

# Metabolomics and flux balance analysis

Priyanka Narad, G. Naresh and Abhishek Sengupta

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida, India

## 21.1 Introduction

Living organisms have complicated metabolic systems, composed mainly of numerous enzymes catalyzing enormous reactions. With the advent of bioinformatics, there is a huge possibility to completely analyze the entire genomes of the organism by fabricating metabolic networks of many organisms through full system models (Xia, Broadhurst, Wilson, & Wishart, 2009). These models contain lot of information of chemical and metabolic reactions occurring within the organism along with their compartmentalized location within the organelles of the cells (Xia et al., 2009). Besides, it contains information about transport reactions occurring between intracellular compartments and between the organism and its environment. Since there are more than 1200 transport and metabolic reactions occurring in simple unicellular organisms, such as *Escherichia coli*, it is computationally challenging for transient modeling of these systems (Orth et al., 2011). It is also a herculean task to analyze the in vivo reaction rates with great precision and confidence since these intracellular metabolites are often present at quite low concentrations to be accurately measured. Metabolomics is an emerging field that aims to quantify and identify cellular metabolites through the applications of sophisticated analytical technologies. It uses various statistical and multivariate methods for processing, extraction, and interpretation of data (Putri et al., 2013). All of these omics, such as transcriptomics, metabolomics, proteomics, and genomics, are some of the important tools to understand the biology of an organism and its response to genetic perturbation and environmental stimuli. This field is emerging as a genetic architecture for broadcasting signals from the surrounding environment. Methods, such as nuclear magnetic resonance (NMR), mass spectrometry (MS), and chromatographic, such as liquid chromatography (LC) and gas chromatography (GC), are used to analyze the metabolic data. Since metabolomics data are obtained in the large amount due to large differences in abundance and increased diversity of chemical structures, there has not been a single technology to analyze the entire metabolomes (Putri et al., 2013). Therefore a suitable technique has to be developed for the identification, detection, extraction, and quantification of metabolites. To deal with large sets of data, methods, such as heatmaps, statistical tools, comparative overlays, multivariate data analysis tools, pathway mapping, and cluster analysis, have been developed to distinguish between real sample information and noise (Orth, Thiele, & Palsson, 2010). Flux balance analysis (FBA) is the most widely used modeling technique that helps to overcome these issues. There are numerous biological and mathematical models developed to identify the capabilities of metabolic networks based on kinetic and stoichiometric details. The new hypothesis can be created from the predicted outcomes and used within metabolic networks to optimizing pathways (Kuang, Marney, Cuevas, Edwards, & Forsberg, 2020). On the other hand, metabolic flux analysis (MFA) can be used to predict single solutions; however, lot of information is required for governing enzymes in the organism. Since it requires lot of kinetic information that is not yet available, this becomes a major drawback of this method. Carbon flux analysis is the second approach, which is based on the mathematical and biological modeling of  $^{13}\text{C}$ -labeled molecules. Predictions would be made according to the measurement of the fluxes within the model where the molecules are located (Wiechert, 2001). The third approach is FBA, which is a constraint-based approach. Depending on the stoichiometric information of the model, FBA creates a solution space. With the aid of environmental and biochemical constraints, the solution space can be reduced (Shlomi, Berkman, & Ruppin, 2005). Flux distribution can be analyzed using the objective function as the optimization technique. Besides these, there are other gene regulation methods developed to reduce the solution space and the outcomes can be predicted under various conditions.

## 21.2 Definition of metabolomics

Metabolomics is a branch of science that deals with the complete set of metabolites within a biological sample. It mainly focuses on the characterization of water-soluble metabolites (Xia et al., 2009). Since it is related to genomics and proteomics, it is considered as an essential resource for systems biology. Biological function is regulated by up- and downregulation of proteins and genes. Thus a complete set of system's behavior can be obtained from the measurement of metabolomics, proteomic, and transcriptomic data. Biological samples are collected for metabolomics studies from two or more experimental groups and the metabolites are extracted. They are then estimated using various instrumental techniques, such as NMR, MS, and LC–MS. After this, metabolites are distinguished and metabolic pathways are constructed based on the obtained data. This technique is also used to estimate the untargeted metabolites under various genetic and environmental stresses. In contrary to this, targeted metabolomics are analyzed from a completed set of specific metabolites based on available knowledge about the system (Nalbantoglu, 2019). Thus targeted metabolomics differ from untargeted metabolomics on the basis of higher quantitative and sensitive data with a lower false positive rate and higher reproducibility. Since enzymes have a rapid turnover rate with respect to chemical stability and temperature, it affects the metabolomics samples. It is essential to optimize the protocol of sample preparation (Nalbantoglu, 2019).

It is an evolving branch that aims to identify and quantify the exogenous and endogenous metabolites of low molecular weight using high-throughput techniques in a biological system (Putri et al., 2013). The relationship between genotype and environment can be determined by metabolomics as it is closely related to physiology and genotype. It is the study of the whole metabolome of an organism and is widely applicable in the field of personalized and molecular medicine, toxicology, clinical chemistry, new-born screening, transplant monitoring, and pharmacology (Putri et al., 2013). The metabolome is a term comprised of small molecules whose interactions within a biological system bring environmental, genetic, and nutritional changes. The metabolome represents the final downstream product of an organism that includes protein expression, gene expression, and the related changes (Putri et al., 2013). It represents the molecular phenotype of disease and health. Human Metabolome Database (HMDB) is a repository comprised of vast information of diverse compounds, such as organic acids, nucleotides, lipids, steroids, carbohydrates, and amino acids, including both of which are lipid and water soluble metabolites (Putri et al., 2013). The metabolomics is widely divided into two categories: “targeted” and “untargeted.” Exometabolome and endometabolome are also included in the systematic quantification and identification of metabolites. Untargeted metabolomics is discovered by a hypothesis-driven technique that allows full scanning of metabolome also known as metabolic fingerprinting and identification of pattern. Targeted metabolomics is based upon hypothesis testing and it is mainly used to validate the untargeted analysis (Toya & Hiroshi, 2013). The hypothesis-driven metabolomics is divided into finger printing, nontargeted profiling, and foot printing. Nontargeted metabolomics is also referred by comprehensive small metabolite/molecule analysis. It identifies features putatively by semiquantitative analysis. Metabolomics fingerprinting is a method of generating a snapshot of the entire intracellular metabolome of an organism, classifying, and screening them (Toya & Hiroshi, 2013). In contrast to this, metabolic foot printing refers to the global snapshot of the entire extracellular metabolome of an organism. To obtain an absolute and accurate quantification and identification of metabolites, both targeted and untargeted approaches should be performed (Toya & Hiroshi, 2013).

## 21.3 MS- and NMR-based techniques in metabolomics

The type of instrumental platform is considered to be one of the most important choices in metabolomics. Among numerous instruments, such as infrared, Raman spectroscopy, capillary electrophoresis, MS, and NMR, are widely used in metabolomics. NMR generates results with more reproducible and quantitative outcomes, and moreover, the sample preparation protocols for the sample are low (Riekeberg & Powers, 2017). However, it has certain limitations, such as low sensitivity toward measuring the abundant metabolites in the sample. Ionized metabolites cannot be easily detected by MS detectors as they rely on ionization processes (Riekeberg & Powers, 2017). Contaminants present within the sample may affect the reproducibility of MS data as they may change the efficiency of ionization of metabolites. Coupling of MS with a chromatographic method, such as LC and GC, may help to overcome the above limitations by resolving overlapping peaks and it helps to identify metabolite based on the properties of stationary phase and the retention time. GC–MS is widely used to measure volatile metabolite mixtures and it requires a minimal amount of sample for operation (Riekeberg & Powers, 2017). Reliability of GC–MS on analyte volatility is the known disadvantage that may lead to the destruction of metabolites with low temperature stability and low volatility. Although derivatization is a time-consuming protocol, it overcomes the issues of limited metabolite volatility. The concentration of metabolites

thus can be perturbed due to the differences in the stability of derivatized metabolites (Sailwal et al., 2020). This may lead to an erroneous biological result. LC is considered to be a more suitable method than GC because it makes the metabolites intact and does not modify the concentration of metabolites. LC generates accurate results of reactive and thermally unstable metabolites. However, it brings drastic changes in retention times due to the introduction of a liquid phase (Sailwal et al., 2020). While MS identifies low abundance metabolites, NMR identifies changes in metabolic trends across the core metabolic pathways. Thus synergetic use of MS and NMR will provide large amounts of dataset size that adds complexity to analysis, processing, and interpretation of the data types (Sailwal et al., 2020).

## 21.4 Data processing in metabolomics

The use of advanced and specialized tools is required for handling large data sets of metabolomics experiments. Various statistical analysis and data processing software packages are used for processing these data sets. Statistical analysis of MS and NMR data requires a unique set of algorithms and processing tools for modeling. To combine the datasets of MS and NMR, a single coherent approach is used. First, the samples are analyzed independently by each method (Sailwal et al., 2020). The data sets generated by each method are then compared for identifying consistencies in the result. Second, the data sets generated by each method are then integrated using multiblock analysis into a single statistical model. The multiblock analysis comprised of a variety of methods to combine the multiple data sets generated by MS and NMR. It is also used to combine datasets obtained from different omics disciplines (Sailwal et al., 2020). It tends to perform independent analysis and produces results with greater resolving power and predictive ability. However, the multiblock analysis relies on crude and custom sets of preprocessing data obtained from multiple software packages (Sailwal et al., 2020).

### 21.4.1 Nuclear magnetic resonance spectroscopy

NMR works on the magnetic properties of special nuclei that possess spin, for example,  $^{13}\text{C}$  and  $^1\text{H}$ .  $^1\text{H}$  is widely used in metabolomics. Superconductors are used in NMR to generate strong magnetic fields. This leads to the formation of two spin states, such as one aligned with the magnetic field and another aligned against the magnetic field. The two spin states produce energy differences (Sailwal et al., 2020). When the sample is labeled with these nuclei, it gets excited with a radio frequency pulse. The nuclei with a lower energy spin state get excited into a higher energy state and a resonance state is achieved due to energy differences, which give chemical structure (Sailwal et al., 2020). The reference signal is created by tetramethylsilane and 3-(trimethyl-silyl) propionic acid. Chemical shifts are generated from the local electron shielding. The resonance signals of NMR are displayed proportional to the concentration of the sample.  $^1\text{H}$  NMR has several advantages, such as nondestructive, noninvasive, little sample preparation, robust, reproducible, and quantitative (Sailwal et al., 2020).

### 21.4.2 Workflow of MS-based metabolomics

MS procedure is widely accepted as compared to NMR because of its ability to detect metabolites at very low concentrations even at femto to atto molar range. It is used to identify and quantify metabolites with high sensitivity, resolution, and dynamicity. The following are the steps involved in MS-based analysis of metabolites: preparation of the sample, separation of the sample by chromatographic, capillary electrophoresis and extraction, ionization of sample, conversion, and detection of metabolites with respect to their mass to charge ratio ( $m/z$ ) (Sailwal et al., 2020).

#### 21.4.2.1 Sample preparation

Biological samples, such as cell culture, tissue, and biofluids, such as saliva, blood, bile, urine, and seminal fluid, can be used for metabolome analysis. It is crucial to do sample preparation carefully along with its management in the form of bio-banking and biorepositories with proper labeling. It is required to generate high-throughput, reproducible, optimum output with enriched metabolite coverage and recovery (Sailwal et al., 2020).

#### 21.4.2.2 Extraction

The prepared samples are extracted using nonpolar and polar compounds based on targeted and untargeted approaches. Methanol–water–chloroform is an approach widely used to extract hydrophobic and hydrophilic compounds. The mixture is then subjected to centrifugation, which yields a biphasic mixture of lower (organic) and upper (aqueous) layers.

Other polar organic solvents, such as a mixture of methanol or acetonitrile with water, are also used. Dichloromethane is used to perform organic extraction (Sailwal et al., 2020).

#### 21.4.2.3 Separation

Separation of the two phases is done using a variety of techniques, such as GC and LC coupled with the MS system. Sometimes the mixture is injected into real-time MS. For polar metabolites, capillary electrophoresis coupled with MS is used for separation and profiling in biological samples. For nonpolar metabolites, reversed-phase LC with C18 columns is used (Sailwal et al., 2020). GC/MS or LC/MS is preferably used for targeted and untargeted metabolomics, which produces high chromatographic resolution. GC/MS is used to analyze volatile organic compounds (VOCs) in biological samples, such as organic acids and fatty acids. VOCs can be separated by the solid phase microextraction technique (Sailwal et al., 2020). It separates organic compounds from solid, aqueous, and gaseous materials. While GC–MS is used to analyze less polar biomolecules, such as perfumes, lipids, waxes, and essential oils, LC–MS is used to analyze more polar biomolecules, such as nucleotides, organic acids, and polyamines. Both GC–MS and LC–MS are used to analyze amino acids, steroids, fatty acids, alcohols, etc. (Sailwal et al., 2020).

#### 21.4.2.4 Ionization

The samples are ionized using positively or negatively charged ions in the gas phase. It is done by injecting dry nitrogen and heat followed by evaporation (Sailwal et al., 2020). The samples are then converted into positively and negatively charged ions. Depending upon the polarity of the molecules, various types of ionization sources are used, such as electrospray ionization sources, matrix-assisted laser desorption ionization, and chemical ionization.

#### 21.4.2.5 Detection

The ions are detected by high resolution mass analyzers depending on the mass to charge ratios of the fragmented ions. Ion trap, time of flight (TOF), quadrupole TOF, etc., are some of the mass analyzers widely used. It produces a mass spectrum with high resolution, sensitivity, and accuracy (Sailwal et al., 2020).

#### 21.4.2.6 Analysis and identification

The enormous amount of raw data extracted from MS contains specific metabolic signals, which can be analyzed using specialized software for the data interpretation and identification of metabolite of interest. Free available software and tools are used for the processing of data, assessment, and their quantification (Sailwal et al., 2020). Preprocessing of raw data, such as retention time correction, peak detection, spectral filtering, peak alignment, noise elimination, and peak normalization, is required. The data are then prepared for integrity checking, identifying compound name and normalization using clustering, statistical, and multivariate and univariate analyses (Sailwal et al., 2020). Following all these steps, processes, such as enrichment analysis, pathway mapping, pathway analysis, and functional interpretation, are performed. Tools, such as MetaCore, 3Omics, Progenesis, and MetaboAnalyst, can be used for processing raw mass spectrum data and connecting with the databases, such as Chemical Entities of Biological Interest, In Vivo/In Silico Metabolites Database, Madison Metabolomics Consortium Database, HMDB, Kyoto Encyclopedia of Genes and Genomes (KEGG), Metabolite and Tandem MS Database, Reactome, BioCyc, ChemSpider, PubChem, MetaCyc, and Metabolomics Workbench MetaboLights (Sailwal et al., 2020).

