Advances in Experimental Medicine and Biology 1336
Proteomics, Metabolomics, Interactomics and Systems Biology

Ana Valéria Colnaghi Simionato *Editor*

Separation Techniques Applied to Omics Sciences

From Principles to Relevant Applications

Springer

Advances in Experimental Medicine and Biology

Proteomics, Metabolomics, Interactomics and Systems Biology

Sub-Series Editor

Daniel Martins-de-Souza
University of Campinas (UNICAMP)
Institute of Biology
Laboratory of Neuroproteomics
Campinas, Brazil

Advances in Experimental Medicine and Biology presents multidisciplinary and dynamic findings in the broad fields of experimental medicine and biology. The wide variety in topics it presents offers readers multiple perspectives on a variety of disciplines including neuroscience, microbiology, immunology, biochemistry, biomedical engineering and cancer research.

Advances in Experimental Medicine and Biology has been publishing exceptional works in the field for over 30 years and is indexed in Medline, Scopus, EMBASE, BIOSIS, Biological Abstracts, CSA, Biological Sciences and Living Resources (ASFA-1), and Biological Sciences. The series also provides scientists with up-to-date information on emerging topics and techniques.
2020 Impact Factor: 2.622 5 Year Impact Factor: 2.324

This series of volumes focuses on concepts, techniques and recent advances in the field of proteomics, interactomics, metabolomics and systems biology. Recent advances in various 'omics' technologies enable quantitative monitoring of myriad various biological molecules in a high-throughput manner, and allow determination of their variation between different biological states on a genomic scale. Now that the sequencing of various genomes, from prokaryotes to humans, has provided the list and linear sequence of proteins and RNA that build living organisms, defining the complete set of interactions that sustain life constitutes one of the key challenges of the postgenomic era. This series is intended to cover experimental approaches for defining protein-protein, protein-RNA, protein-DNA and protein-lipid interactions; as well as theoretical approaches dealing with data analysis, integration and modeling and ethical issues.

More information about this series at http://www.springer.com/series/15040

Ana Valéria Colnaghi Simionato
Editor

Separation Techniques Applied to Omics Sciences

From Principles to Relevant Applications

Springer

Editor
Ana Valéria Colnaghi Simionato
Analytical Chemistry Department
University of Campinas, Institute of Chemistry
Campinas, São Paulo, Brazil

Proteomics, Metabolomics, Interactomics and Systems Biology
ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-030-77251-2 ISBN 978-3-030-77252-9 (eBook)
https://doi.org/10.1007/978-3-030-77252-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my beloved children, Isabela and Rafael: my love for you is the size of the sky!

Preface

Omic sciences have been of ultimate importance to comprehend the complex biochemical reactions and related events that occur in a biological system. The classical central dogma of molecular biology, which states that genetic information flows unidirectionally from DNA to RNA and then to proteins, has been gradually replaced and complemented by the systems biology approach. This multidisciplinary approach tries to explain the biological system as a whole, where the entire organism is influenced by a variety of internal events as well as the environment, showing that each level of the biological information flux may influence the previous or the subsequent one.

Separation techniques constitute the first primordial dimension to obtain comprehensive data on biological samples analyses. The second dimension method has often been the hybridization of separation techniques with mass spectrometry and, more rarely, nuclear magnetic resonance.

This book presents liquid chromatography, gas chromatography, and capillary electrophoresis, the three main separation techniques lately available, applied to key omic sciences, such as proteomics, metabolomics, peptidomics, glycomics, and foodomics. Additionally, important directions on proteomics and metabolomics large set of data analyses are also approached. The fundamentals of each technique will not be covered herein. Instead, the recent advances in such techniques will be presented focusing on the application to omics analyses and unique aspects in each case. Therefore, this book intends to offer wide ranging options available to researchers on omics sciences, and how to integrate them in order to achieve the comprehension of a biological system as a whole.

Campinas, São Paulo, Brazil — Ana Valéria Colnaghi Simionato

Acknowledgments

Firstly, I would like to thank Daniel Martins de Souza for the invitation to edit this book and for trusting my ability to fulfill the expected quality of this series. The acknowledgments are extended to Springer's editorial team.

