. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* 1043, 150–157.
- Patra, B., Sharma, M., Hale, W., Utz, M., 2021. Time-resolved non-invasive metabolomic monitoring of a single cancer spheroid by microfluidic NMR. *Sci. Rep.* 11 (1), 53.
- Patti, G.J., Yanes, O., Siuzdak, G., 2012. Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13 (4), 263–269.
- Pawliszyn, J., 2012. Comprehensive Sampling and Sample Preparation.
- Petras, D., Jarmusch, A.K., Dorrestein, P.C., 2017. From single cells to our planet—recent advances in using mass spectrometry for spatially resolved metabolomics. *Curr. Opin. Chem. Biol.* 36, 24–31.
- Petucci, C., Zelenin, A., Culver, J.A., Gabriel, M., Kirkbride, K., Christison, T.T., Gardell, S.J., 2016. Use of ion chromatography/mass spectrometry for targeted metabolite profiling of polar organic acids. *Anal. Chem.* 88 (23), 11799–11803.
- Phan, A., Blank, L., 2020. GC-MS-based metabolomics for the smut fungus *Ustilago maydis*: a comprehensive method optimization to quantify intracellular metabolites. *Front. Mol. Biosci.* 7, 211.
- Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 14 (11), 2367–2376.
- Pičmanová, M., Moses, T., Cortada-García, J., Barrett, G., Florance, H., Pandor, S., Burgess, K., 2022. Rapid HILIC-Z ion mobility mass spectrometry (RHIMMS) method for untargeted metabolomics of complex biological samples. *Metabolites* 18 (3), 16.
- Pinu, F.R., Villas-Boas, S.G., 2017. Extracellular microbial metabolomics: the state of the art. *Metabolites* 7 (3).
- Pinu, F.R., Villas-Boas, S.G., Aggio, R., 2017. Analysis of intracellular metabolites from microorganisms: quenching and extraction protocols. *Metabolites* 7 (4), 53.

- Plaimas, K., Mallm, J.P., Oswald, M., Svava, F., Sourjik, V., Eils, R., König, R., 2008. Machine learning based analyses on metabolic networks supports high-throughput knockout screens. *BMC Syst. Biol.* 2, 67.
- Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., Gross, S.S., Dorfman, L., McLean, C.Y., DePristo, M.A., 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36 (10), 983–987.
- Popp, D., Gentler, F., 2020. μ BialSim: constraint-based dynamic simulation of complex microbiomes. *Front. Bioeng. Biotechnol.* 8.
- Prakadan, S.M., Shalek, A.K., Weitz, D.A., 2017. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nat. Rev. Genet.* 18 (6), 345–361.
- Prasanna, C.B., Jaiswal, D., Davis, R., Wangikar, P.P., 2018. An improved method for extraction of polar and charged metabolites from cyanobacteria. *PLoS One* 13 (10), e0204273.
- Protsyuk, I., Melnik, A.V., Nothias, L.F., Rappez, L., Phapale, P., Aksenov, A.A., Bouslimani, A., Ryazanov, S., Dorrestein, P.C., Alexandrov, T., 2018. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. *Nat. Protoc.* 13 (1), 134–154.
- Purwaha, P., Silva, L.P., Hawke, D.H., Weinstein, J.N., Lorenzi, P.L., 2014. An artifact in LC-MS/MS measurement of glutamine and glutamic acid: in-source cyclization to pyroglutamic acid. *Anal. Chem.* 86 (12), 5633–5637.
- Pyke, J.S., Callahan, D.L., Kanojia, K., Bowne, J., Sahani, S., Tull, D., Bacic, A., McConville, M.J., Roessner, U., 2015. A tandem liquid chromatography–mass spectrometry (LC–MS) method for profiling small molecules in complex samples. *Metabolomics* 11 (6), 1552–1562.
- Qian, H., Beard, D.A., 2005. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys. Chem.* 114 (2–3), 213–220.
- Quest, D.J., Land, M.L., Brettin, T.S., Cottingham, R.W., 2010. Next generation models for storage and representation of microbial biological annotation. *BMC Bioinf.* 11 (Suppl. 6), S15. Suppl. 6.
