PREFACE

Metabolomics is the logical progression of the study of genes, transcripts and proteins. Nutrients, gut microbial metabolites and other bioactive food constituents interact with the body at system, organ, cellular and molecular levels, and effect the expression of genome at several levels, and subsequently, the production of metabolites. This book presents an overview of nutrigenomics and metabolomics tools, and their perspective in livestock health and production. In addition, this book describes how lists of masses (molecular ions) and mass unit bins of interest are searched within online databases for compound identification, the extra biochemical data required for metabolite confirmation, how data are visualized and what the putative and protein sequences are associated with observed metabolic changes. Moreover, environmental metabolomics is the application of metabolomics to the investigation of both free-living organisms directly obtained from the natural environment or laboratory conditions. This book outlines some of the advances made in areas of plant environmental metabolomics. The applications of microbial metagenomics, the use of genomics techniques to the study of communities of directly in their diverse natural environments, are explored as well. Other chapters examine the abnormalities in metabolism of cancer cells, which could play a strategic role in tumour initiation and behavior.

As explained in Chapter 1, metabolism represents a complex system characterized by a high variability in metabolites' structures, concentrations and regulation ratios. Metabolic information can be stored in and analysed from metabolomic matrix consisting of concentrations of different metabolites analysed in different individuals (subjects). From such a matrix, different relationships can be highlighted between metabolites through a correlation analysis between their levels. When the set of all the metabolites are considered, their levels can be converted into ratios representing their metabolic regulations by reference to their metabolic profile. The complexity of network resulting from all the metabolic profiles can be structured by classifying the different profiles into different homogeneous groups representing different metabolic trends. Beyond the correlations between metabolites and their associations to different metabolic trends, a third variability can be observed consisting of atypical or original profiles in the population due to atypical values for some metabolites. Such cases provide information on extreme states in the studied population or on new emergent populations. Extreme cases are detected by combining analysis of variables with that of profiles leading to the outlier diagnostics. These three statistical aspects of variability analysis of metabolomic datasets are detailed in this chapter by different numerical examples and illustrations. Additionally to these correlation and distance matrices-based approaches,

the chapter gives a background on different other metabolomic approaches based on other criteria/constraints/information stored in other types of matrices. According to the context, such matrices can contain (a) binary codes formulating the adjacencies between metabolites, (b) stoichiometric coefficients of metabolic reactions, (c) transition probabilities between different metabolic states, (d) partial derivatives of the system according to small perturbations, (e) contributions of different metabolic pathways, etc. Such matrices are used to describe/handle the complex structures, processes and evolutions of metabolic systems. General applications and interests of these different matrix-based approaches are illustrated in a first general section of the chapter, followed by a second detailed section on the correlation and distance-based analyses.

As discussed in Chapter 2, during the last decades compelling evidence has accumulated indicating that abnormalities in metabolism of cancer cells could play a strategic role in tumour initiation and behaviour. Abnormalities in metabolism are likely a consequence of several alterations in the complex network of signal transduction pathways, which may be caused by both genetic and epigenetic factors. An aberrant energy metabolism was recognized as one of the prominent features of the malignant phenotype, since the pioneering work of Warburg. It is now well established that the majority of tumours is characterized by a high glucose consumption, even under aerobic conditions, in absence of the Pasteur Effect, i.e. the lack of inhibition of glycolysis when cancer cells are exposed to normal oxygen consumption. Several investigators provided experimental data in support of a specific structure of the metabolic network in cancer cells. The 'tumour metabolome' has been defined as the metabolic tumour profile characterized by high glycolytic and glutaminolytic capacity and a high channelling of glucose carbons toward synthetic processes.