### 21.4.3 Limitations of NMR and MS methods

Although MS and NMR are widely used in metabolomics, they have certain limitations that should be taken into account (Xia et al., 2009). The number of samples detected by NMR is lesser than MS because of its low sensitivity. Hence, in lipidomics, NMR has less application due to insufficient resolution. This can be overcome by utilizing higher magnetic field strength (Riekeberg & Powers, 2017). MS when coupled with chromatography would cause moderate reproducibility of the sample. Unlike NMR, the sample cannot be used for multiple analyses and cannot be recovered. The analysis time for the sample varies with the protocol used from few minutes to hours. Thus it is time consuming as compared to NMR. It is costlier than NMR as it requires sample preparation and separation techniques (Riekeberg & Powers, 2017).

## 21.4.4 Recent advances in MS- and NMR-based metabolomics

### 21.4.4.1 Dynamic nuclear polarization

To overcome the problem of overlapping in 1D spectra, 2D NMR methods are required to process data with a high confidence level in metabolite identification. Hence, isotopic labeling of  $^{13}\text{C}$  and  $^{15}\text{N}$  is required as they have a low natural abundance. Thus dynamic nuclear polarization (DNP) has emerged as a promising structural biology tool to process data of NMR in metabolomics (Xia et al., 2009). In this method, a frozen, solid sample is polarized at 1.5K temperature with the aid of microwave irradiated free radicals. This leads to the introduction of temporary hyperpolarization in the nuclei with an active spin state by transferring electrons to nuclei polarization. It is essential to rapidly melt the sample and transfer it to a NMR spectrometer (Xia et al., 2009). This increases the sensitivity of the sample up to 10,000-fold. Such a drastic increase in sensitivity curbs the requirement of isotope labeling for in vivo samples. It also aids in the detection of metabolites with low abundance. However, there are certain limitations associated with the DNP methods, such as limitations by T1 relaxation rates, the requirement of advanced hardware accessories, and modifications (Xia et al., 2009).  $^{13}\text{C}$  MFA is done with the DNP method that uses tracer elements for in vivo imaging. Despite the limitations in sample preparation and reproducibility, DNP is emerging as a promising tool for metabolomics.

### 21.4.5 Challenges and affecting factors

Challenges in the arena of data processing, analysis, pathway analysis, and identification of metabolites need to be taken into account even though extensive tools and databases are available (Nalbantoglu, 2019). This requires high-throughput bioinformatics tools and computational techniques for metabolomics profiling. Since MS uses an ionization process, it affects MS signals, identification of compounds, chromatographic resolution, etc. Isomers with identical masses are another challenge to distinguish between the samples (Nalbantoglu, 2019). To maintain the robustness, integrity, and sensitivity of the results, accurate monitoring of retention time points, MS signals, and chromatographic peaks is required. Another problem is the contamination due to matrix compounds during sample preparation. Control samples and batch samples need to be labeled properly and randomization has to be performed on both (Nalbantoglu, 2019).

## 21.5 Applications of metabolomics

Metabolomics has proven to be widely applicable in various fields, such as animal, plant, and microbial systems, medicine, synthetic biology, and food sciences (Putri et al., 2013). The various applications of metabolomics are discussed below:

### 21.5.1 Microbial science

Microbes are an important source of metabolomics as they can be easily modified according to the experimental procedures. However, microbial metabolomics requires high resolution analysis, controllable environmental conditions for metabolite measurement, and sample preparation. The two essential processes are extraction and quenching in sample preparation for microbial metabolomics. The process of extracting metabolites from the cells is called extraction and the process of halting biological reactions in the cells is called quenching. For producing results with high reproducibility, quenching of sample is done at a specific point to estimate the actual quantity of metabolites at a given time. Quenching is validated by two major factors—stopping of biological reaction at a short time frame and limited leakage of metabolite. The microbial cells are subjected to the extraction process depending upon the robustness of the cell membrane, cell properties, reactivity of enzymes, and chemical properties of the target analyte. Depending upon the ability of microbial cells to survive the harsh conditions, high temperature, methanol, and chloroform or free thaw is applied. Microbial metabolomics is generally carried out by stable isotopes. For example, in *Saccharomyces cerevisiae*,  $^{14}\text{C}$  glucose was used to study the relationship between global regulation and cellular metabolism.  $^{13}\text{C}$ -labeled intermediates have also been developed to study the core central metabolism using MFA. Thus microbial metabolomics will be useful for the study of higher organisms.

### 21.5.2 Plant science

Metabolomics has a vital role in plant science to decode the functions of many essential genes and understand the complex cellular systems (Toya & Hiroshi, 2013). Transcriptional regulation in plants plays a key role in determining the

metabolic state of plants under various developmental and environmental conditions. However, the mechanism underlying the regulatory functions in gene expression and metabolic phenotype is still elusive (Toya & Hiroshi, 2013). Hence, metabolomics helps botanists to study in detail the dynamic behavior of plant metabolic systems. Accumulation of metabolites and expression of genes can be determined by plant metabolomics. AtMetExpress dataset was developed to study the metabolic accumulation and gene expression in the *Arabidopsis thaliana* plant (Putri et al., 2013). It was reported that the genomic sequence of *A. thaliana* contained many essential metabolic genes for the production of commercially important phytochemicals. The data revealed that the plant contained 167 metabolites and 1589 metabolic signals for various phytochemicals. It has been postulated that transcription of metabolites and their regulation determines the diversity of secondary metabolism of plants and the origin of dynamics (Putri et al., 2013). The application of metabolomics to breeding and crop sciences is also commendable. To improve the nutritional value of the crops, genetic factors playing a vital role in regulating metabolic levels should be determined. Since the metabolism in plants is controlled by the relationship between biomass/yield and the composition of metabolite, metabolomics is widely used to study these relationships (Putri et al., 2013). Besides these, metabolomics plays a pivotal role in the efficient production of genetically modified crops and the degree of risk management required. The technological advancements in metabolomics have helped scientists to produce phytochemicals in large amounts for various purposes. The plant metabolomics requires three important strategies—MS spectral data for metabolite structural elucidation, a method to estimate false discovery rates and a comprehensive mass spectral database of phytochemicals (Toya & Hiroshi, 2013).

### 21.5.3 Animal science

Metabolomics technology has proven as a useful technique to study the biological phenomena of an essential model organism, such as fruit fly and zebrafish. A comprehensive study of metabolism occurring in these organisms may provide intense information on pathological, physiological, and developmental processes (Riekeberg & Powers, 2017). Zebrafish fish or *Danio rerio* is a commonly used model organism to study the biomedical, behavioral, and biological processes occurring during organogenesis and embryogenesis in the development of vertebrates. It has been widely used for research in drug and disease discovery as it can be easily bred in large amounts with comparatively low maintenance cost to other model organisms (Riekeberg & Powers, 2017). With the aid of the metabolomics technique, it is possible to determine the relationship between embryogenesis and metabolome, which serves as a fingerprint for analyzing the developmental process, thereby aids in drug treatments. Another model organism *Caenorhabditis elegans* is widely used for the study of lifespan, physiology, genetic diseases, pharmacological studies, aging, and drug toxicity screening (Riekeberg & Powers, 2017). *Drosophila melanogaster*, also known as fruit fly, is widely used for studying physiology and genetics of aging because of its short lifespan, similarities with human aging, and ease to breed. It has been used as an ideal model organism for the metabolomics study of developmental biology, effects of phenobarbital, pesticide resistance, oxidative stress, etc. It can withstand the extreme conditions of hypoxia, cold shock, and mutations (Putri et al., 2013).

### 21.5.4 Medical science

Metabolomics has widespread applications in the medical field. It is widely used to determine the mechanism of drug action and to explore biomarkers present in the bodily fluids. Metabolomics is widely used in drug treatment and medical therapy by monitoring the changes in metabolites in the biofluids (Toya & Hiroshi, 2013). Since metabolites act as biomarkers for diseases, the change in their concentration in bodily fluids indicates a diseased state. Thus multiple metabolites provide information about treatment response with high selectivity and sensitivity. Moreover, metabolomics is widely used to predict the changes in the response of a drug to a particular disease and to estimate the prognosis of disease progression. It is also used to predict the alternative treatments of a patient (Toya & Hiroshi, 2013).

Metabolomics has tremendous applications in disease and health monitoring, such as single cell, metabolic phenotyping, personalized medicine, precision metabolomics, metabolome-wide association studies (MWAS), and epidemiological population studies. It helps in the identification and expression of metabolites, small molecules, and high-dimensional characterizations of individual cells. It has been reported that metabolomics is widely applicable in clinical biomarkers identification through metabolomics fingerprinting, profiling, footprinting, MWAS, etc. (Putri et al., 2013).

It is also used to study numerous metabolic syndromes, such as life-threatening diseases caused by sugar and lipid metabolism, for example, cardiovascular diseases, cancer, and cerebrovascular disease. It is used in the pathophysiological study of metabolites and biomarkers for the early diagnosis of fatal diseases. Trimethylamine oxide, which is a biomarker of cardiovascular diseases, was detected by LC–MS-based metabolomics technique (Toya & Hiroshi, 2013). It

is a potential biomarker associated with diseases, such as myocardial infarction, coronary, and peripheral artery disease. It is widely used for the study of cancer, its early diagnosis, and its efficient prognosis.  $^{13}\text{C}$  stable isotope-based metabolomics technique has been used for research on lung cancer cells using  $^{13}\text{C}$  glucose as tracers to determine metabolic flux. It revealed overproduction of alanine, lactate, and glutamine, which are major agents for causing cancer (Putri et al., 2013). Thus isotopomer-based metabolomics is used to understand biochemical changes in cancer cells. It is a highly sensitive, noninvasive, and promising diagnostic tool for cancer. Besides these, it is used to understand diseases of the central nervous system and psychiatric disorders (Putri et al., 2013).  $^1\text{H}$  NMR-based metabolomics was used to study cerebrospinal fluid as it serves as a promising biomarker for neural metabolomics. It is also used to study Parkinson's disease and Alzheimer's disease. Psychiatric diseases, such as depression and schizophrenia, can also be studied with the aid of neural metabolomics. Schizophrenia is caused due to changes in neurotransmitter systems and neuronal membrane phospholipids (Putri et al., 2013). Since alterations in lipid metabolism lead to schizophrenia, lipidomic analysis is widely used to identify the potential biomarkers causing pathogenesis. Thus metabolomics is used in early diagnosis, drug toxicity, biochemical alterations in mood disorders, and therapeutics. A better understanding of the linkage between pathological conditions and the changes occurring in the body can be obtained by integrating metabolomics with other omics techniques (Putri et al., 2013).

### 21.5.5 Food and herbal medicines

Metabolomics has emerged as a promising technique for perusing the safety and quality of food and herbal medicines. Processes, such as genetic modifications, milling, and atmospheric storage, affect the quality of food products. Metabolomics is used to control the quality of final food products and ensure the safety of the herbal medicines by a method called sensory evaluation (Putri et al., 2013). It is a scientific method used to analyze, evoke, interpret, and measure the product quality through a canonical sense of touch, hearing, sight, taste, and smell. It has great importance to monitor the quality and prices of food products in the food industry. MS-based metabolomics tools are used to evaluate the quality of food products, such as fruits, cereals, crops, and beverages. Food metabolomics is categorized into two concepts- organizing taste active compounds and mimicking human senses (Putri et al., 2013). A branch of food metabolomics called sensomics is used to mimic the human sense of hearing, taste, sight, smell, and touch for evaluation of food quality. Techniques, such as NMR and GC-MS, are employed in food metabolomics. These are also used in herbal medicinal research, industrial and pharmaceutical purposes. They are used in the evaluation of toxic and pharmacological effects. Therefore metabolomics is emerging as a robust, reliable, sensible, and promising tool in quality control and sensory chemistry (Putri et al., 2013).

## 21.6 Flux balance analysis

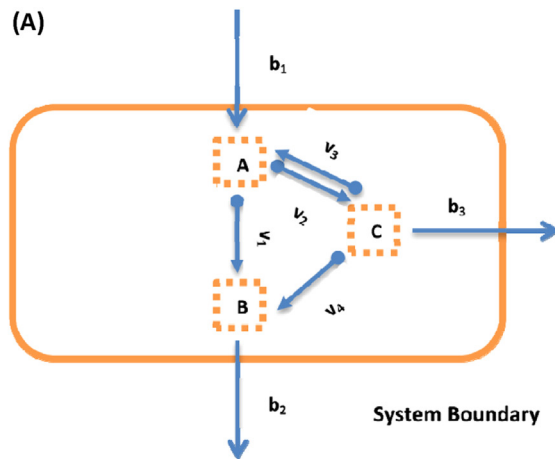
Numerous mathematical models are used to model metabolic networks that are based upon the calculation of a single solution describing all the fluxes flowing through a metabolic network (Kauffman & Edwards, 2003). FBA is a constraint based analysis that provides a basic understanding of the structure and function of a metabolic network. These constraints are represented in form of mathematical expressions (Kauffman & Edwards, 2003).

Fig. 21.1 depicts how a metabolic network can be analyzed using FBA. With the knowledge of the reactions involved, a set of differential equations can be formulated describing cellular metabolism. The following is the general of the equation (Kauffman & Edwards, 2003):

$$\frac{dC_i}{dt} = \sum_j^n a_j v_j = \sum_j^n k_j C_i^{a_i} \prod_k^m C_K^{a_k} \quad (21.1)$$

where  $C$  represents the concentration of species  $i$  and is time dependent and is the sum of the  $n$  number of reactions' fluxes  $a_j v_j$ . which is defined as the product of the reaction rate constant  $k_j$  and the  $m$  number of stoichiometrically weighted species involved in the reaction. Since these equations are quite complex to measure kinetic constants, a computational solution is required (Kauffman & Edwards, 2003).