- Radiivojević, T., Costello, Z., Workman, K., Garcia Martin, H., 2020. A machine learning Automated Recommendation Tool for synthetic biology. *Nat. Commun.* 11 (1), 4879.
- Rappez, L., Stadler, M., Triana, S., Phapale, P., Heikenwalder, M., Alexandrov, T., 2019. Spatial single-cell profiling of intracellular metabolomes in situ. *bioRxiv*, 510222.
- Ravikirthi, P., Suthers, P.F., Maranas, C.D., 2011. Construction of an E. Coli genome-scale atom mapping model for MFA calculations. *Biotechnol. Bioeng.* 108 (6), 1372–1382.
- Reardon, P.N., Marean-Reardon, C.L., Bukovec, M.A., Coggins, B.E., Isern, N.G., 2016. 3D TOCSY-HSQC NMR for metabolic flux analysis using non-uniform sampling. *Anal. Chem.* 88 (5), 2825–2831.
- Ren, J.L., Zhang, A.H., Kong, L., Wang, X.J., 2018. Advances in mass spectrometry-based metabolomics for investigation of metabolites. *RSC Adv.* 8 (40), 22335–22350.
- Rizzi, M., Baltes, M., Theobald, U., Reuss, M., 1997. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnol. Bioeng.* 55 (4), 592–608.
- Roume, H., Heintz-Buschart, A., Müller, E.E.L., Wilmes, P., 2013. Chapter eleven - sequential isolation of metabolites, RNA, DNA, and proteins from the same unique sample. In: DeLong, E.F. (Ed.), *Methods in Enzymology*, vol. 531. Academic Press, pp. 219–236.
- Rubakhin, S.S., Romanova, E.V., Nemes, P., Sweedler, J.V., 2011. Profiling metabolites and peptides in single cells. *Nat. Methods* 8 (4), S20–S29.
- Saa, P.A., Nielsen, L.K., 2017. Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. *Biotechnol. Adv.* 35 (8), 981–1003.
- Sans, M., Feider, C.L., Eberlin, L.S., 2018. Advances in mass spectrometry imaging coupled to ion mobility spectrometry for enhanced imaging of biological tissues. *Curr. Opin. Chem. Biol.* 42, 138–146.
- Sasidharan, K., Soga, T., Tomita, M., Murray, D.B., 2012. A yeast metabolite extraction protocol optimised for time-series analyses. *PLoS One* 7 (8), e44283.
- Schaber, J., Liebermeister, W., Klipp, E., 2009. Nested uncertainties in biochemical models. *IET Syst. Biol.* 3 (1), 1–9.
- Schrimpe-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D., McLean, J.A., 2016. Untargeted metabolomics strategies-challenges and emerging directions. *J. Am. Soc. Mass Spectrom.* 27 (12), 1897–1905.
- Selvarajoo, K., 2011. Macroscopic law of conservation revealed in the population dynamics of Toll-like receptor signaling. *Cell Commun. Signal.* : CCS 9, 9.
- Selvarajoo, K., 2021. The need for integrated systems biology approaches for biotechnological applications. *Biotechnol. Note.* 2, 39–43.
- Selvarajoo, K., Tomita, M., Tsuchiya, M., 2009. Can complex cellular processes be governed by simple linear rules? *J. Bioinf. Comput. Biol.* 7, 243–268.
- Sengupta, A., Krishnaiah, S.Y., Rhoades, S., Growe, J., Sflaff, B., Venkataraman, A., Olarerin-George, A.O., Van Dang, C., Hogenesch, J.B., Weljie, A.M., 2016. Deciphering the duality of clock and growth metabolism in a cell autonomous system using NMR profiling of the secretome. *Metabolites* 6 (3), 23.
- Shah, A.D., Bartlett, J.W., Carpenter, J., Nicholas, O., Hemingway, H., 2014. Comparison of random forest and parametric imputation models for imputing missing data using mice: a caliber study. *Am. J. Epidemiol.* 179 (6), 764–774.
- Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E., 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* 3, 101.
- Si-Hung, L., Troyer, C., Causon, T., Hann, S., 2019. Sensitive quantitative analysis of phosphorylated primary metabolites using selective metal oxide enrichment and GC- and IC- MS/MS. *Talanta* 205, 120147.
- Siegel, D., Permentier, H., Reijngoud, D.J., Bischoff, R., 2014. Chemical and technical challenges in the analysis of central carbon metabolites by liquid-chromatography mass spectrometry. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 966, 21–33.
- Sitnikov, D.G., Monnin, C.S., Vuckovic, D., 2016. Systematic assessment of seven solvent and solid-phase extraction methods for metabolomics analysis of human plasma by LC-MS. *Sci. Rep.* 6 (1), 38885.
- Srinivasan, S., Cluett, W.R., Mahadevan, R., 2015. Constructing kinetic models of metabolism at genome-scales: a review. *Biotechnol. J.* 10 (9), 1345–1359.
- Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., 1998. Chapter 8 - metabolic flux analysis. In: Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J. (Eds.), *Metabolic Engineering*. Academic Press, San Diego, pp. 309–351.
- Steuer, R., Gross, T., Selbig, J., Blasius, B., 2006. Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 103 (32), 11868–11873.
- Strutz, J., Martin, J., Greene, J., Broadbelt, L., Tyo, K., 2019. Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain. *Curr. Opin. Biotechnol.* 59, 24–30.
- Struys, E.A., Jansen, E.E., Verhoeven, N.M., Jakobs, C., 2004. Measurement of urinary D- and L-2-hydroxyglutarate enantiomers by stable-isotope-dilution liquid chromatography-tandem mass spectrometry after derivatization with diacetyl-L-tartaric anhydride. *Clin. Chem.* 50 (8), 1391–1395.
- Tabatabaei Anaraki, M., Simpson, M.J., Simpson, A.J., 2018. Reducing impacts of organism variability in metabolomics via time trajectory in vivo NMR. *Magn. Reson. Chem.* 56 (11), 1117–1123.
- Teusink, B., Passarge, J., Reijenga, C.A., Esgalham, E., van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., van Dam, K., Westerhoff, H.V., Snoep, J.L., 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* 267 (17), 5313–5329.
- Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., Reuss, M., 1997. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae* : I. Experimental observations. *Biotechnol. Bioeng.* 55 (2), 305–316.
- Thiele, I., Palsson, B.O., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5 (1), 93–121.
- Tomar, N., De, R.K., 2013. Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene* 521 (1), 1–14.
- Tongman, S., Chanama, S., Chanama, M., Plaimas, K., Lursinsap, C., 2017. Metabolic pathway synthesis based on predicting compound transformable pairs by using neural classifiers with imbalanced data handling. *Expert Syst. Appl.* 88, 45–57.
- Töpfer, N., Kleessen, S., Nikoloski, Z., 2015. Integration of metabolomics data into metabolic networks. *Front. Plant Sci.* 6, 49.
- Traxler, M.F., Watrous, J.D., Alexandrov, T., Dorrestein, P.C., Kolter, R., Davies, J.E., 2013. Interspecies interactions stimulate diversification of the streptomycetes coelicolor secreted metabolome. *mBio* 4 (4), e00459-13.
- Tummler, K., Klipp, E., 2018. The discrepancy between data for and expectations on metabolic models: how to match experiments and computational efforts to arrive at quantitative predictions? *Curr. Opin. Struct. Biol.* 8, 1–6.
- Uchimiyama, M., Tsuboi, Y., Ito, K., Date, Y., Kikuchi, J., 2017. Bacterial substrate transformation tracked by stable-isotope-guided NMR metabolomics: application in a natural aquatic microbial community. *Metabolites* 7 (4), 52.
- Unthan, S., Radek, A., Wiechert, W., Oldiges, M., Noack, S., 2015. Bioprocess automation on a Mini Pilot Plant enables fast quantitative microbial phenotyping. *Microb. Cell Factories* 14, 32.