To all authors and co-authors who have accepted my invitation to contribute for the completion of this work, my sincere acknowledgment. I had devised initial topics in my mind and you have made them come true with excellence.

To my students who have directly or indirectly assisted me, either co-authoring some of the chapters with me, or performing their research projects with the expected quality so that our group has been acquiring more and more practical knowledge about metabolomics and separation techniques.

To all my teachers and professors (from kindergarten to graduation): it is not possible nor fair to cite the names of all of them (cause maybe I could forget some), but each one has profoundly contributed to my professional achievements and to my personal formation.

To my parents (Itamar Simionato and Maria Clélia Colnaghi Simionato), my daughter (Isabela Simionato Cantu), and my son (Rafael Simionato Cantu) for all support and love.

To the National Council for Scientific and Technological Development (CNPq), São Paulo Research Foundation (FAPESP), and Coordination for the Improvement of Higher Education Personnel (CAPES), who have supported research in Brazil, and, especially, my research group.

Contents

The New Omics Era into Systems Approaches: What Is the Importance of Separation Techniques?

Flávia da Silva Zandonadi, Fábio Santos Neves, Elisa Castañeda Santa Cruz, Alessandra Sussuilini, and Ana Valéria Colnaghi Simionato

Abstract

Omics sciences have been facing challenges in different fields, especially in life sciences. One of these challenges involves assessing biology into systems interpretation. With the advance of genomics, molecular biology has been projected into the realm of systems biology. In a different direction, systems approaches are making definitive strides toward scientific understanding and biotechnological applications. Separation techniques provided meaningful progress in the omics era, conducting the classical molecular biology to contemporary systems biology. In this introductory chapter, the relevance of these techniques to the development of different omics sciences, within the systems biology context, will be discussed.

Keywords

Systems biology · Genomics · Proteomics · Metabolomics · Separation techniques

1 Systems Biology

How are biologists building strategies to understand life? This question has been methodically surveyed by exploring the characteristics of living organisms in different ways. Moreover, as a reflection of the necessity in deciphering the biological dynamics, the progress of cost-effective technologies able to comprehensively assess DNA, RNA, protein, and metabolites, molecules that orchestrate all the biological dynamics, has also been promoted. Before defining systems biology and omics sciences, it is important to describe the main approaches applied in life sciences since the beginning of these studies.

In 2004 Westerhoff and Palsson developed a series of arguments about two scientific schools considering their origin in the expansion of molecular biology to genome-wide analyses [1]. The idea of biology using integration tools is not new. The first regulatory circuit within the molecular biology context was described more than 40 years ago [1,

F. da Silva Zandonadi · F. S. Neves · E. C. S. Cruz
Department of Analytical Chemistry, Institute of Chemistry, University of Campinas, Campinas, SP, Brazil

A. Sussuilini
Department of Analytical Chemistry, Institute of Chemistry, University of Campinas, Campinas, SP, Brazil

National Institute of Science and Technology for Bioanalytics, Institute of Chemistry, University of Campinas, Campinas, SP, Brazil

A. V. C. Simionato (✉)
Analytical Chemistry Department, University of Campinas, Institute of Chemistry, Campinas, São Paulo, Brazil
e-mail: avsimionato@unicamp.br

A. V. Colnaghi Simionato (ed.), *Separation Techniques Applied to Omics Sciences*, Advances in Experimental Medicine and Biology 1336, https://doi.org/10.1007/978-3-030-77252-9_1

2]. These studies were conducted focusing on the regulatory mechanisms, admittedly on a small scale. Molecular biologists began to apply systems approaches to unravel the molecular components and the logic that underlie cellular processes, often in parallel with the characterization of individual macromolecules. High-throughput technologies have made the scale of such inquiries much larger, enabling the view of the genome, for example, as the "system" in study [3–5].