Where, substrate concentrations are represented by A, B, and C. Fluxes related to exchange of substrates and transformation of substrates into products from in out of the cellular compartments via cell membrane is represented by symbols  $b_1$ ,  $b_2$ , and  $b_3$ . Fluxes  $v_1$ ,  $v_2$ , and  $v_3$  are internal substrate fluxes. This problem can be overcome using FBA. It assumes that the internal metabolites will be at a pseudo steady state when they are constant environmental conditions, being consumed and produced at the same rate (Covert, Schilling, & Palsson, 2001). Stoichiometries of these cellular



**FIGURE 21.1** Construction of a metabolic model by (A) designing a metabolic network and (B) transforming the network into a set of differential equations.

(B)

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

← s

reactions impart constraints to the distribution of metabolites in the system (Covert et al., 2001). This overcomes the problem of transient ordinary differential equations (ODEs) by introducing linear algebraic equations. These sets of reactions are classified in the form of a matrix  $Sv = 0$ . Here,  $S$  is a matrix representing stoichiometric coefficients and  $v$  represents the vector of fluxes of associated reactions (Covert et al., 2001).

These sets of reactions consist of growth equations representing the final consumption of metabolites necessary for cell replication (Lee, Gianchandani, & Papin, 2006). There are other equations representing the uptake of essential nutrients and excretion of molecules to the extracellular space. The flux of metabolites throughout the network of reactions is constrained by these equations. If the system consists of more reactions than metabolites, it is known as underdetermined reactions. Hence, it becomes arduous to estimate the accurately low concentration of these metabolites (Lee et al., 2006).

## 21.7 Metabolic networks and model construction

The metabolic network of an organism tells us the information about the construction of products from their corresponding substrates. Flux distributions cannot be fully described by the genomic data obtained to construct the metabolic network (Papin et al., 2004). The rate of all the fluxes flowing through the metabolic network is represented by a vector. A simplified representation of a metabolic network is shown in Fig. 21.2.

It is simple to determine the exchange fluxes. They are represented in the above diagram as the reaction of the transformation of sugar to ethanol and carbon dioxide (Segre, Vitkup, & Church, 2002). However, it is quite challenging to determine the internal fluxes since they are difficult to measure without interrupting the biological processes occurring within the cells as they are hidden inside the cell (Segre et al., 2002). Hence, with the aid of computational modeling and mathematical simulations, we could predict and interrupt the metabolic flux distributions flowing through a metabolic network. Detailed concentration and kinetic information are required to be included in most of the existing methods, such as various cofactors and enzyme activity (Segre et al., 2002).

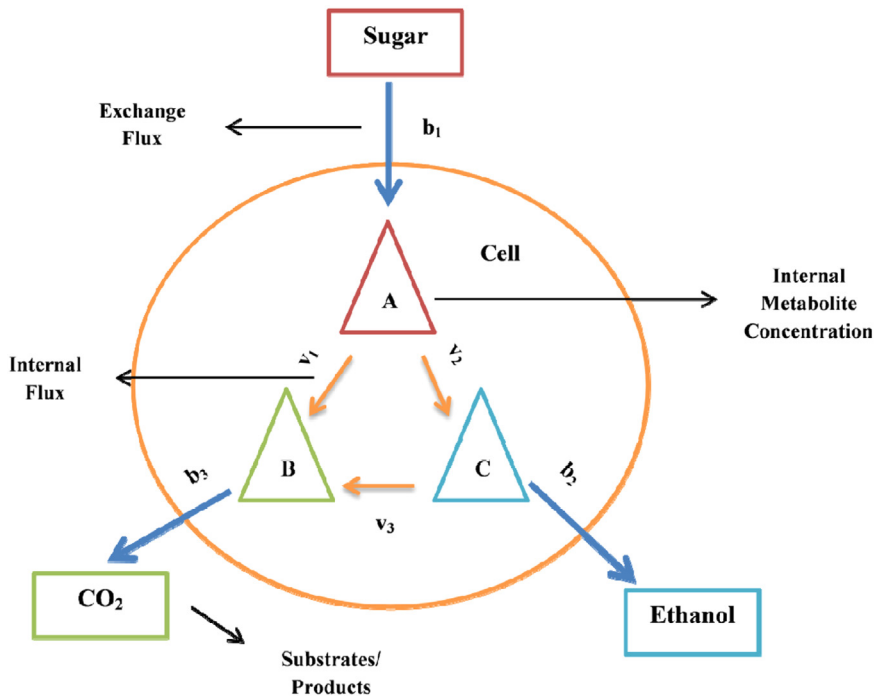


FIGURE 21.2 Representation of construction of the metabolic network.

### 21.7.1 Metabolic model: construction and refinement

#### 21.7.1.1 Initial model development

The construction of a metabolic network requires the following information: (1) reactions being catalyzed by the enzyme, (2) the stoichiometries of the products and reactants for each associated reaction, (3) the localization of the reaction to the cellular compartment, and (4) the reversibility of the reaction (Mahadevan, Burgard, Famili, Van, & Schilling, 2005). In addition to this information, it is mandatory to identify the transport reactions representing cellular uptake and excretion mediated by membrane proteins (Mahadevan et al., 2005). The information of the enzymes involved in these reactions is obtained from the annotated genomic sequences. With the advent of Homology searches, it has become possible to compare known genes to the unknown to identify the enzymes involved in metabolism. Table 21.1 shows the list of some of the useful databases of metabolic and genomic information for model construction (Mahadevan et al., 2005).

While constructing the model, it is significant to figure out the relationships from genes to enzymes to reactions. Depending upon the enzymes the genes encode, reaction rules provide the framework for connecting genes to reactions (Tomar & De, 2013). The expression of individual genes to the reaction in the network is indicated by Boolean statements. The basic reaction rule is an one-to-one true/false connection between a gene and the reaction the associated enzyme performs (Tomar & De, 2013). It works in the following way: the reaction is available in the model if the gene is expressed and vice versa happens when the gene is not active due to genetic modifications, such as mutations, transcription regulation, and knockout. Isozymes are the enzymes that perform the same reaction on the same substrate in the same cellular compartment. “Or” statement indicates the presence of isoenzymes in the reaction (Tomar & De, 2013). If any single gene or multiple isoenzymes are present in a reaction, the reaction is included in the model network. The remaining reaction rule deals with enzyme complexes. Enzyme complexes are formed by a group of proteins that together perform a specific function. If any of these enzymes are absent, then the functional activity of the enzymatic complex is altered. “And” statement indicates types of associated reactions. It is essential to include all the associated enzymes in the model (Tomar & De, 2013).

The following are the processes involved in the reaction networks:

- Substrate uptakes from the surroundings.
- Substrate catabolism and formation of intermediate metabolites.
- Conversion of these metabolites into biologically essential molecules (Wiechert, 2002).

**TABLE 21.1** Databases and tools for metabolic network analysis and construction.

S. no.	Databases/tools	Description	References
Tools			
1.	MetaFluxNet	Metabolic flux analysis	Lee, Yun, and Park (2003)
2.	Yana	Reconstruction, analysis, and visualization of networks	Schwarz et al. (2005)
3.	System Biology Research Tool	Analysis of stoichiometric networks through multiple methods	Wright and Wagner (2008)
4.	Constraint Based Reconstruction and Analysis Toolbox	Works with MATLAB® for metabolic network analysis, gene deletions, etc., using FBA	Becker et al. (2007)
5.	PathwayAnalyzer	Uses MoMA, FBA for gene deletion studies and metabolic networks	Raman and Chandra (2009)
6.	SBML Software Guide	Model databases	Hucka et al. (2015)
7.	CellNetAnalyzer	Databases of functional and structural analysis	Klamt, Saez-Rodriguez, and Gilles (2007)
8.	SNA—Stoichiometric Network Analysis	Mathematic toolbox for metabolic networks	Urbanczik (2006)
9.	GNU linear programming toolkit	Solves LP problems	Lee, Palakornkule, Domach, and Grossmann (2000)
10.	LINDO	Solves optimization problems	Lee et al. (2000)
11.	CPLEX	Optimization software package	Savinell and Palsson (1992)
Databases			
1.	BRENDA	Information of molecular and biochemical pathways on enzymes	Schomburg, Chang, and Schomburg (2002)
2.	BioCyc	Databases of pathways for several organisms	Karp, Billington, and Caspi (2019)
3.	Reactome	Curated databases of biological processes in humans	Fabregat et al. (2017)
4.	PEDANT	Genome annotations	Riley et al. (2007)
5.	Biomodels	Databases of kinetic models of pathways	Li et al. (2010)
6.	Kyoto Encyclopedia of Genes and Genomes	Databases of pathways for several organisms	Kanehisaa and Goto (2000)
7.	SABIO-RK Database	Database for the analysis of biochemical reaction kinetics	Wittig, Kania, and Golebiewski (2012)
<i>FBA</i> , Flux balance analysis.			

A biomass reaction equation is used to represent the reproduction of organisms in the model after consumption of metabolites, which is denoted in the form of a stoichiometric ratio of all necessary metabolites, such as nucleic acids, proteins, and lipids (Wiechert, 2002). The catabolism of macromolecules into essential metabolites is represented in the growth equation as being consumed from their general pools. To estimate the performance of the model, it is significant to determine an accurate biomass equation (Wiechert, 2002).

Genomic data render a framework for constructing the metabolic model. However, there might be some errors regarding substrate specificity, the activity of the enzymes, etc., which may lead to network gaps. Hence, a thorough literature search is required to avoid any erroneous results (Wiechert, 2002).

### 21.7.1.2 Optimization method to model the system

In this method, mathematical methods of optimization are used to determine a unique solution. An objective function can be maximized or minimized using optimization by manipulating certain bound constraints (Zomorodi, Suthers,

Ranganathan, & Maranas, 2012). The optimum values of the objective function and its variables are required to define the optimum state. The general problem in the optimization of the metabolic model appears as follows:

$$\begin{aligned} \max c_v^T v \\ \text{s.t. } S \cdot v &= 0 \\ v_i &\geq 0 \\ v_i &\leq \text{UB} \\ b_i &\geq \text{LB} \\ b_i &\leq \text{UB} \end{aligned}$$

where LB represents lower bound limits, UB represents upper bound limits,  $v_i$  represents a vector of species  $i$ ,  $S$  represents stoichiometric limitations, and  $b_i$  represents bound constraints of species  $i$  (Zomorodi et al., 2012). These variables represent the fluxes through transport reaction and each metabolite. These constraints are divided into equality and inequality equations, those representing stoichiometric limitations ( $S$ ) are represented as equality constraints, which consist of growth and metabolic reactions, and those representing upper and lower bounds of the system are represented as inequality constraints (Zomorodi et al., 2012). A maximum feasible substrate uptake rate and a minimum flux through an energy consumption equation are some of the examples of these constraints representing maintenance energy required for cellular functions other than growth (Zomorodi et al., 2012). To maximize the growth rate, the objective function is used, which represents the growth of fast growing organisms over a period of time. Objective functions also include maximizing the production of energy or product (Zomorodi et al., 2012). Linear programming techniques can be used to solve the flux distribution of a metabolic model in FBA. There are numerous tools available to perform FBA, such as Constraint Based Reconstruction and Analysis (COBRA) Toolbox for MATLAB and algebraic modeling software GAMS (Zomorodi et al., 2012).

### 21.7.1.3 Model improvement by experimental validation

Growth prediction inconsistencies are the major types of errors found in a model. In such cases, the model conditions according to a particular set of conditions would be contrary to that of the actual experimental results (Reed et al., 2006). For instance, if the model would predict the growth of organisms, then it might be possible to have no growth experimentally (G/NG). These inconsistencies may lead to network gaps, which must be considered to bring the accuracy with the reality. Hence, it is recommended to keep literature and experimental results judiciously to verify the model against experimental results (Reed et al., 2006).

Growth of organisms on different substrates is the first experimental condition to be verified. This is depicted in the model by manipulating various exchange reactions of substrate uptake (Mahadevan et al., 2005). For example, if we would like to stimulate the growth of organisms on sucrose, then substrate consumption reaction would be set up to some real value and other substrate reaction would be set to zero as follows:

$$\begin{aligned} \text{Ex\_Suc}_{\text{LB}} &\geq -10; \text{Ex\_Suc}_{\text{UB}} \leq 0 \\ \text{Ex\_Glu}_{\text{LB}} &\geq 0; \text{Ex\_Glu}_{\text{UB}} \leq 0 \\ \text{Ex\_Fru}_{\text{LB}} &\geq 0; \text{Ex\_Fru}_{\text{UB}} \leq 0 \end{aligned}$$

where  $\text{Ex\_Suc}_{\text{LB}}$  represents the lower bound limits and is set to negative integer for substrate consumption as they are depicted in models as conversion to extracellular from intracellular species and  $\text{Ex\_Suc}_{\text{UB}}$  represents the upper bound limits (Mahadevan et al., 2005).

During substrate consumption, there would be some reactions to deal with the formation of biomass precursors. Thus the model will not predict flux through these reactions. Such paths are called “dead end” metabolites that would be dangerous for the organism if the limit exceeds (Savinell & Palsson, 1992). GapFind is an algorithm used to determine the “dead end” metabolites and GapFill is an algorithm used to determine the modifications needed to reconnect the metabolite with the rest of the network. GapFill algorithm works by four ways: adding known reactions to model organisms found in various databases, adding transport reactions within intracellular compartments, reversing the directionality of the reactions, and adding extracellular uptake reactions (Savinell & Palsson, 1992). Binary variables are used by both of these algorithms, GapFill returns a true value if the particular change is made to the model and GapFind returns a true value if a metabolite contains reactions leading to or away from it (Mahadevan, Edwards, & Doyle, 2002).

#### 21.7.1.4 Thermodynamic constraints to metabolic model

Metabolic models require thermodynamic modalities to simplify the reversibility of the reactions. The more the thermodynamic constraints are applied to the model, the more will be the accuracy of the model predictions (Pandey, Hadadi, & Hatzimanikatis, 2019). Thermodynamic constraints are applied according to Kirchhoff's laws where a change in chemical potential leads to a change in electrical potential difference across the circuit (Pandey et al., 2019). Hence, this application potentially eliminated all the thermodynamically infeasible reactions through the flux, and constraint was applied to the closed reaction loop where a change in chemical potential led to the sum equal to zero for feasible reactions. However, there is an issue with the chemical potential constraints since they are nonlinear (Pandey et al., 2019). Hoppe et al. established realistic reversibility constraints by relating specific metabolic concentrations to the change in Gibb's free energy for the reaction. It deals with the metabolic pool size associated with the reaction and the actual reversibility of the reaction could be determined (Pandey et al., 2019).