- Vavricka, C.J., Hasunuma, T., Kondo, A., 2020. Dynamic metabolomics for engineering biology: accelerating learning cycles for bioproduction. *Trends Biotechnol.* 38 (1), 68–82.
- Vert, J.-P., Qiu, J., Noble, W.S., 2007. A new pairwise kernel for biological network inference with support vector machines. *BMC Bioinf.* 8 (10), S8.
- Villas-Bóas, S.G., Hojer-Pedersen, J., Åkesson, M., Smedsgaard, J., Nielsen, J., 2005a. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* 22 (14), 1155–1169.
- Villas-Bóas, S.G., Mas, S., Åkesson, M., Smedsgaard, J., Nielsen, J., 2005b. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.* 24 (5), 613–646.
- Vivek, V., Eka, F.N., Chew, W., 2022. Mixing studies in an unbaffled bioreactor using a computational model corroborated with in-situ Raman and imaging analyses. *Chem. Eng. J. Adv.* 9, 100232.
- Volk, M.J., Lourentzou, I., Mishra, S., Vo, L.T., Zhai, C., Zhao, H., 2020. Biosystems design by machine learning. *ACS Synth. Biol.* 9 (7), 1514–1533.
- Volkova, S., Matos, M.R.A., Mattanovich, M., Marín de Mas, I., 2020. Metabolic modelling as a framework for metabolomics data integration and analysis. *Metabolites* 10 (8).
- Vorreiter, F., Richter, S., Peter, M., Baumann, S., Von Bergen, M., Tomm, J., 2016. Comparison and optimization of methods for the simultaneous extraction of DNA, RNA, proteins, and metabolites. *Anal. Biochem.* 508.
- Wang, Q., Chen, X., Yang, Y., Zhao, X., 2006. Genome-scale in silico aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. *Appl. Microbiol. Biotechnol.* 73 (4), 887–894.
- Wang, B.L., Ghaderi, A., Zhou, H., Agresti, J., Weitz, D.A., Fink, G.R., Stephanopoulos, G., 2014. Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption. *Nat. Biotechnol.* 32 (5), 473–478.
- Wang, X., Yi, L., Mukhitov, N., Schrell, A.M., Dhumpa, R., Roper, M.G., 2015. Microfluidics-to-mass spectrometry: a review of coupling methods and applications. *J. Chromatogr. A* 1382, 98–116.
- Watrous, J.D., Dorrestein, P.C., 2011. Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* 9 (9), 683–694.
- Weiß, J., 1987. The chromatography of ions — an advancement of inorganic analytical chemistry. *Fresenius' Z. für Anal. Chem.* 327 (1), 25–26.
- Wishart, D.S., 2007. Current progress in computational metabolomics. *Briefings Bioinf.* 8 (5), 279–293.

- Wordofa, G.G., Kristensen, M., Schrübbers, L., McCloskey, D., Forster, J., Schneider, K., 2017. Quantifying the metabolome of *Pseudomonas taiwanensis* VLB120: evaluation of hot and cold combined quenching/extraction approaches. *Anal. Chem.* 89 (17), 8738–8747.
- Wu, Z., Kan, S.B.J., Lewis, R.D., Wittmann, B.J., Arnold, F.H., 2019. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. USA* 116 (18), 8852–8858.
- Wu, Q., Wang, J.-Y., Han, D.-Q., Yao, Z.-P., 2020. Recent advances in differentiation of isomers by ion mobility mass spectrometry. *TrAC, Trends Anal. Chem.* 124, 115801.
- Xu, P., Rizzoni, E.A., Sul, S.-Y., Stephanopoulos, G., 2017. Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth. Biol.* 6 (1), 148–158.
- Yamanishi, Y., Vert, J.-P., Kanehisa, M., 2005. Supervised enzyme network inference from the integration of genomic data and chemical information. *Bioinformatics* 21 (Suppl. 1_1), i468–i477.
- Yang, Z., Michailidis, G., 2016. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* 32 (1), 1–8.