As mentioned before, the dynamics of life is assessed by the studies of DNA, RNA, proteins, and metabolites. Behind these molecules, studies were addressed under investigation of multiple escalation levels, i.e., molecular, cellular, organism, and ecological organization. Survey by reductions, as defined by the classical scientific method, explores complexity in its individualized parts. The biological system has been dissected into their constituent parts and explained according to the chemical basis of numerous living processes [6, 7], thus producing multifaceted and disconnected knowledge. Beyond the philosophy of biology, reductionism method could be encompassed, according to Brigandt and Love [7]:

> … a set of ontological, epistemological, and methodological claims about the relations between different scientific domains. The basic question of reduction is whether the properties, concepts, explanations, or methods from one scientific domain (typically at higher levels of organization) can be deduced from or explained by the properties, concepts, explanations, or methods from another domain of science (typically at lower levels of organization).

From this definition, the multifaceted and disconnected knowledge of the biological dynamics could no longer be questioned under reductionist pragmatism. Automation, miniaturization, and multiplexing of various assays led to the generation of additional omics data types [8]. These enormous amounts of information that come from the omics sciences (such as genomics, proteomics, and metabolomics) no longer could be interpreted under compartmentalization contexts but as an integrated system.

For this reason, a more formal and mechanistic framework was required to analyze multiple high-throughput data types systematically [9, 10]. At this moment, the structure of scientific theories, the relations between scientific disciplines, the nature of explanation, the diversity of methodology, and aspects of biological complexity, especially the regulation process under environmental influences, changed to system status. Not only biology but other areas reached the center of the status of a new concept, the science of complex systems.

The knowledge from classical chemistry has provided human and technological resources, from analytical techniques to recent developments in high-throughput approaches and bioinformatics. The biological investigation from the bewildering diversity of interactions and regulatory networks has produced a formal and mechanistic framework necessary to analyze multiple high-throughput data types systematically [10], as represented in a summarized timeline inserting omics sciences into systems biology (Fig. 1).

Systems biology describes changes and connects variables over time, taking the chaotic, unpredictable, or counterintuitive contrasting with much simpler linear system properties from the reductionist methods. Defined as a new level of understanding and capturing of the dynamics of large sets of interacting components, the field combined the molecular and cell biology approaches with a stronger commitment toward quantitative experimentation under physiological conditions and with formal mathematical modeling (e.g., R-Theory) [11].

Unquestionably, the mathematical models are crucial for handling the associated complexity as formal representations to system-level understanding, but the main idea of the next topics is to describe the progress of main separation techniques, developments, and important milestones of the evolution of molecular biology into systems biology.

2 Omics Sciences and the Systems Biology Era

The identification, qualification, and application of diagnostic and prognostic biomarkers remain the holy grail of the current omics paradigm. Genomics, moving on to proteomics and metabolomics, premise, and promise of systems biology, has provided a powerful motivation for scientists