#### 21.7.1.5 Regulatory controls and dynamic simulations

In FBA, regulatory controls play an important role in cellular operations. Boolean operators are used to introduce regulation constraints in metabolic models, which represent flux constraints of transcriptional regulation associated with the specific environmental conditions (Ramakrishna, Edwards, McCulloch, & Palsson, 2001). The expression of each gene is analyzed to initiate the Boolean rules. The constraints are then imposed by a standard FBA model to regulate the model and a new set of environmental conditions is determined. The process is then repeated using a new set of environmental conditions each time until an equivalent relationship between metabolic and regulatory systems is attained (Ramakrishna et al., 2001). This eliminates unnecessary fluxes through the pathways to obtain a feasible result. The transcriptional regulatory process is obtained by cycling between these two models through a simulation of time. This approach is applied in *E. coli* and *S. cerevisiae* model to determine the growth on glucose lactose substrate, suppression of other catabolic pathways and amino acids, etc. The ODE method was introduced by Covert et al. to regulate the metabolic model (Ramakrishna et al., 2001). This model provides a framework for the integration of kinetic information into the metabolic model. Probabilistic regulation of metabolism (PROM) is a method developed to automate the regulation of metabolic networks by integrating high throughput data into genome scale (Ramakrishna et al., 2001). However, this approach requires an enormous amount of data, such as the framework of the regulatory network, which is comprised of transcription factors and their targets, a fully reconstructed metabolic network, and gene expression data under variable environmental conditions and gene knockouts. This approach was applied in *E. coli* and *Mycobacterium tuberculosis* models to predict the lethal knockout phenotypes (Radhakrishnan, Edwards, & Doyle, 2002).

### 21.7.2 Mass balance

Mass balance equations are widely used to represent the steady state conditions of metabolic networks. These equations represent the change in concentration on the fluxes over time through the metabolic network (Papin et al., 2004). The change in concentration of the fluxes is determined by the difference between the rate at which metabolite is consumed and the product is formed. It follows the principle of the law of conservation of mass, that is, sum of all the variables internal fluxes equal the external fluxes. An example of a mass balance equation in Fig. 21.2 is given by the following equations:

$$\begin{aligned}\frac{dA}{dt} &= -v_1 - v_2 + b_1 \\ \frac{dB}{dt} &= -b_3 + v_1 + v_3 \\ \frac{dC}{dt} &= -v_3 - b_2 + v_2\end{aligned}\tag{21.2}$$

Substrate concentrations are represented by boxes *A*, *B*, and *C*. Fluxes related to the exchange of substrates and transformation of substrates into products from in out of the cellular compartments via cell membrane are represented by the symbols  $b_1$ ,  $b_2$ , and  $b_3$ . Fluxes  $v_1$ ,  $v_2$ , and  $v_3$  are internal substrate fluxes. A stoichiometric matrix *S* and a flux vector *v* is generally used to represent the mass balance equations. A stoichiometric matrix is generally represented by an  $m \times n$  matrix where *m* denotes the number of metabolites and *n* denotes the number of chemical reactions or fluxes

occurring in the metabolic network (Feist & Palsson, 2010). The relationship between the products and metabolites is represented by the matrix. It helps us to identify the metabolites and their quantity required to create a specific product. The amount of fluxes flowing through the metabolic network is represented by the vector  $v$ . It denotes the rate at which the chemical conversion of one metabolite into another takes place (Papin et al., 2004). The below equation shows the relationship between the stoichiometric matrix and the metabolites (Feist & Palsson, 2010):

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} v1 \\ v2 \\ v3 \\ b1 \\ b2 \\ b3 \end{bmatrix} \quad (21.3)$$

Values other than zeroes and ones can also be used within the stoichiometric matrices. These are used to represent balanced chemical reactions (Feist & Palsson, 2010).

### 21.7.3 Steady state

Steady state is one of the most important assumptions in FBA. This denotes that the concentrations of all metabolites in the model are constant or change in concentration will be zero over time. This assumption must be taken into account since the metabolic transients are more rapid than both dynamic changes in the environment and the cellular growth rates (Schilling, Letscher, & Palsson, 2000). The transients present in the metabolic activities of an organism are extremely rapid, and they lead to steady-state conditions for the entire system in a few seconds. Metabolic changes mediated by gene regulation will be noticed only after a few minutes or hours. This type of assumption is called quasi steady-state approximation, which denotes that the change in concentration of metabolites over time will be approximately zero and is denoted by the equation:  $Sv = 0$ . Thus internal fluxes can now be easily calculated (Schilling et al., 2000).

### 21.7.4 Types of constraints

The constraints are applied to the system by the transporter capacities or enzymes once mass balance equations have been defined. In these constraints, reaction rates and maximum substrate uptake are considered and incorporated into the model. Since the cells in vivo conditions have to undergo lot of constraints, FBA works on these constraints to identify the limitations that the cells have to deal with under various environmental conditions (Covert & Palsson, 2003). Numerous constraints are being developed to use to limit possible solutions. Following are some of the constraints used in FBA.

#### 21.7.4.1 Physico-chemical constraints

These constraints are called hard constraints on cell functions as they will not change according to environmental conditions. Examples of such constraints are energy, mass, and momentum that are conserved in the cell. These constraints will be considered equal to their initial values during the experiment (Covert & Palsson, 2003). Mass balance equations will follow the conservation of mass, and they form important constraints of the model. Same principles are also applied to energy balance equations (Beard, Liang, & Qian, 2002). The determination of chemical reaction rates can be done by the substrate concentration and kinetic properties of the enzymes. As we know that millions of reactions are catalyzed by the enzymes; hence,  $V_{max}$  is used to denote the maximum capacity of an enzyme to catalyze the reactions. At this stage, substrates occupy all the active sites of the enzyme (Beard et al., 2002). The velocity of the enzyme is thus determined by the amount of substrate present in it. Michaelis–Menten constant ( $k_m$ ) is one of the most important parameters used for enzymes, which represent the amount of substrates required for an enzyme to reach half its maximum velocity (Beard et al., 2002). Another parameter to be considered is the enzyme turn over number, which is the number of substrates required to convert into products per unit time when full saturation of the enzyme with the substrate occurs. This number is generally below  $10^4 \text{ s}^{-1}$ . Hence, kinetic parameters and substrate concentration are widely used to determine the chemical reaction rate (Beard et al., 2002).

#### 21.7.4.2 *Topobiological constraints*

These constraints are determined by the topobiological or 3-dimensional structures of the molecules crowding in the cells. An example of this is DNA that is tightly packed within the nucleus. Since DNA stretch is 1000 times larger than the size of the cell, it has to be present in large quantities for transcription (Feist, Herrgard, Thiele, Reed, & Palsson, 2009).

#### 21.7.4.3 *Environmental constraints*

These constraints are dependent on time and environmental conditions, such as temperature, nutrient concentrations, and pH value. In each experiment, the amount of media used and the essential environmental conditions are monitored. Therefore proper conditions need to be maintained in different laboratories as their data are included in silico modeling (Kumar, Dasika, & Maranas, 2007).

Since many of these constraints are quite useful, however, they require mathematical definitions to be implemented. Moreover, the topobiological constraints are extremely difficult to be incorporated in the model for FBA. Environmental constraints, such as pH value, nutrient concentrations, and temperature, are tending to affect the chemical reaction rate and flux balance model. These constraints are implemented in FBA in terms of vectors and matrixes (Gombart & Nielsen, 2000).

### 21.7.5 Optimization

Optimization is a method used in FBA and is represented in terms of mathematical forms in the metabolic network (Burgard & Maranas, 2003). They are used to determine the optimal flux distribution in a network based on an objective function. Experimental results and the response of cells according to the environment are determined by the flux distribution (Burgard & Maranas, 2003).

#### 21.7.5.1 *Linear programming*

Linear programming is used to apply constraints to solve algebraic equations. The main aim of this is how to make optimization of a certain function that is constrained by a number of inequality functions (Lee et al., 2000). The simplex method is commonly used in linear programming. It was the first method developed to solve linear programs. It converts the linear programming into linear equations, which are the first approach. Constraints and mass balance equations are used in the case of FBA (Lee et al., 2000).

#### 21.7.5.2 *Multiinteger linear programming*

An objective function of some algebraic function is used when more than one optimal value is to be determined. These are called alternate optima. Mixed integer linear programming is an algorithm widely used to find all other alternate optima based on recursive linear programming (Lee et al., 2000).

#### 21.7.5.3 *Alternative objective functions*

The growth rate of an organism is often taken as the objective function in FBA as it has shown as the optimizing condition for the experimental data of both wild type and knockout mutants. There are numerous tools developed to optimize the growth of an organism. Minimization of metabolic adjustment (MOMA) is a tool used to minimize the variation in the distribution of the flux through the metabolic network. It uses quadratic programming to optimize the flux distribution relative to wild type and generates accurate predictions for knockout strains (Zamboni, Fendt, Rühl, & Sauer, 2009). There are many alternative objective functions developed to optimize the growth, such as maximization of ATP synthesis for mitochondrial metabolism modeling. Minimization of either redox potential or ATP production leads to maximum cellular energy efficiency (Zamboni et al., 2009). Production of the desired metabolite can be increased by altering the objective function. The accuracy of the model predictions can be increased by the inclusion of metabolic flux data. These objective functions often follow stoichiometric constraints to measure fluxes, such as the release of different products and uptake of substrates. Radioactive isotopes, such as  $^{13}\text{C}$  are used to measure difficult intracellular flux distribution (Zamboni et al., 2009). The cells are allowed to grow on such radioactive isotope-labeled substrate until a steady growth rate is attained. During the course of time, the labeled C atom is integrated well with the amino acid synthesis pathways and is disturbed throughout the proteins.

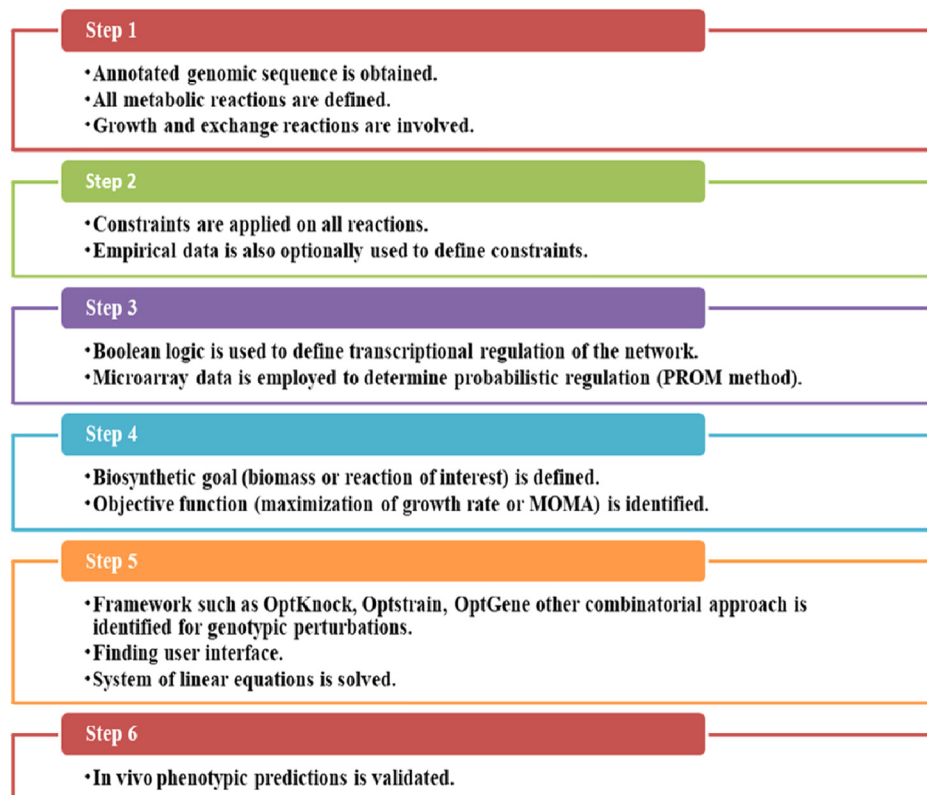
The distribution of radioactive C atom throughout the proteins is then analyzed through nuclear magnetic radiation and MS techniques and the results of flux distribution of the model are generated in the form of curves (Zamboni et al., 2009).

Fine refinement of the metabolic model can be done from the data of flux distribution obtained by labeled carbon atom experiment. Like MOMA, another objective function of regulatory on/off minimization (ROOM) is used to minimize the total number of fluxes through wild type (Zamboni et al., 2009).

### 21.7.6 Steps in FBA

FBA is one of the most important methods in predicting metabolic phenotypes from metabolic mode based on the genomic scale. It works on the principle of the law of conservation of mass on reaction fluxes and metabolites (Orth et al., 2010). An objective function is required by this method to solve the optimal solution. Perturbations, such as overexpression, environmental conditions, gene knockouts, and heterologous gene integration, can be used to predict optimal conditions by inducing phenotypic change and flux distributions (Orth et al., 2010). The six fundamental steps of FBA are explained in Fig. 21.3.

1. Construction of the metabolic model by defining all necessary stoichiometric reactions occurring within the organism with the aid of literature data and genomic sequence.
2. Constraints are applied to each metabolic reaction to obtain meaningful solutions to the set of linear equations.
3. Probabilistic or Boolean logic is used to incorporate the optional transcriptional regulation to the model.
4. Defining mathematical objective function depending upon the objective of the biotechnological goal, organism, and the hybrid between the two.
5. The objective function is maximized or minimized using an appropriate solution algorithm and suitable software to solve the set of linear equations.
6. Validation of predicted phenotypic results with that of in vivo (Orth et al., 2010).