- Yang, Y.-L., Xu, Y., Straight, P., Dorrestein, P.C., 2009. Translating metabolic exchange with imaging mass spectrometry. *Nat. Chem. Biol.* 5 (12), 885–887.
- Yang, J.H., Wright, S.N., Hamblin, M., McCloskey, D., Alcantar, M.A., Schrübbers, L., Lopatkin, A.J., Satish, S., Nili, A., Palsson, B.O., Walker, G.C., Collins, J.J., 2019a. A white-box machine learning approach for revealing antibiotic mechanisms of action. *Cell* 177 (6), 1649–1661 e9.
- Yang, K.K., Wu, Z., Arnold, F.H., 2019b. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* 16 (8), 687–694.
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., Baker, D., 2020. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.* 117 (3), 1496–1503.
- Yasemi, M., Jolicoeur, M., 2021. Modelling cell metabolism: a review on constraint-based steady-state and kinetic approaches. *Processes* 9 (2).
- Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* 7 (7), 445–452.
- Yip, K.Y., Cheng, C., Gerstein, M., 2013. Machine learning and genome annotation: a match meant to be? *Genome Biol.* 14 (5), 205.
- Yizhak, K., Benyamini, T., Liebermeister, W., Rupp, E., Shlomi, T., 2010. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* 26 (12), i255–i260.
- Zampieri, G., Vijayakumar, S., Yaneske, E., Angione, C., 2019. Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput. Biol.* 15 (7), e1007084.
- Zelezniak, A., Vowinkel, J., Capuano, F., Messner, C.B., Demichev, V., Polowsky, N., Müllleder, M., Kamrad, S., Klaus, B., Keller, M.A., Ralser, M., 2018. Machine learning predicts the yeast metabolome from the quantitative proteome of kinase knockouts. *Cell systems* 7 (3), 269–283 e6.
- Zenobi, R., 2013. Single-cell metabolomics: analytical and biological perspectives. *Science* 342 (6163), 1243259.
- Zhang, B., Powers, R., 2012. Analysis of bacterial biofilms using NMR-based metabolomics. *Future Med. Chem.* 4 (10), 1273–1306.
- Zhang, X.-C., Wei, Z.-W., Gong, X.-Y., Si, X.-Y., Zhao, Y.-Y., Yang, C.-D., Zhang, S.-C., Zhang, X.-R., 2016. Integrated droplet-based microextraction with ESI-MS for removal of matrix interference in single-cell analysis. *Sci. Rep.* 6 (1), 24730.
- Zhang, W., Hankemeier, T., Ramautar, R., 2017. Next-generation capillary electrophoresis–mass spectrometry approaches in metabolomics. *Curr. Opin. Biotechnol.* 43, 1–7.
- Zhang, X., Quinn, K., Cruickshank-Quinn, C., Reisdorph, R., Reisdorph, N., 2018a. The application of ion mobility mass spectrometry to metabolomics. *Curr. Opin. Chem. Biol.* 42, 60–66.
- Zhang, C., Chen, X., Lindley, N.D., Too, H.P., 2018b. A "plug-n-play" modular metabolic system for the production of apocarotenoids. *Biotechnol. Bioeng.* 115 (1), 174–183.
- Zheng, Z.-Y., Guo, X.-N., Zhu, K.-X., Peng, W., Zhou, H.-M., 2017. Artificial neural network – genetic algorithm to optimize wheat germ fermentation condition: application to the production of two anti-tumor benzoquinones. *Food Chem.* 227, 264–270.
- Zhong, F., Xu, M., Metz, P., Ghosh-Dastidar, P., Zhu, J., 2018. A quantitative metabolomics study of bacterial metabolites in different domains. *Anal. Chim. Acta* 1037, 237–244.
- Zhou, Z., Shen, X., tu, J., Zhu, Z.-J., 2016. Large-scale prediction of collision cross-section values for metabolites in ion mobility — mass spectrometry. *Anal. Chem.* 88.
- Zhuang, K., Vemuri, G.N., Mahadevan, R., 2011. Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* 7, 500–500.