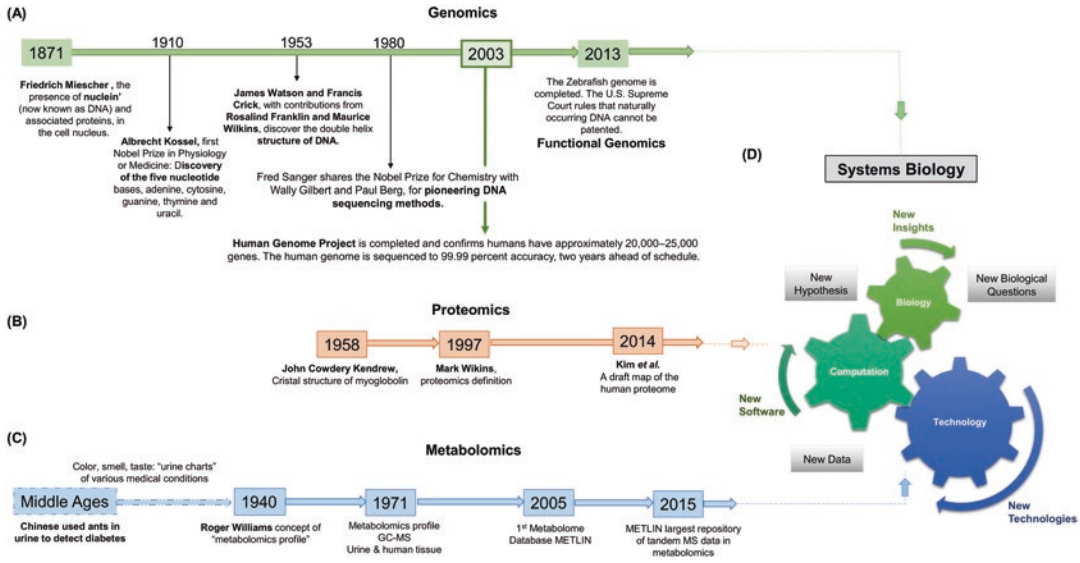

Fig. 1 High-throughput omics timeline to systems biology. From classical to modern science, how the high-throughput studies and multidisciplinary interpretations provide the resurgence of entering biology as a system. Classical genetics to genomics (**a**), protein biochemistry to proteomics (**b**), metabolomics and emergence of the next field (**c**), systems biology (**d**)

to combine the data generated from multiple omics approaches (e.g., genomics, transcriptomics, proteomics, and metabolomics) to create a holistic understanding of cells, organisms, and communities, relating to their growth, adaptation, development, and progression to disease. In this section, some tools and advanced techniques in genomics, proteomics, and metabolomics are summarized. These techniques are essential tools to the new challenges in life sciences, highlighting the systems biology field into the paradigm from omics sciences.

2.1 Genomics to Systems Biology, Post-Genomics Era

Around many definitions related to the genomics field, the most classic and simplistic one is the study of the complete genome of organisms, aiming at decoding and identifying relationships among the gene set, growth, and development of the organisms. This science is dedicated to determine the complete sequence of organisms' DNA, or mapping a smaller genetic scale.

DNA sequencing history began in 1977 when Frederick Sanger and colleagues described a methodology for determining the sequential order of nucleotides that make up the structure of DNA, based on the principle of controlled termination of dideoxynucleotide replication [12]. In its original version, this method was not a suitable tool for sequencing complex genomes, covering thousands, millions, or even billions of base pairs in complex organisms (e.g., mammals). Genome assembly was used to reconstruct the exact gene disposition and to locate other genome components in the chromosome, since there was no computer software available to analyze the generated sequences and sort them correctly.

One of the most significant challenges of these first steps into the genomics field has been the development of new DNA sequencing going through different achievements along all over the next 30 years. From these first 10 years, genome studies have made rapid progress in three genera-

tions of gene separation and sequencing techniques.

2.1.1 First-Generation Sequencing (Gel-Based Sequencing)

(a) Manual slab gel: Sanger et al. [12] and Gilbert and Maxam [13] developed sequencing by chemical fragmentation techniques or chain termination, coupled with gel electrophoresis-based size separation [13, 14]. The method required a labeled DNA primer, which could be labeled by fluorescence or radiation. After DNA fragments were separated and the bands visualized, the sequence was manually read from the pattern of the four parallel runs.

(b) Automatic slab gel: Developed by Smith and colleagues [15]. The key differences between this method and the former one were the tagging and the number of reactions. By tagging four different dyes of different fluorescent emission wavelengths instead of one, the four reactions were reduced to a single one [15], which demonstrated optimization in gel (such as diameter and length), electrophoresis conditions (such as current and temperature), optics and electronics used in data acquisition, and software used in data reduction (72 Kb/h/slab).

(c) Capillary gel electrophoresis: Cohen et al. [16] demonstrated the use of polyacrylamide gel-filled in capillaries used to reach single-nucleotide separation of DNA oligonucleotide markers by UV detection [16], which was improved by using ultrasensitive fluorescence by Swerdlow and Gesteland [17], verifying enhanced speed, resolution, and efficiency comparing with the former methods [17]. Furthermore, DNA sequencing using capillary array electrophoresis was developed by Huang et al. [18], which performs rapid, parallel separation followed by on-column detection using multicolor, confocal fluorescence scanner.

Genomics began only in the late 1980s, after Sanger's method was modified to allow automated sequencing and integration with a computerized reading system [19], enabling these processes to be carried out on a large scale within a high-performance platform. Thus, a leading project that drove genomics into the research race entitled the Human Genome Project was initiated, which is still considered one of the boldest scientific projects in history. It began in 1984 and was developed in the subsequent years by a consortium of scientists from the United States, the United Kingdom, Japan, France, Germany, and China, with the financial support of their respective governments [20–22]. The next two further methods are known as *next-generation sequencing* (NGS) due to its parallel and fast high-throughput sequencing platforms promoting the degree of sequence coverage and accuracy of individual reads compared to Sanger's one.