Despite no archetypal cancer cell genotype exists, facing the wide genotypic heterogeneity of each tumour cell population, some malignant features (i.e. invasion, uncontrolled growth, apoptosis inhibition, metastasis spreading) are virtually shared by all cancers. This paradox of a common clinical behaviour despite marked both genotypic and epigenetic diversity needs to be investigated by a Systems Biology approach and suggests that cancer phenotype should be considered as a sort of "attractor" in a specific space phase defined by thermodynamic and kinetic constraints. This is not the only phase space cancer cells are embedded into: in principle cancer cells, like any living entity travel along an integrated set of genetic, epigenetic or metabolomic parameters. A fractal dimension formalism can be used in a prospective reconstruction of cancer attractors. Studies conducted on MCF-7 and MDA-MB-231 breast cancer cells, exposed to different morphogenetic fields, show that metabolomic profile correlates to cell shape: modification of cell shape and/or architectural characteristics of the cancer- tissue relationships, induced through manipulation of environmental cues, are followed by significant modification of the cancer metabolome as well as of the fractal dimensions at both single cell and cell population level. These results suggest how metabolomic shifts in cancer cells need to be considered as an adaptive modification adopted by a complex system under environmental constraints defined by the non-linear thermodynamic of the specific attractor occupied by the system. Indeed, characterization of cancer cells behaviour by means of both metabolomic and fractal parameters could be used to build an operational and meaningful space phase, that could help in evidencing the transitions boundaries as well as the singularities of cancer behaviour. Hence, by revealing tumour-specific metabolic shifts in tumour cells, metabolic profiling enables drug developers to identify the metabolic steps that control cell proliferation, thus

aiding the identification of new anti-cancer targets and screening of lead compounds for anti-proliferative metabolic effects.

As discussed in Chapter 3, molecular biology has recently concentrated on the determination of multiple gene-expression changes at the RNA level (transcriptomics), and into determination of multiple protein expression changes (proteomics). Similar developments have been taking place at metabolite small-molecule level, leading to the increasing expansion in studies now termed metabolomics. This approach can be used to provide comprehensive and simultaneous systematic profiling of metabolite levels in biofluids and tissues, and their systematic and temporal changes. Analysis of metabolites is not a new field; long prior to the development of the various "omics" approaches, the simultaneous analysis of the plethora of metabolites seen in biological fluids had been carried out largely, but historically it has been limited to relatively small numbers of target analytes. However, the realization that metabolic pathways do not act in isolation but rather as part of an extensive network has led to the need for a more holistic approach to metabolite analysis.

The main analytical techniques employed for metabolomics studies are based on NMR spectroscopy and mass spectrometry (MS), that, in turn, can be considered complementary each other. Neverthless, MS measurement following chromatographic separation offers the best combination of sensitivity and selectivity, so it is central to most metabolomics approaches. Either gas chromatography after chemical derivatization, or liquid chromatography (LC), with the newer method of ultrahigh-performance LC being used increasingly, can be adopted. Capillary electrophoresis coupled to MS has also shown some promises. Analyte detection by MS in complex mixtures is not as universal as for NMR and quantitation can be impaired by variable ionization and ion-suppression effects. A LC chromatogram is generated with MS detection, usually using electrospray ionization (ESI), and both positive- and negative-ion chromatograms can be recorded. The utilization of nano-ESI can reduce ionization suppression effects due to the increased ionization efficiency. Mass analyzer able to produce high mass resolution, mass accuracy, and tandem MS, such as quadrupole-time-of-flight (Q-TOF) or high-resolution ion trap instruments, are employed.

Direct infusion (DI)-MS/MS using Fourier transform ion cyclotron resonance mass spectrometers provides a sensitive, high-throughput method for metabolic fingerprinting. Unfortunately, DI-MS analysis is particularly susceptible to ionization suppression arising from competitive ionization. In metabolomics, matrix assisted laser desorption-ionization (MALDI) has largely been confined to the targeted analysis of high-molecular weight metabolites due to the substantial signals generated by the matrix in the low-molecular-weight region (<1,000 m/z). Recent advancements in laser desorption techniques include desorption-ionization MS from porous silicon chips and matrices that have minimal background signals in the low-molecular-weight region. These offer new opportunities for the utilization of MALDI ionization in metabolite screening and fingerprinting employing MALDI-TOF/TOF. However, the technique is still subject to ion suppression and yields poor quantitative detection. Desorption ESI (DESI), a new ambient, soft-ionization technique that combines features from both ESI and desorption-ionization methods, allows the direct analysis of animal and plant tissues. However, DESI experimental conditions typically require optimization for each sample type, so time must be invested initially in optimizing the experimental parameters.