**FIGURE 21.3** Important steps involved in flux balance analysis for the reconstruction of the metabolic model.

## 21.8 Metabolic control analysis and isotopic steady state/carbon flux analysis

### 21.8.1 Metabolic control analysis

Metabolic control analysis is a type of analysis based on the regulatory structures present within the cells that contribute to some fluxes in a metabolic network. Enzymes play a major role in this type of metabolic analysis (Antonie $\acute{w}$ icz, 2015). Besides describing the internal fluxes present in the model, this approach explains the reasons for the occurrences of these fluxes.

In metabolic control analysis, single enzyme-catalyzed reaction steps are used to form a metabolic network. Formation of new products and modifications of substrates in the cells are carried out by these reaction steps. Depending upon the kinetic properties of the enzymes, modeling of these reaction steps is done through mathematical functions. Lot of information is required for complex reaction steps as they may involve inhibitors and other cofactors. Hence, it is difficult to determine the kinetic formulas (Antonie $\acute{w}$ icz, 2015).

#### 21.8.1.1 Data

Data are usually measured in different strains and different organisms under different conditions. There is a difference in the magnitude of estimated kinetic parameters, leading to difficulty in modeling the process. Vector  $e$  is used to represent active enzyme concentrations and it also corresponds to the rate of formation of products by the enzymatic activity (Antonie $\acute{w}$ icz, 2015). Since it depends upon the expression level of the enzymes, it is difficult to measure. Experiments dealing with the measurement of the concentrations of the enzymes from the cell extracts can also determine the vector  $e$ . However, these measurements by cell disruption may influence the results. The following equation shows the formation of a model from the information consisting of the enzyme concentrations, substrate concentrations, kinetic parameters, and stoichiometric information. Under steady-state stationary conditions, the concentrations of the metabolites in the system are constant (Antonie $\acute{w}$ icz, 2015) and represented as given below.

$$Nv(\alpha, s, e; x) = 0$$

where  $N$  represents the stoichiometric matrix,  $\alpha$  represents the vector of all enzyme kinetic parameters,  $s$  represents the vector of extracellular substrate concentrations,  $e$  represents the vector of enzyme concentrations, and  $x$  represents the vector of intracellular metabolite concentrations (Antonie $\acute{w}$ icz, 2015).

### 21.8.2 Carbon flux analysis

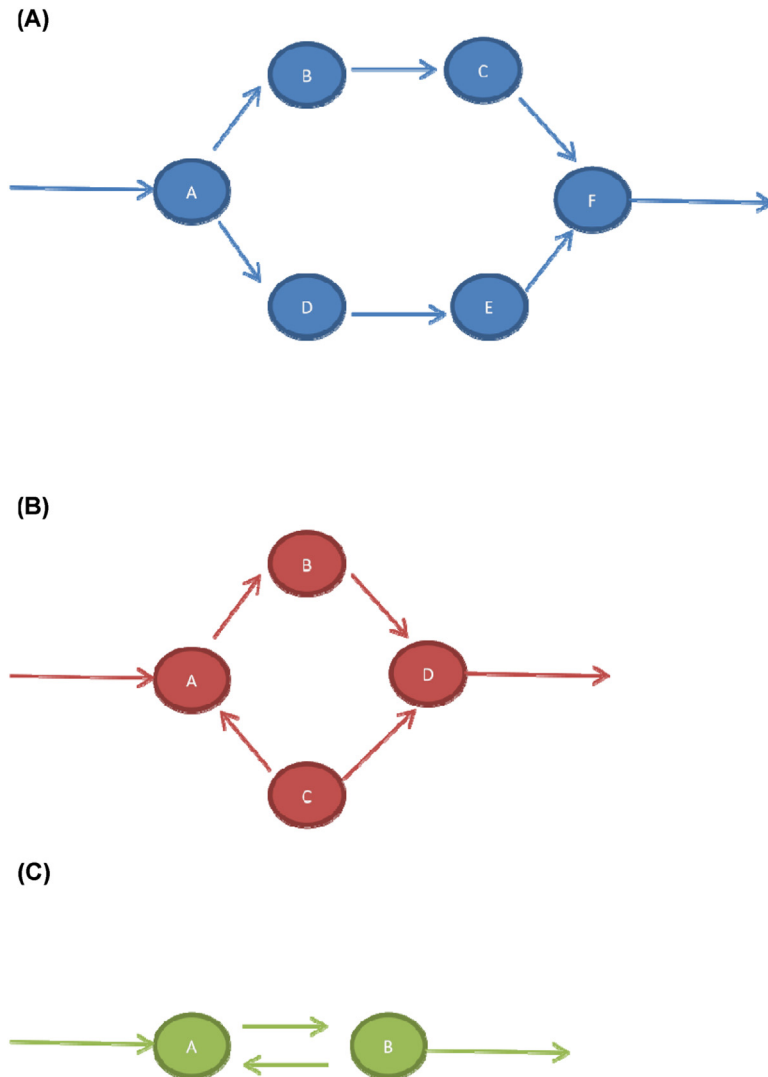
Carbon flux analysis is a type of FBA that is based on the combination of mathematical and biological models to analyze the metabolic networks. Because of some of the drawbacks of stoichiometric based approaches, this method has been widely used. Some of the drawbacks of stoichiometric based methods are depicted in Fig. 21.4A given below, such as they are not functional in bidirectional steps and cannot resolve cyclic fluxes (Wiechert, 2001). When fluxes flowing through branch point A are unknown, then their individual routes cannot be measured and determined. Another case of the circular metabolic pathway is depicted in Fig. 21.4B where fluxes between the branch points A and B and A–C if unknown, then it is difficult to determine the individual fluxes with only stoichiometric information (Wiechert, 2001). In Fig. 21.4C, bidirectional reaction steps are shown. In such cases, it is difficult to determine the two fluxes by stoichiometric-based approaches and only net flux can be determined by the sum of two fluxes (Wiechert, 2001).

Hence, additional information is required to overcome this drawback and to determine the extracellular flux data.  $^{13}\text{C}$ -labeled substrate is widely used in a metabolic stationary state. It is fed into the cells by certain modifications to distinguish it from other carbon molecules present in the metabolic network (Wiechert, 2001). It is generally assumed in carrying of the  $^{13}\text{C}$  experiment that the enzymes within the sense cannot differentiate between the modified and normal carbon molecules.  $^{13}\text{C}$  is distributed throughout the metabolic network (Zamboni et al., 2009). With the aid of NMR and MS, the intracellular distribution of labeled carbon can be measured and the intracellular fluxes can be calculated from the result obtained in a large amount of additional information from these instruments (Zamboni et al., 2009).

#### 21.8.2.1 Algorithm

The working of the algorithm is shown in Fig. 21.5, which is used to evaluate a carbon labeling experiment (CLE) as follows (Zamboni et al., 2009):

1. A flux distribution that fulfills the stoichiometric balance equations is selected and constraint is applied where the sum of incoming fluxes must be equal to the sum of outgoing fluxes.



**FIGURE 21.4** (A) Branch point network, (B) circular network, and (C) bidirectional network.

2. Based on this flux distribution and the input substrate with isotope labeling, a CLE is stimulated.
3. The outgoing isotopomer distribution is computed from the data of measured values if the system shows the guessed fluxes.
4. The measurement difference between the predicted data and the actual data is computed and the discrepancy measured due to the differences is measured by the standard deviation of the sum of squares of every single residual value.
5. Applying the optimization algorithm, systematic variation of the fluxes is calculated from the computed discrepancy. Again, steps 2–4 are repeated.

The CLE is evaluated by the distribution of labeled material throughout the metabolic network using a complicated mathematical model. In this, the intracellular fluxes are assumed to be already known. The distribution of  $^{13}\text{C}$  is calculated based on the known input substrate concentration and the obtained flux values (Zamboni et al., 2009).

## 21.8.3 Different types of flux balance analysis at different conditions

### 21.8.3.1 Analysis of flux under metabolic steady-state

It is widely used to calculate metabolic fluxes in living cells at steady state conditions. In a metabolic network, MFA tends to balance fluxes of intracellular metabolites. This approach is used in full and overdetermined systems. Fully

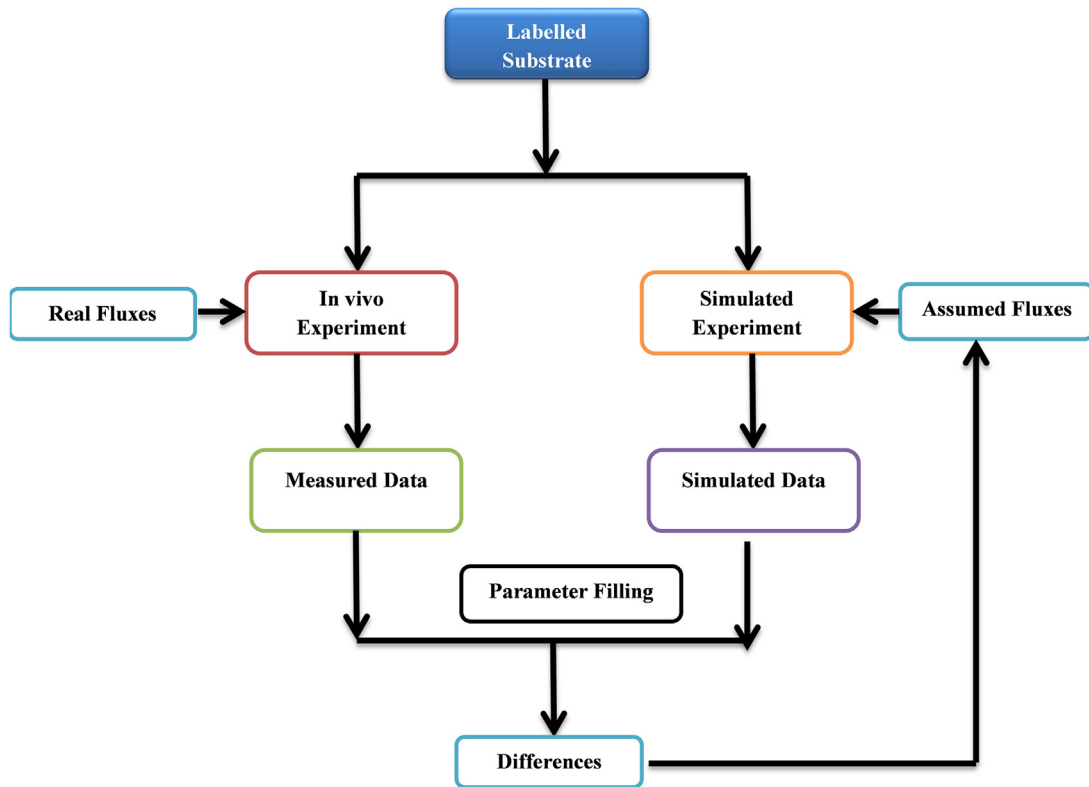


FIGURE 21.5 Flowchart showing the working of the algorithm of  $^{13}\text{C}$  labeling experiment of metabolic flux analysis.

determined consists of all the essential external rate measurements and overdetermined consists of a redundant set of external rate measurements (Zamboni et al., 2009). It is easy to apply and is widely used to measure robust extracellular metabolites. However, it is limited by the number of constraints and it cannot determine the intracellular metabolic pathways (Zamboni et al., 2009).

### 21.8.3.2 Analysis of flux under metabolic and isotopic steady state

It is a more advanced method of determining metabolic fluxes in systems at a metabolic steady state. It uses  $^{13}\text{C}$  labeled tracers in combination with isotopic balancing, isotopomer balancing, and metabolite balancing through techniques, such as tandem MS, NMR, and MS (Zamboni et al., 2009). In this, the cells are allowed to grow for a certain duration of time in specifically labeled  $^{13}\text{C}$  tracer, for example, glucose with the incorporation of  $^{13}\text{C}$  atoms and then the intracellular metabolic fluxes can be determined by their distribution into metabolic intermediates and products. It is assumed that the metabolic steady state and isotopic steady state are constant over time. Isotopic steady state is dependent on several factors, such as tracer substrate, metabolic activities of the cells, and the composition of the medium (Zamboni et al., 2009). The main advantage of this method is that a large number of redundant measurements can be determined from flux distribution thereby increasing the accuracy and precision of estimated fluxes (Zamboni et al., 2009). It is also used in complex metabolic network models, such as bidirectional reversible fluxes, cyclic pathways (e.g., pyruvate cycle), and parallel metabolic pathways (e.g., glycolysis and pentose phosphate pathway). Tools based on  $^{13}\text{C}$  MFA are OpenFlux,  $^{13}\text{CFLUX2}$ , Metran, and INCA (Zamboni et al., 2009).

### 21.8.3.3 Analysis of flux under isotopic nonsteady state

In this method, the metabolic fluxes are determined at pseudo steady state, that is, transient  $^{13}\text{C}$  labeling data, and the metabolites pool size is used for the quantification of fluxes along with partial constant metabolic fluxes and pool size conditions. It overcomes the limitations of  $^{13}\text{C}$  MFA (Wiechert, 2001). In this method, ODEs are used with isotopomer balances to estimate the isotopomer distributions as a function of time. This method has a higher computational time than  $^{13}\text{C}$  MFA. Nonlinear least squares regression techniques are used to estimate labeling transients (Wiechert, 2001).

#### 21.8.3.4 Analysis of flux under metabolic nonsteady state

This method relies on temporal transients in metabolic fluxes over the change in time. It is used to determine the metabolic shifts at metabolic nonsteady state (Shlomi et al., 2005). It is assumed that the transients of fluxes are relatively slow, that is, order of hours while the intracellular metabolism takes an order of seconds to minutes to reach metabolic pseudo steady state. Following steps are involved in dynamic metabolic flux analysis: (1) dividing experiments into discrete time intervals, (2) calculation of average external rates by derivatives of external concentrations for each time interval, and (3) applying classical MFA to calculate average fluxes for each time interval (Shlomi et al., 2005). Flux transients are framed into a time profile after obtaining the results from the steady state models at multiple time points. Methods, such as linear and polynomial fitting, derivations, and splines, are applied on extracellular measurements for data smoothing (Shlomi et al., 2005). Since it does not rely on the manual selection of time intervals it serves as an advantage of this method. Detecting the level of dynamic information is done automatically by the rigorous statistical criteria. It helps in obtaining information on metabolic transient with a less experimental and computational effort, which could not be possible in the case of MFA. However, it has the same limitations with resolving reversible reactions, cyclic, and parallel pathways as that with MFA since it is based on stoichiometric metabolite balancing in the metabolic model (Shlomi et al., 2005).