2.1.2 Second-Generation Sequencing

The sequence by synthesis (SBS) method requires the direct action of DNA polymerase to produce the visible result. Considering Sanger's method is also based on sequence by synthesis, Nyrén [23] developed a new technique by using the luminescent method for measuring pyrophosphate synthesis known as pyrosequencing [23]. A few years later, Ronaghi et al. [24] and Nyrén [23] performed real-time sequencing by synthesis within a proper choice of enzyme and substrate in a solid-phase format [24]. This approach was then bought by 454 Life Science (2005), nowadays Roche, which developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using pyrosequencing protocol for solid support and picoliter scale volumes [25].

After that, in 2005, Turcatti et al. developed the "Illumina" sequencing platform composed of four companies, Solexa and Illumina among them. This approach is based on fluorescent reversible terminators for sequencing [26, 27].

2.1.3 Third-Generation Sequencing

Braslavsky and colleagues [28] developed a single-molecule sequencer (SMS), later commercialized by Helicos BioSciences (2009). This technology is capable of sequencing single molecules without DNA amplification through poly-

merase chain reaction (PCR). Moreover, it enhanced the sequencing speed and reduced costs [28, 29]. Puglisi and collaborators developed single-molecule sequencing in real time (SMRT) in 2010. This process enables the observation of DNA synthesis as it occurs in real time, generating very long reads of sequences up to 10 kilobases long, which is useful for de novo genome assemblies [30].

Although the nanopore DNA sequencing methodology started within the second-generation sequencers in the 1990s [31], Oxford Nanopore announced a third-generation single-molecule platform in 2012 that represented a clear step onto sequencing of single DNA molecules using this technology, based on the principle of minute changes in electric current across the nanopore immersed in a conducting fluid with voltage applied when a moving nucleotide, or DNA strand, passes through it and the ion current probes the base identity [32]. All these approaches have brought the cost of human genome sequencing down from US$ 300 million in 2001 to US$1500 in 2015 [33, 34]. Within the improvements in DNA sequencing, physicians can identify a particular type of cancer, enabling them to make better choices for treatments.

In the last three decades, the field was markedly characterized by the proliferation and evolution of technologies, especially for those technologies able to provide new possibilities to decipher the genome from several species and further locate and identify regulatory patterns in gene code. The genomics in their structural fields starts to dive into the necessity in how to elucidate their functions aspects within biological systems and to begin to understand the mechanisms that control interactions [35]. At this time, the gene pool is immediately associated to the conception of a practically static set, while its products, represented by a messenger RNA (transcriptome) and protein (proteome), have a dynamic character, showing continuous changes in response to internal and external stimuli [36]. However, it is known that there is a complex process of regulation, and even at the advanced genome sequencing stage, new research platforms have begun to emerge to integrate individual functional genes and their products (RNAs and proteins) into a global context – a new field in biology named functional genomics [37].

Functional genomics attempts to describe functions and interactions regarding encoded genes and proteins by making use of genome-wide approaches, in contrast to the gene-by-gene approach of classical molecular biology techniques [38]. Biological high-throughput methods were probably the first start for the studies of genes and their regulatory molecules, by applying a data combination derived from the various processes related to DNA sequence, gene expression, and protein function, such as coding and noncoding transcription, protein translation, protein-DNA, protein-RNA, and protein-protein interactions [39].

The classical genomics toward functional genomics studies opened investigations based on the interaction between genes and their products. As mentioned before, high-throughput analytical techniques were the most important tools that allowed biology to have a new perspective as systems rather than study their elements one by one or few at a time, or even as structural elements. Proteins, among the encoded gene products, are vital to living organisms, as they comprise the machinery required for the operation of metabolic pathways.

Genomics and transcriptomics research have progressed due to advances in microarray technology, but protein studies (proteomics) coined to describe the set of proteins encoded by the genome. It was evident, among the omics, that proteomics emerged during the genomics progress as a complementary field regarding to the functional genomics approach. Even though mass spectrometry (MS) is the most common technique used for the detection of analytes in proteomics and metabolomics research, the microarray-based expression [40, 41] and small molecule-based array [42, 43] techniques have been widely used to integrate gene and protein information.