It was quoted in 1953 at the 'Changing flora of Britain' conference that 'we should mobilize a team which could tackle the problems, genetical, cytological, physiological,

ecological and chemical, and see whether out of the available mass of material we can not only reach a settled nomenclature... but make a serious contribution to the problems of evolution' (Raven 1953). Nearly 60 years later, we are now starting to assemble such genomic and post-genomic teams with the appropriate infrastructure, technology and bioinformatic power to answer questions in plant ecology and evolution. Of course, the chemical component of the team can now be termed environmental metabolomics and is progression of the study of genes (genomics), mRNA (transcriptomics) and proteins (proteomics).

The main intention of plant metabolomics research is to provide an unbiased assessment of metabolism across multiple pathways. Ideally, all plant metabolites should be identified and quantified at a relevant temporal and spatial scale by untargeted metabolomic fingerprinting using mass spectrometry or NMR or by targeted, quantitative metabolite profiling; to provide a comprehensive view of metabolism. Such global screening of the metabolites has been termed biochemical, or metabolic phenotyping. This approach builds on the much valid work carried out by plant biologists such as Richard Dixon and Jeffrey Harborne to name but a very few. However, the ease of application and software to analyse results, alongside the increase in interdisciplinary science, has opened up such technology to more research fields to answer a wider range of questions.

Chapter 4 will outline some of the advances made in such areas of plant environmental metabolomics.

As explained in Chapter 5, despite enormous advancements in microbial culturing methods, more than 95% of the global microbial diversity still remains cryptic. Microbial metagenomics- the applications of modern genomics techniques to the study of communities of microbes directly in their diverse natural environments, bypassing the need for isolation, is changing our comprehension of the biosphere. Advances in technologies designed to access this wealth of genetic information through environmental nucleic acids extraction and analysis have provided the means of overcoming the limitations of conventional culture-dependent microbial exploitation. Further developments and applications of these methods promise to provide opportunities to link distribution and identity of gut microbes in their natural habitats, and explore their use for promoting livestock health and industrial biotechnological applications.

Nutrition exhibits the most important life-long environmental impact on health. Nutrients, gut microbial metabolites and other bioactive food constituents interact with body at system, organ, cellular and molecular levels, and affect the expression of genome at several levels, and subsequently, the overall production of metabolites. Direct measurement of cellular metabolites is essential for the study of biological processes, and may allow causes of disease, toxicological progression, and novel disease-biomarkers to be identified. Advances in analytical techniques and the algorithms for management of the data has allowed a precise and global analysis of biological substances such as DNA (genomics), RNA (transcriptomics), proteins (proteomics) and smaller molecules (metabolomics). Holistic "**omics**" approaches are indispensable to cover the complex nutrient-cell and gut microbial-host interactions. Chapter 6 presents an overview of nutrigenomics and metabolomics tools with reference to their perspective in livestock health and production.

Metabolomics is a rapidly growing field with the goal of measuring and interpreting the complex time and condition dependent concentration, activity or flux of metabolites in cells, tissues and other biosamples. On the other side, the integrated approach to studying biological

systems in Systems Biology has led to significant improvement of our understanding of such systems. Since biological circuits are hard to model and simulate, many efforts are being made to develop computational models that can handle their intrinsic complexity. However, a large part of the biological networks remains unknown and hard to understand and Metabolomics technology that allows simultaneous acquisition of many metabolite measurements can lead to further analysis for discovering novel pathway components and unknown network relationships. Metabolic networks are structurally complex and behave in a stochastic fashion. In Chapter 7 the authors describe how symbolic-statistical machine learning techniques can be used to reconstruct metabolic networks from metabolic profiling data. The authors show that symbolic machine learning methods have the power to model structural and relational complexity while statistical machine learning ones provide principled approaches to uncertainty modeling. They apply a symbolic-statistical learning framework to analyze sequences of reactions for biologically active paths in metabolic networks. The authors show through experiments that their approach provides a robust methodology for machine reconstruction of metabolic networks from metabolomic data.