#### 21.8.3.5 Analysis of flux under metabolic and isotopic nonsteady state

This method is used to incorporate isotope labeling measurements to estimate the fluxes of reversible, parallel, and cyclic pathways (Varma & Palsson, 1994). However, this method is still in the infant stage and further research and development are required to make this method more efficient and accessible, and less cumbersome to implement (Varma & Palsson, 1994).

#### 21.8.3.6 Recent advances in metabolic flux analysis

Highly precise metabolic fluxes can be estimated by the use of parallel labeling techniques together with the combination of rigorous data integration. Instead of a single  $^{13}\text{C}$ -MFA, multiple tracer experiments can be performed in parallel (Antoniewicz, 2015). A single flux map of all different experiments is used to label data by combining and integrating rigorously. COMPLETE-MFA is used to produce more precise information than  $^{13}\text{C}$ -MFA. It is considered as the golden technique in  $^{13}\text{C}$ -based flux estimation (Antoniewicz, 2015).

## 21.9 Some important tools of flux balance analysis

There are numerous tools available for FBA. Some of the important tools are as follows:

### 21.9.1 OptKnock

A typical metabolic model consists of a large number of genes ( $\sim 1000$ ). Therefore it becomes computationally intensive to carry out an exhaustive search of knockout sets as the size of the set increases (Bugard, Pharkya, & Maranas, 2003). OptKnock works on duality theory, which states that for every primal LP problem, there is a unique dual LP problem that equals the objective function of the primal. The dual problem is used to maximize biomass production and add constraint by defining the objective functions equal to each other (Bugard et al., 2003). A single mixed integer linear programming (MILP) problem is used to couple the maximum product yield with maximizing growth. Optx and OptKnock algorithms are wide as a standalone program to carry out knockout prediction (Bugard et al., 2003).

### 21.9.2 OptGene

Nonoptimization of nonlinear objective functions is a drawback of OptKnock. With a large number of knockouts, it becomes a computationally intensive task with MILP problem (Bugard et al., 2003). OptGene overcomes these drawbacks through a genetic algorithm as follows:

- A set of random optimal conditions is generated.
- MOMA and FBA are used to solve the metabolic model for each set.
- A score is assigned to each member of the set depending upon some function of metabolic state.
- The highest score of the optimal condition is determined.
- Step 2 is repeated till the best score is attained (Bugard et al., 2003).

### 21.9.3 OptStrain

Although OptGene and OptKnock tools are considered very powerful in predicting knockouts, their working space is limited to the alterations in the metabolic reactions. OptStrain overcomes this problem by developing a database of bio-transformation to improve the prediction of heterologous pathways (Bugard et al., 2003). Following is the approach for the working of OptStrain:

- LP is applied to determine the maximum possible product formation. It forms the baseline of the yield.
- A minimum number of heterologous genes is identified by MILP to match the baseline yield. It is necessary to make an assumption of maximizing the product yield, not growth.
- The identified genes in step 2 are incorporated into the stoichiometric model. Optimization is done by the OptKnock method.
- Product formation is improved by a set of knockouts and insertions (Bugard et al., 2003).

### 21.9.4 COBRA Toolbox

The COBRA Toolbox is a software package of MATLAB (Becker et al., 2007). It is widely used to perform the MOMA analysis and maximization of growth. A large amount of flexibility and simplicity to use is an advantage of COBRA and it uses various scope functions to optimize the model, such as objective functions and solution algorithms (Schellenberger et al., 2011).

### 21.9.5 MetaboAnalyst 4.0

It is a tool developed for metabolic data analysis, functional interpretation, and visualization. It produces reproducible and transparent analysis using the R package. Functional enrichment analysis is used to regulate metabolic pathways (Chong et al., 2018). Untargeted metabolomics data are determined by the mummichog algorithm. It supports biomarker meta-analysis and multiomics data integration. It consists of 12 different modules divided into four categories: (1) data integration and systems biology, (2) exploratory statistical analysis, (3) data processing and utility functions and (4) functional analysis (Chong et al., 2018). Data gathered from targeted or untargeted metabolites are accepted by the exploratory statistical analysis. MS data of pathway activity prediction are contained in the functional analysis category (Chong et al., 2018). Biomarker meta-analysis, network explorer, and joint pathway analysis are three modules included in the data integration and systems biology category. Data processing tools that include web based tools, compound ID conversion, etc., are included in the data processing and other utilities category. MetaboAnalyst is fast, robust, and intuitive and it produces reproducible data (Chong et al., 2018).

### 21.9.6 OptFlux

OptFlux operates in a similar way to the COBRA Toolbox. It performs both MOMA optimization and growth maximization. It works on the JavaScript platform with a simple user friendly interface instead of MATLAB. OptKnock algorithm is used by OptFlux along with the Boolean logic. However, it is not flexible for simple one time changes (Aurich, Fleming, & Thiele, 2016). It provides open source software packages to users. It is an open source platform where users can freely work and it helps in the modeling of background informatics easily. It is modular and compatible with SBML and other file formats. It works well with FBA, MOMA, and ROOM (Aurich et al., 2016). It works on an AIBench application that incorporates three well-defined objects into the modeling, such as datatype views, data types, and operations. It helps users to perform MFA under various environmental and genetic conditions (Aurich et al., 2016). It provides several operations, such as import, export, and visualization of stoichiometric metabolic models, such as equations, metabolic reactions, and gene reaction associations. It is compatible with databases, such as BioModels and BiGG database, and software tools, such as CellDesigner. An explicit definition is used to incorporate external metabolites and identify biomass formation reactions (Aurich et al., 2016). OptFlux uses three methods to conduct simulations, such as MOMA, FBA, and ROOM. FBA method uses LP formulation for the calculation of fluxes of wild type or mutant strains. MOMA uses quadratic programming while ROOM uses MILP and LP. Optimization is done by OptKnock and two meta-heuristic algorithms—EA and SA (Aurich et al., 2016). In the first method, the user has to specify the allowed number of genes and reactions while in the second method, the optimum value is determined automatically. It currently allows the optimization of two objective functions—biomass product coupled yield and total yield. Total yield is targeted on yield by the desired compound while biomass product coupled yield is focused on

higher productivity (Aurich et al., 2016). EFM Tool is used for the state of the EFM calculation. BioVisualizer is a plugin used by OptFlux for the graphical visualization of graphs, nodes, edges, and connections. It provides users with good usability, intuitive, and simple user friendly tools (Aurich et al., 2016).

### 21.9.7 OpenFlux

It is an intuitive and versatile user-friendly spreadsheet-based interface developed to regulate the isotopomer and metabolite-based models. It is used in FBA, flux estimation, sensitivity analysis, and implementation of large-scale metabolic models and networks (Quek, Wittmann, Nielsen, & Krömer, 2009). Elementary metabolite units (EMU) decomposition algorithm is used in OpenFlux to generate an isotopomer balance model. It is computationally more efficient. It is validated against the  $^{13}\text{C}$  MFA results as it is an efficient and generic tool for  $^{13}\text{C}$  MFA. It is intuitive, simple, and faster than  $^{13}\text{C}$  Flux. Unknown free fluxes in large-scale metabolic models can be easily determined using statistical evaluation (Quek et al., 2009). It uses Monte-Carlo and nonlinear simulation methods to evaluate flux solutions, thereby making it a robust method. It is useful in exploring various aspects of metabolomics, such as flux distribution and network topology. It provides users with a simple text interface to regulate the metabolic models (Quek et al., 2009). It supports the use of various labeled substrates, such as ATP and NADPH. The text-based spreadsheet interface is regarded as one of the most efficient means of disintegrating the metabolic model into modeling assumptions and network topology in a single model definition file (Quek et al., 2009).

### 21.9.8 CellNetAnalyzer

It uses MATLAB like COBRA Toolbox. It operates in a simple graphical user interface unlike the MATLAB command window (Cheng, 2012). It uses significant Boolean logic to the model making an easy way of using various visualization and interactive tools. It does not use MOMA or any other advanced solving techniques other than the maximization of growth (Cheng, 2012).

### 21.9.9 SBRT

The Systems Biology Research Tool is a software package used in FBA that runs in JavaScript. It is an open source and plug-in-ready software package (Wright & Wagner, 2008).

### 21.9.10 Escher-FBA

Escher-FBA is a user friendly and versatile visualization tool developed by Escher to study the metabolic pathways. It is a quick and easy way to mapping GEM containing reactions in the model and visualizing both metabolites and reactions. Numerous flexibilities are provided by Escher FBA to users, such as objective function, flux bounds, and reaction knockouts, to modify parameters and visualization of the results. It works on various platforms, including mobile devices (Rocha, Maia, & Evangelista, 2010). Hence, it is widely used in academic labs to visualize, explore, and learn FBA simulations. It provides users to load, modify, and save their maps that are stored as JSON files. It provides interactive tooltips to modify the parameters of FBA simulation. GNU Linear Programming Kit is used in FBA simulations. The tooltip is used to adjust the lower and upper bounds of the reaction (Rocha et al., 2010). These lower and upper bounds are used to obtain precise values. Knockout of the associated reaction is simulated by a knockout button by setting lower and upper bounds to zero. The reset button is used to reset the bounds to their original values. Maximization and minimization of the objective function are done by maximize and minimize buttons. These controls are used to control the outcomes of the visible effects within the system. Custom maps and models are supported by Escher FBA (Rocha et al., 2010). It provides users to create their own maps and models and in silico experiments can be performed with their own data. BiGG Model is a database used to download additional models and maps. GEMs are supported by Escher FBA in COBRA JSON format. COBRApy is used to convert files in other formats to JSON. It supports files with SBML and FBC extension (Rocha et al., 2010).

### 21.9.11 MetaFluxNet

MetaFluxNet is a tool to construct a metabolic network and perform MFA. It is an easy, user friendly, and customized way of performing dynamic visualization. It is written in C# with .NET platform and supports packages of Java libraries

and VGJ for graph layout, JAMA for various matrix operations, and LP for linear programming. It helps to work under a well-established construction environment where they can create their own metabolic network (Lee et al., 2003). It contains information, such as EC number, concentrations of substrates and products, and gene name participating in the reaction. It allows users to edit, store, and modify these fields (Lee et al., 2003). It is used to construct metabolic models by defining stoichiometric equations at steady state, application of redundancy and determinacy methods, and optimizing flux distribution by LP approach (Lee et al., 2003). It allows users to obtain the effect of gene addition or deletions in the metabolic network using optimal metabolic flux distribution. Thus the physiological and metabolic changes in the cell can be studied under different conditions. MetaFluxNet is also used to design new and advanced metabolic engineering strategies (Lee et al., 2003). Various labels are used to represent the results more attractively. The comparative MFA results are displayed in one window. Labels, such as “added or deleted” and “measured,” are used to specify the gene modification. “Calculated” and “bound” labels are used to specify nonmeasured metabolic fluxes. Designing and evaluating the metabolically engineered *in silico* strains can be done by the MetaFluxNet using the features of bound fluxes and changing the genotypes (Lee et al., 2003). It displays pathways in a dynamic and interactive graphical user interface. Spring embedder layout algorithm is an algorithm used to visualize pathways in an automated and dynamic manner. The MetaFluxNet tool is under intensive upgradation to be integrated with the dynamic simulation of comprehensive metabolic networks, simulation and modeling, structural pathway analysis and database management system (Lee et al., 2003).

## 21.10 Applications, challenges, and future perspectives of FBA

### 21.10.1 Applications

There are numerous applications of FBA; one of the most important is a simulation of systems under varying experimental conditions. They are widely used in the identification of drug targets, annotation of genomes, and optimization of bioprocess in industries (Lakshmanan, Koh, Bevan, Chung, & Lee, 2012). Fig. 21.6 shows various applications of FBA.

#### 21.10.1.1 Identifying “essential” and “synthetic lethal” genes

One of the important applications of FBA is the determination of “essential” gene and “synthetic lethal” gene pairs. Cell growth is mediated by the essential genes, and to determine essentiality, FBA is performed on a new reaction network after each gene is knocked out (Kuang et al., 2020). A gene that had been knocked out and designated as an essential gene is determined when the growth rate is low especially between 0.1 and  $1e^{-9}$ . On the other hand, synthetic lethal genes are those consisting of two genes: when both are knocked out, they stop the growth and when they are knocked out individually, they do not stop the growth (Kuang et al., 2020). They are determined in a way similar to essential genes. To determine the entire set of essential genes, single gene deletion is performed first. After that, each pair is sequentially removed; FBA is performed using the same criteria as that of essential genes starting from the set of nonessential genes. Similarly, higher-order genes and synthetic triplets can be determined by this method; however, this process becomes computationally intensive and prohibitive for full metabolic networks (Kuang et al., 2020).

#### 21.10.1.2 Maximizing production of biomolecules

Another major application of FBA in metabolic engineering is to produce the desired product by manipulating the genome of an organism. FBA provides deep insights into various metabolic engineering tools of these aspects, such as up- and downregulation of certain genes, reaction knockouts, the inclusion of nonnative reaction pathways, and alternative path analysis (Lakshmanan et al., 2012). Among these, gene knockouts have been proved as an effective technique to maximize the production of the desired product. This method tends to eliminate other competing pathways having the same precursor metabolites as target compounds. Since the steady state system curbs chemical accumulation, the driving force through the desired reaction is increased by increasing those metabolite precursors (Lakshmanan et al., 2012). It is advisable to have prior knowledge of synthetic lethal and essential genes while designing knockout strategies to avoid unfruitful attempts.