2.2 Proteomics and the Advance in Systems Biology Studies

Proteins, the molecular products of genes, are vital to living organisms, as they comprise the machinery required for the operation of metabolic pathways. Protein expression depends on cellular and environmental conditions. For nearly two decades, proteomics research has attempted to provide the identity and expression level of a large numbers of proteins in different physiological states in cells, body fluids, or tissues. The expectation is that this information will improve the understanding of biological functions and provide molecular signatures for particular health and disease states.

In contrast to mRNA expression analysis, proteomics indicates actual, rather than potential, functional states of a biological system. The bottleneck in proteomics is that there is no amplification step, like a PCR amplification for DNA. The low abundance and high dynamic range of proteins in biological samples, as well as data acquisition and analysis time, remain a challenge in this area. Therefore, proteomics approach drives the continuous development of analytical techniques and bioinformatics tools aiming to deepen the biological functions comprehension.

Proteomics was first defined in 1995, as the protein content complementary to a genome [44]. This concept and new field of science started after the human genome was almost fully sequenced and 24,000 encoded genes was reported [45]. This information brought new challenges into this science area, especially involving the measurement and identification of the amount of gene products, considering the complexity and different regulatory processes. Besides, the gene expression step can be modulated from transcription to the posttranslational modification (PTM) of proteins. Considering this intricate regulatory mechanism, the development of a technique that is able to identify the entire proteome in a single analysis is very challenging [46, 47].

From the definitions established in 1995 to our days, proteomics not only complements the genome but also provides a better biological, phenotypical, and functional understanding of the entire physiology. Considering the emerging omics fields, since proteomics encompasses the quantitative, functional, structural, and PTM characterization of proteins, new areas that make possible the determination of the protein relationships (interactome) and systems biology were developed [48].

In the beginning, a typical proteomics experiment resulted in a list of identified proteins, with no information regarding abundance, distribution, or stoichiometry. Abundance information is critical considering the expression regulation dynamics, reflecting the balance among the entire biochemistry in order to comprise the life dynamics. All the gaps and questions, especially for the stoichiometry balances [49], have been answered according to the advanced analytical capabilities. MS-based proteomics has been extensively used to identify the components of biological systems, and it is the method of choice to consistently quantify the effects of network perturbation in time and space [50–52]. Besides the protein level, life biochemistry operates in multidimensional space. In this way, two important questions have been opened in omics fields: (1) how is this balance achieved and (2) to what extent each of these processes contributes to the regulation of cellular protein abundances. Moreover, back to the regulation events, the biochemistry of living organisms builds up an inside-outside regulatory molecular function system that is the main reservoir for these questions for the research community.

Due to the high complexity of the proteome, there is no particular standard method for sample preparation [53–56]. For the sample preparation procedure based on proteomics, the first and conventional methods used for sample fractionation from complex mixtures are one- and two-dimensional electrophoresis (1-DE and 2-DE, respectively) [56]. Although they are widely used as fractionation techniques, at least in the early days, both suffer from disadvantages including limited dynamic range [57], poor solubilization leading to poor resolution at extremes of isoelectric point (pI) [58], and inability to identify proteins in low abundance [59].

1-DE is traditionally used as a protein fractionation method, mainly when molecular mass is used as separation factor. The technique uses sodium dodecyl sulfate (SDS), a detergent that solves the poor protein solubilization issue. This method is simple, fast, and reproducible and can separate proteins in a broad spectrum based on the molecular mass. 1-DE technique has been used as a step for protein fractionation [60], even though the resolving power is limited when used for separation of high complex protein mixtures. Historically, MS-based proteomics began with the use of 2-DE separation. 2-DE uses two physicochemical proteins properties to separate complex samples obtained from cells and tissues among other biological sources. Thus, this technique occurs in two steps: in the first step, the isoelectric focusing (IEF) separates proteins according to the respective pI; while in the second step, polyacrylamide gel electrophoresis with SDS (SDS-PAGE) separates proteins according to the respective molecular masses (MM). Thousands of proteins can be separated simultaneously. Moreover, information on pI, MM, and relative abundance can be obtained, as well as posttranslational modifications, since they generally cause altered electrophoretic mobility [61]. In general, the generated spots correspond to a single polypeptide chain present in the sample. The development of the gel may determine the number of polypeptide chains, and the amount of each one may be defined with dyes and subsequent densitometric analysis [62].