As discussed in Chapter 8, generally, a large proportion of the genes in any genome encode enzymes of primary and specialized (secondary) metabolism [1]. Not all primary metabolites, those that are found in all or most species, have been identified and only a small portion of the estimated hundreds of thousand specialized metabolites, those found only in restricted lineages, have been studied in any species [1]. Fridman and Pichersky [1] noted that the correlative analysis of extensive metabolic profiling and gene expression profiling had proven a powerful approach for the identification of candidate genes and enzymes, particularly those in secondary metabolism [2]. It is rapidly becoming possible to measure hundreds or thousands of metabolites in small samples of biological fluids or tissues. Arita [3] said that metabolomics, a comprehensive extension of traditional targeted metabolite analysis, had recently attracted much attention as the biological missing pieces that can complement transcriptome and proteome analysis. Metabolic profiling applied to functional genomics (metabolomics) is in an early stage of development [4]. Fridman and Pichersky [1] said that the final characterization of substrates, enzymatic activities, and products requires biochemical analysis, which had been most successful when candidate proteins have homology to other enzymes of known function. To facilitate the analysis of experiments using post-genomic technologies, new concepts for linking the vast amount of raw data to a biological context have to be developed [5]. Visual representations of pathways help biologists to understand the complex relationships between components of metabolic network [5].

Organ function can only be completely understood through knowledge of molecular and cellular processes within the constraints of structure-function relations at the tissue level [6]. Knowledge on integrative computational physiology is required. Cellular components interact with each other to form networks that process information and evoke biological responses [7]. Today different database systems for molecular structures (genes and proteins) and metabolic pathways are available. All these systems are characterized by the static data representation [8]. For progress in biotechnology the dynamic representation of this data is important. The metabolism can be characterized as a complex biochemical network [8]. A deep understanding of the behavior of these networks requires the development and analysis of mathematical models [7]. Computer modeling of metabolic networks can help better understand complex metabolism [9 - 10]. As previously mentioned, mathematical modeling is

one of the key methodologies of metabolic engineering [11]. Based on a given metabolic model different computational tools for the simulation, data evaluation, systems analysis, prediction, design and optimization of metabolic systems have been developed [11]. More details on mathematical modeling can be seen in another specific chapter in this book. In additional to mathematical model, graph-based analysis of metabolic networks is another widely used technique in metabolomics [12].

Various types of evidence have implicated estrogens in the etiology of human breast cancer [1-8]. They are generally thought to cause proliferation of breast epithelial cells through estrogen receptor-mediated processes [4]. Rapidly proliferating cells are susceptible to genetic errors during DNA replication, which, if uncorrected, can ultimately lead to malignancy. While receptor-mediated processes may play an important role in the development and growth of tumors, accumulating evidence suggests that specific oxidative metabolites of estrogens, if formed, can be endogenous ultimate carcinogens that react with DNA to cause the mutations leading to initiation of cancer [6-9]. Thus, estrogen metabolites, specifically catechol estrogen-3,4-quinones, are hypothesized to be endogenous initiators of breast, prostate and other human cancers.

Several lines of evidence, including metabolism and carcinogenicity studies by Liehr and coworkers, led to the recognition that the 4-hydroxylated estrogens play a major role in the genotoxic properties of estrogens [1-3]. In Chapter 9, the authors have hypothesized that the estrogens estrone (E_1) and estradiol (E_2) initiate breast and other human cancers by reaction of their electrophilic metabolites, catechol estrogen-3,4-quinones [$E_1(E_2)$-3,4-Q], with DNA to form depurinating adducts [5-8]. These adducts generate apurinic sites leading to mutations that may initiate breast, prostate and other human cancers [6-9].