OptKnock is an algorithm widely used to frame knockout techniques for maximizing the production of the desired metabolite. It works on the application of a bi-level optimization program that operates in maximizing both the production of metabolites and growth rate simultaneously (Lakshmanan et al., 2012). It fixes the production of the desired

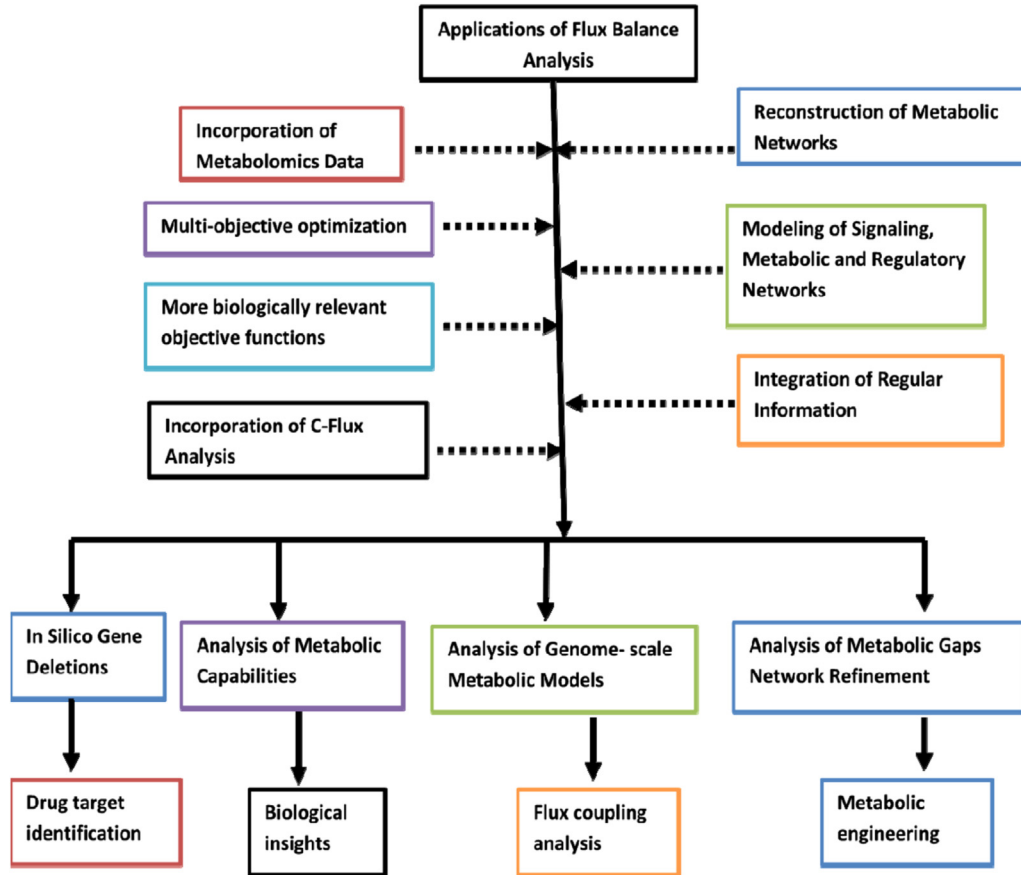


FIGURE 21.6 Applications of flux balance analysis.

metabolite to produce biomass precursors if genes for deletion are identified. Production of the metabolite is directly proportional to the production of the biomass precursor due to stoichiometric constraints. Another algorithm OptForce is widely used to identify strategies for maximizing the production of a product of interest by up- and down-regulating the genes with addition to knockouts. Lists of similar optimization algorithms are summarized in the table given below that are used for performing metabolic manipulations (Lakshmanan et al., 2012).

### 21.10.1.3 Analysis of genome-scale metabolic networks

FBA is widely used to construct and analyze genome-scale metabolic networks (GSMNs). They are also applicable in various organisms, such as archaea, bacteria, and eukaryotes. Metabolic networks of humans have also been constructed by GSMNs (Raman & Chandra, 2009). They are used in the identification of potential drug targets, improving the strains of microbes and deep understanding of microbial metabolism, detection of gene deletion, identification of pathogenic organisms, and metabolic engineering. GSMNs can be integrated with high-throughput transcriptomics to obtain data on phenotypes and metabolic capabilities. These techniques are called rFBA (Raman & Chandra, 2009).

### 21.10.1.4 Flux coupling analyses

With the aid of reconstructed networks, structure and topology can be easily determined. A couple of fluxes such as in the case of extreme pathway analysis and elementary flux mode analysis can be possible. A tool called flux coupling finder is used to analyze different types of coupling (Covert et al., 2001). It globally identifies the blocked reactions in the network, which are not capable of carrying out flux reactions under certain conditions. The knockout reactions that force reactions to zero due to deletion can also be determined by flux coupling finder (Papin et al., 2004).

### 21.10.1.5 Refinement of metabolic networks

FBA is also widely applicable in the refinement of the metabolic networks. It helps in improving the knowledge of understanding the metabolism of an organism (Kuang et al., 2020). During the simulations of metabolic networks, FBA is used to refine “metabolic gaps” occurring due to inconsistencies in experimental data. An optimized based algorithm could be employed to identify the missing gaps and reconcile the inconsistencies between experiment and reconstructed metabolic networks (Raman & Chandra, 2009). The gaps can also be resolved with the aid of a systematic network (Papin et al., 2004).

### 21.10.1.6 Predicting novel regulatory mechanisms

Predicting novel regulatory mechanisms is possible by estimating the discrepancies between experimentally observed phenotypes and predicted growth phenotypes. These discrepancies are raised due to missing regulatory effects in the model (Raman & Chandra, 2009).

## 21.10.2 Challenges and future perspectives

Although there are numerous applications of FBA, still there are a number of areas where FBA continues to be refined. Metabolic networks are being constructed for a large number of organisms with the availability of high-throughput data analysis from FBA to analyze the genomic information (Raman & Chandra, 2009). There are many errors to be found in these reconstructions of the model due to inaccurate annotations and insufficient data. With the aim of developing accurate, distributable, and applicable metabolic models, it is crucial to keep in mind to carry out meticulous experimental studies and judicious record keeping (Lakshmanan et al., 2012). A static picture of the metabolic network is used to analyze these models to obtain insight into the functioning of the organism. Dynamic functions of certain organisms have been made by bringing changes in the traditional FBA; however, their scope is limited. Organisms, for example, photosynthetic autotrophs do not function at a constant steady state. Since they undergo transition through dark and light reaction cycles, the operational modes of these organisms are highly alternated (Radhakrishnan et al., 2002). Hence, there is a need to develop better strategies to grasp and boost the predictive schemes to understand such kinds of dynamics. More representative objective functions need to be created to understand the cyclical nature of some organisms (Raman & Chandra, 2009). The application of FBA in the maximization of biomass has been proven to show promising results for some organisms to the evolutionary adaptation to the environmental conditions. Intensive research has been carried out to identify the best objective functions to apply under specific circumstances by the inclusion of experimental data (Lakshmanan et al., 2012).

### 21.10.2.1 Limitations of FBA

Since constraints are applied in FBA, the solution obtained is also good to build the model. However, it requires a lot of computational time and effort to construct a metabolic network (Kuang et al., 2020). It sometimes contains incomplete annotation of proteins or genes in a genome. It focuses mainly on the enzymes that help in catalyzing reactions in the cell. Moreover, zero fluxes could be obtained in certain cases due to the incomplete annotation aroused due to non-characterization of upstream and downstream regulators and several other metabolites or metabolic gaps (Kuang et al., 2020).

### 21.10.2.2 Biologically relevant objective functions

Sometimes more biologically relevant objective functions are required in the genome scale reconstructions of higher organisms. Biological Objective Solution Search is a method developed to resolve the problem of selecting a suitable objective function similar to ROOM and MoMA (Lakshmanan et al., 2012). In this method, a new stoichiometric reaction of the biological objective reaction is added to the stoichiometric matrix. Thus the existing constraints are optimized by adding these reactions. This minimizes the variation between the available experimental data and the existing flux distribution (Lakshmanan et al., 2012). It is also used to compare the experimental flux data with various objective functions thus enabling synergetic relationships between the biochemical experiments and in silico simulations that aim to improve the framework of metabolic models (Kuang et al., 2020). The objective function needs to be selected with respect to the conditions of the simulation. The choice of selecting the objective function is considered to be an important measure in FBA and it is essential to select a biologically suitable objective function. This method requires two important improvements: first, an appropriate objective function needs to be identified and second, there is

a requirement of detailed large-scale quantitative experimentation to generate function with high resolution under various conditions (Kuang et al., 2020).

### 21.10.2.3 Impact of high-throughput experiments

Large amounts of data can be generated from high through techniques for estimating metabolomics data. This is helpful in identifying more constraints. It can overcome the drawbacks of incompleteness and uncertainty in metabolomics data. Metabolomics experiments can also be used to identify metabolic gaps in FBA (Kauffman & Edwards, 2003). The recent advancements in high-throughput carbon flux analysis have revolutionized the way of estimating the intracellular fluxes, improving the choice of objective functions and constraints in FBA models, and generating more data for hypothesis validation. Genome scale transcriptomics data can also be used in the integrated reconstruction of regulatory and metabolic networks (Kauffman & Edwards, 2003).

## 21.11 Case study: applications of metabolomics and flux balance analysis in industrially important microorganisms

### 21.11.1 *Lactococcus lactis*

*Lactococcus lactis* is a strain of lactococci bacteria widely used in fermentation industries. It is important because of its ability to undergo process development. *L. lactis* is used for the production of diverse recombinant proteins using a genetic engineering approach. It uses an extracellular secretion system (Nicolas, Dirk, & Peter, 2013). It is used as an important raw material in the pharmaceutical industry and moisturizing agent in the cosmetic industry. Metabolic models of *L. lactis* have been developed to understand the flavor forming pathways and nitrogen metabolism for industrial purposes. A genome scale reconstruction of the constraint based model was developed for *L. lactis* comprising of 650 metabolites, 518 genes, and 754 reactions. It has been reported that 59 reactions out of 754 are involved either directly or indirectly in the flavor flux formation (Nicolas et al., 2013). Flux distributions across these reactions were determined by flux variability analysis and FBA. The energy parameters were determined by anaerobic carbon limited cultures. Various models have been developed to optimize the nitrogen metabolism, such as coupling of redox reactions with branched chain amino acid catabolism, which plays an important role in the formation of various flavor compounds (Oliveira, Nielsen, & Förster, 2005). *L. lactis* MG1363 was developed to model nitrogen and carbon metabolism. It was developed with aim of developing potential applications in dairy and cheese manufacturing industries by modifying the metabolic networks of flavor formation. The entire metabolic model along with transport networks was developed using constraint based genome scale metabolic models (Oliveira et al., 2005). These models were developed to reconstruct whole transport and metabolic networks through the linkage of transport and enzymatic reactions occurring in various anabolic and catabolic pathways. These were used to model the flavor forming pathways from  $\alpha$ -ketoglutarate precursors. Various models of *L. lactis*, such as NCDO712, MG1363, and IL1403, have been developed that favors flavor compounds (Oliveira et al., 2005). These models were comprised of information related to enzymatic and chemical reactions using the AUTOGRAPH algorithm. Experiments consisting of continuous culture were designed using constraint-based modeling techniques. This involved curations of manually created metabolic networks for flavor forming reactions (Liu, Chan, Chen, Solem, & Jensen, 2019). This approach has proven to improve agro industrial applications of *L. lactis*. All of these models accurately measure ammonia production and amino acid consumption rates. Therefore the genome scale metabolic model of *L. lactis* proved to be an important tool for the thorough study of nitrogen and carbon metabolism, which are essential for fermentation industries especially for the production of flavor compounds (Pham et al., 2019).

### 21.11.2 *Saccharomyces cerevisiae*

*S. cerevisiae* is an important strain of yeast used in the fermentation industry, bioremediation, dairy industry, pharmaceutical industry, and biofuel production. It is widely used for the production of industrially important compounds, such as recombinant proteins, polyphenols, and isoprenoids (Deutscher, Meilijson, Kupiec, & Rupp, 2006). *S. cerevisiae* is suitable for the large-scale fermentation process as it is fast growing and cost effective. It can be easily subjected to posttranslational modifications, such as glycosylation, acylation, and phosphorylation. It is therapeutically more useful for the production of insulin and human serum albumin. It can be easily modified using genetic tools (Deutscher et al., 2006). Genomic scale metabolic models of *S. cerevisiae* have been developed using FVA and FBA and tools, such as OptKnock. These models have been extensively used to increase the production of vanillin, ethanol, and succinate.

Identification of lethal and essential genes can be done using these models (Duarte, Herrgard, & Palsson, 2004).  $^{13}\text{C}$  MFA can be used to determine the intracellular fluxes, oxygen availability, and capability to produce desired products, catabolic repression, and gene knockouts. Improved production of ethanol can be possible by a thorough study of glycine biosynthesis and glyoxylate shunt pathway using this approach (Duarte et al., 2004). Specific pathways of *S. cerevisiae*, such as the  $\alpha$ -aminoadipate pathway, which is an important pathway for lysine biosynthesis, can be subjected to atom mapping and the information obtained from this is then integrated into the databases, such as KEGG, yeast metabolic database, Metacyc, and MetRxn. MFA analysis of *S. cerevisiae* is generally carried out using metabolomics techniques of NMR, LC–MS, and GC–MS (Gopalakrishnan & Maranas, 2015). GC–MS helps to determine 12 essential amino acids, LC–MS helps to determine compounds of glycolysis pathway, such as pyruvate, 3-phosphoglycerate, and phosphoenol pyruvate, and NMR helps to determine all amino acids except glutamine, tryptophan, asparagine, and cysteine.  $^{13}\text{C}$  MFA can also be used to determine fatty acids and nucleotides using algorithms, such as OptMeans. However, certain issues need to be addressed, such as loss of resolution can be raised in pathways, such as the TCA cycle, methylglyoxal, glycolysis, and  $\gamma$ -aminobutyrate. The reliability of  $^{13}\text{C}$  MFA strongly depends on the isotopic steady state (Gopalakrishnan & Maranas, 2015). Genome scale reconstruction of the *S. cerevisiae* model is carried out using the EMU algorithm, which accurately measures fluxes and minimizes any deviation of metabolite labeling patterns from predicted and experimental data. However, certain challenges are associated with the GSM model (Herrgard, Lee, Portnoy, & Palsson, 2006). Since the metabolic model comprised of 2223 metabolites and 3494 reactions, of which 856 reactions are unable to carry out aerobic glycolysis, it becomes a computationally herculean task for determining intracellular fluxes. Since there are no such tools available for estimating flux using the GSM model, it affects the scalability of biomass production. These challenges can be addressed by some useful algorithms, such as pFBA. It significantly reduces the computational time (Herrgard et al., 2006).

### 21.11.3 *Escherichia coli*

*E. coli* is widely considered to be one of the suitable organisms for the optimal production of natural products as it is highly productive and can be easily manipulated using genetic engineering tools (Chandrasekaran & Price, 2010). It is a suitable host for the production of recombinant proteins. It has certain advantages in bioprocess industries, such as high versatility, simple culture requirements, cost effectiveness, and fast growth (Chandrasekaran & Price, 2010). Since *E. coli* is not suitable for posttranslational modifications, synthetic PTMs can be added to obtain desired proteins or recombinant proteins, such as IFN  $\alpha$ -2a and  $\alpha$ -2b, insulin, and growth hormone. *E. coli* is an enteric bacterium that causes urinary tract infections and it is gram negative and nonphotosynthetic. The metabolic pathways occurring in *E. coli* can be accessed from databases, such as EcoCyc, MetaCyc, and KEGG (Covert, Markus, & Palsson, 2002). The whole genome of *E. coli* can be obtained from the techniques, such as microarrays, genome-wide mRNA expression analysis, and genomic sequence analysis of novel mutants.  $^{13}\text{C}$  MFA is applied to *E. coli* for the identification of amino acids and all other essential metabolites. It is also subjected to kinetic flux profiling (Covert et al., 2002). Advanced computational techniques, such as steady-state FBA, can be used for the detection of unknown fluxes. For *E. coli*, the maximum amount of carbon or glucose consumption is used as an objective function (Edwards, Ibarra, & Palsson, 2001). The linear optimization method in optimal determination fluxes has been known to match with assumptions and predictions made. Complex metabolomics data can be obtained from the *E. coli* model using matrix decomposition techniques, such as principal component analysis (PCA) and singular value decomposition (SVD). SVD deals with the breaking of matrix data into corresponding vectors, which represents response patterns while PCA determines the variance between these response conditions (Edwards & Palsson, 2000). With the aid of advanced genomic techniques, such as GeneChips, genome sequencing, and microarrays, the behavior of central metabolic networks can be easily determined. These networks can be quantitatively studied by FBA and cybernetic modeling (Ibarra, Edwards, & Palsson, 2002). Among these, FBA has been shown to generate a relationship between predicted and experimental data in *E. coli*. It has also been used to study diauxic growth using complex mathematical modeling and reprogramming algorithm (Ibarra et al., 2002). Besides these, the cybernetic modeling algorithm has proven to be an optimal method for the modeling of *E. coli*'s diauxic growth on the media containing fumarate, pyruvate, glucose, and succinate (Orth et al., 2011). Dynamic flux balance analysis (DFBA) is a method used to perform dynamic modeling by estimating the changes in flux constraints. It can be used to predict the dynamics underlying the diauxic growth of *E. coli* (Orth et al., 2011). Although classical FBA is used to study diauxic growth on media containing acetate and glucose, it does not predict correctly the reutilization of acetate. Moreover, it does not give us information about the concentration of metabolites in the reaction (Schuetz, Kuepfer, & Sauer, 2007). DFBA overcomes these limitations by introducing kinetic

expressions. It is also used to examine the change in flux constraints, parameters, and diauxic growth of *E. coli*. Hence, it is widely applicable as a quantitative analysis tool in industrial biotechnology (Schuetz et al., 2007).

## 21.12 Conclusion

Metabolomics is a promising field of science that tends to determine the effect of exogenous factors, such as host factors, disease states, and treatments, to bring changes in global metabolic profiles. MS and NMR are the widely used techniques in metabolomics to analyze the various types of metabolites using multivariate statistical analysis. Various metabolic models have been developed to study metabolic fluxes across metabolic networks and reactions. FBA is widely used for metabolic models with stoichiometric equations. The incorporation of thermodynamic constraints makes the predictions more accurate and reliable. Since there is a need for improvement in the estimation of fluxes accurately, numerous algorithms have been developed for flux analysis, model refinement, and development. Constraint-based analysis of FBA is a versatile tool for metabolic engineering and bioprocess technology. It provides solution space for the solution of the network. The stoichiometric matrix provides null space, which is used to create the solution space. This provides information about substrate requirements, consumption, and product formation. MCA is another approach for modeling an enormous amount of data of the system. Isotope labeling is another technique used for the analysis of metabolites in a system. Each of the above discussed modeling approaches has both advantages and disadvantages. As FBA provides a complete range of solutions, it does not provide a precise solution. MCA although being very precise and easily determines each flux, it lacks all the kinetic information. While carbon flux analysis resolves fluxes, however, it is time consuming. <sup>13</sup>C MFA is widely used in metabolic engineering for quantifying fluxes. It has short labeling times and uses efficient algorithms for statistical analysis and parameter fitting. Certain improvements need to be brought to new analytical techniques for better determination of fluxes in complex and compartmentalized metabolic models.

## References

- Antoniewicz, M. R. (2015). Methods and advances in metabolic flux analysis: A mini-review. *Journal of Industrial Microbiology and Biotechnology*, 42(3), 317–325.
- Aurich, M. K., Fleming, R. M., & Thiele, I. (2016). MetaboTools: A comprehensive toolbox for analysis of genome-scale metabolic models. *Frontiers in Physiology*, 7, 327.
- Beard, D. A., Liang, S. D., & Qian, H. (2002). Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83(1), 79–86.
- Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O., & Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. *Nature Protocols*, 2(3), 727–738.
- Bugard, A. P., Pharkya, P., & Maranas, C. D. (2003). Optknoack: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6), 647–657.
- Burgard, A. P., & Maranas, C. D. (2003). Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and Bioengineering*, 82(6), 670–677.
- Chandrasekaran, S., & Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41), 17845–17850.
- Cheng, Q. (2012). Microbial metabolic engineering: Methods and protocols. *Methods in Molecular Biology*, 834.
- Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., ... Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, 46, 486–494.
- Covert, M. W., & Palsson, B. O. (2003). Constraints-based models: Regulation of gene expression reduces the steady-state solution space. *Journal of Theoretical Biology*, 221(3), 309–325.
- Covert, M. W., Markus, W., & Palsson, B. O. (2002). Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *Journal of Biological Chemistry*, 277(31), 28058–28064.
- Covert, M. W., Schilling, C. H., & Palsson, B. O. (2001). Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1), 73–88.
- Deutscher, D., Meilijson, I., Kupiec, M., & Ruppin, E. (2006). Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature Genetics*, 38(9), 993–998.
- Duarte, N. C., Herrgard, M. J., & Palsson, B. O. (2004). Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, 14(7), 1298–1309.
- Edwards, J. S., & Palsson, B. O. (2000). The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), 5528–5533.
- Edwards, J. S., Ibarra, R. U., & Palsson, B. O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19(2), 125–130.

- Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., ... Hermjakob, H. (2017). Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinformatics*, *18*(1), 142.
- Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, *13*(3), 344–349.
- Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L., & Palsson, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, *7*(2), 129–143.
- Gombart, A. K., & Nielsen, J. (2000). Mathematical modelling of metabolism. *Current Opinion in Biotechnology*, *11*(2), 180–186.
- Gopalakrishnan, S., & Maranas, C. D. (2015). Achieving metabolic flux analysis for *S. cerevisiae* at a genome-scale: Challenges, requirements, and considerations. *Metabolites*, *5*, 521–535.
- Herrgard, M. J., Lee, B. S., Portnoy, V., & Palsson, B. O. (2006). Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Research*, *16*(5), 627–635.
- Hucka, M., Bergmann, F. T., Dräger, A., Hoops, S., Keating, S. M., Le Novère, N., ... Wilkinson, D. J. (2015). Systems Biology Markup Language (SBML) Level 2 Version 5: Structures and facilities for model definitions. *Journal of Integrative Bioinformatics*, *12*(2), 271.
- Ibarra, R. U., Edwards, J. S., & Palsson, B. O. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, *420*(6912), 186–189.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30.
- Karp, P. D., Billington, R., Caspi, R., et al. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, *20*(4), 1085–1093.
- Kauffman, K. J., & Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, *14*(5), 491–496.
- Klamt, S., Saez-Rodriguez, J., & Gilles, E. D. (2007). Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, *1*, 2.
- Kuang, E., Marney, M., Cuevas, D., Edwards, R. A., & Forsberg, E. M. (2020). Towards predicting gut microbial metabolism: Integration of flux balance analysis and untargeted metabolomics. *Metabolites*, *10*, 156.
- Kumar, V. S., Dasika, M. S., & Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, *8*(1), 212.
- Lakshmanan, M., Koh, G., Bevan, K., Chung, S., & Lee, D. Y. (2012). Software applications for flux balance analysis. *Briefings in Bioinformatics*, *15*(1), 108–122.
- Lee, D. Y., Yun, H., & Park, S. (2003). MetaFluxNet: The management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics Applications Note*, *19*(16), 2144–2146.
- Lee, J. M., Gianchandani, E., & Papin, J. A. (2006). Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, *7*(2), 140–150.
- Lee, S., Palakornkule, C., Domach, M. M., & Grossmann, I. E. (2000). Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering*, *24*(2), 711–716.
- Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., ... Laibe, C. (2010). BioModels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, *4*, 92.
- Liu, J., Chan, S. H. J., Chen, J., Solem, C., & Jensen, P. R. (2019). Systems biology—A guide for understanding and developing improved strains of lactic acid bacteria. *Frontiers in Microbiology*, *10*, 876.
- Mahadevan, R., Burgard, A. P., Famili, I., Van, D. S., & Schilling, C. H. (2005). Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals. *Biotechnology and Bioengineering*, *10*(5), 408–417.
- Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal*, *83*(3), 1331–1340.
- Nalbantoglu, S. (2019). *Metabolomics: Basic principles and strategies, molecular medicine*. IntechOpen, 10.5772/intechopen.88563.
- Nicolas, A. L., Dirk, E. M., & Peter, J. S. (2013). Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Applied Microbiology and Biotechnology*, *97*, 8729–8739.
- Oliveira, A. P., Nielsen, J., & Förster, J. (2005). Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiology*, *5*, 39.
- Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Molecular Systems Biology*, *7*, 535.
- Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248.
- Pandey, V., Hadadi, N., & Hatzimanikatis, V. (2019). Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Computational Biology*, *15*(5), e1007036.
- Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., & Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, *22*(8), 400–405.
- Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., ... Yoon, Y. J. (2019). A review of the microbial production of bioactive natural products and biologics. *Frontiers in Microbiology*, *10*, 1404.
- Putri, S. P., Nakayama, Y., Matsuda, F., Uchikata, T., Kobayashi, S., Matsubara, A., & Fukusaki, E. (2013). Current metabolomics: Practical applications. *Journal of Bioscience and Bioengineering*, *115*(6), 579–589.
- Quek, L. E., Wittmann, C., Nielsen, L. K., & Krömer, J. O. (2009). OpenFLUX: Efficient modelling software for 13C-based metabolic flux analysis. *Microbial Cell Factories*, *8*, 25.
- Radhakrishnan, M., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal*, *83*, 1331–1340.

- Ramakrishna, R., Edwards, J. S., McCulloch, A., & Palsson, B. O. (2001). Flux-balance analysis of mitochondrial energy metabolism: Consequences of systemic stoichiometric constraints. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 280(3), 695–704.
- Raman, K., & Chandra, N. (2009). Flux balance analysis of biological systems: Applications and challenges. *Briefings in Bioinformatics*, 10(4), 435–449.
- Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., & Palsson, B. O. (2006). Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), 17480–17484.
- Rieckeberg, E., & Powers, R. (2017). New frontiers in metabolomics: From measurement to insight. *Research*, 1(3), 1146–1148.
- Riley, M. L., Schmidt, T., et al. (2007). PEDANT genome database: 10 years online. *Nucleic Acids Research*, 35(1), 354–357.
- Rocha, I., Maia, P., Evangelista, P., et al. (2010). OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 4, 45.
- Sailwal, M., Das, A. J., Gazara, R. K., Dasgupta, D., Bhaskar, T., Hazra, S., & Ghosh, D. (2020). Connecting the dots: Advances in modern metabolomics and its application in yeast system. *Biotechnology Advances*, 44, 107616.
- Savinell, J. M., & Palsson, B. O. (1992). Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *Journal of Theoretical Biology*, 154(4), 421–454.
- Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., & Palsson, B. O. (2011). Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols*, 6(9), 1290–1307.
- Schilling, C. H., Letscher, D., & Palsson, B. O. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203(3), 229–248.
- Schomburg, I., Chang, A., & Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 30(1), 47–49.
- Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3, 119.
- Schwarz, R., Musch, P., Kamp, A., Engels, B., Schirmer, H., Schuster, S., & Dandekar, T. (2005). YANA—A software tool for analyzing flux modes, gene-expression and enzyme activities. *BMC Bioinformatics*, 6(1), 135.
- Segre, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), 15112–15117.
- Shlomi, T., Berkman, O., & Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21), 7695–7700.
- Tomar, N., & De, R. K. (2013). Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene*, 521(1), 1–14.
- Toya, Y., & Hiroshi, S. (2013). Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnology Advances*, 31(6), 818–826.
- Urbanczik, R. (2006). SNA—A toolbox for the stoichiometric analysis of metabolic networks. *BMC Bioinformatics*, 7, 129.
- Varma, A., & Palsson, B. O. (1994). Metabolic flux balancing: Basic concepts, scientific and practical use. *Biotechnology*, 12(10), 994–998.
- Wiechert, W. (2001). 13C metabolic flux analysis. *Metabolic Engineering*, 3(3), 195–206.
- Wiechert, W. (2002). Modeling and simulation: Tools for metabolic engineering. *Journal of Biotechnology*, 94(1), 37–63.
- Wittig, U., Kania, R., Golebiewski, M., et al. (2012). SABIO-RK—Database for biochemical reaction kinetics. *Nucleic Acids Research*, 40, 790–796.
- Wright, J., & Wagner, A. (2008). The Systems Biology Research Tool: Evolvable open-source software. *BMC Systems Biology*, 2, 55.
- Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2009). Translational biomarker discovery in clinical metabolomics: An introductory tutorial: What is metabolomics all about? *BioTechniques*, 46, 363–365.
- Zamboni, N., Fendt, S. M., Rühl, M., & Sauer, U. (2009). (13)C-based metabolic flux analysis. *Nature Protocols*, 4(6), 878–892.
- Zomorodi, A. R., Suthers, P. F., Ranganathan, S., & Maranas, C. D. (2012). Mathematical optimization applications in metabolic networks. *Metabolic Engineering*, 14(6), 672.