Advances in this technique to separate proteins were the development of narrow ranges of pH, besides enhanced software tools for correlation analysis of proteins with pH [63–66]. Afterwards, 2-DE was combined to fluorescence probe labeling techniques, giving rise to 2-DE-DIGE (differential in-gel electrophoresis) [67]. The improvement of separation techniques in 2-DE gels was based on the need to minimize the experimental effects, as well as facilitating the comparison step between the samples. 2-DE-DIGE is a technique based on label sample preparation. Samples are labeled separately with different fluorescence probes (Cy2, Cy3, or Cy5) and combined in the same vial so that the run is performed in the 2D gel, thus minimizing the experimental variation and facilitating the comparison between the same protein from different samples (spot matching) [68].

Since the beginning of proteomics, it has heavily relied on 2-DE for the separation and visualization of proteins. Correlated techniques applied to optimize this separation technique still show many inherent drawbacks. 2-DE is costly, insensitive to low copy proteins, scarcely reproducible, and cannot be used for the entire proteome at the dynamic range view [69].

Over the years, several gel-free proteomics techniques have been developed to either fill the gaps left by 2-DE or to entirely abolish the gel-based techniques. Performing proteomics without gel separation of 2-DE electrophoresis gave rise to the term shotgun protein analysis in 1998. Developed by Yates et al., shotgun proteomics consists of the combination of enzymatic protein digestion, followed by separation by liquid chromatography (LC) coupled to sequential (tandem) mass spectrometry (MS/MS) [70]. This technique provided a significant gain in efficiency and sensitivity in the analysis of complex protein mixtures since it automated the process of protein separation, minimizing sample loss, in addition to using nano-flow scale chromatography [71].

The idea of sample complexity, especially for proteins extracted from biological tissues and the wide dynamic concentration range, is still the main challenge for the technique improvements. In this direction, the selective fractionation of complex proteome is an efficient strategy to optimize the identification coverage in complex proteomes. Among the improvements, the multidimensional methods or systems promote the development and comprehension of the proteomics field. The multidimensional methods [72] means that the idea of combining different separation techniques is fundamental [73]. According to the proteome experts, such methods are the way to understand the inherent challenges in gaining insight beyond the “tip of the proteoberg” (The Multidimensional Future of Proteomics, 2016).

Furthermore, the multidimensional techniques based on chromatography allow the separation of these complex mixtures by using multiple columns with different stationary phases, coupled orthogonally, which means that fractions from the first column can be selectively transferred to other columns for additional separation.

Any liquid chromatography separation mode can be used at the protein level, including ion exchange [74, 75], reverse phase [76, 77], hydrophilic interaction [78], or size exclusion [79], prior to digestion. One of the best known multidimensional techniques, also developed by the Yates group, was Multidimensional Protein Identification Technology (MudPIT). Briefly, this technology uses two liquid chromatography separations modes: in the first dimension, proteins are separated in cation exchange columns, according to analytes charge density, while in the second dimension, the separation occurs in reverse phase columns, based on proteins hydrophobicity [79, 80]. This methodology has brought significant improvement in the dynamic range and coverage of the proteome studies. However, the increased process complexity, the low reproducibility, the longer analysis time, and the high cost of analysis are still the main limitations of its use [80, 81].

In the first decade of 2000, important labeling techniques were developed, minimizing the proteome complexity and the analysis time once the different samples were submitted to the same analysis, reducing the variability and the steps in the process [82]. Besides, liquid chromatography coupled with mass spectrometry (LC-MS) also optimizes the proteolytic product separation (e.g., truncated polypeptides), expanding the protein range identification. However, among the analytical techniques, LC has been following the progress and the necessity to improve proteomics separation resolving power, analysis coverage, sensitivity, and throughput. Especially in proteomics, LC is still the best option regarding fractionation of peptide mixtures to enable and maximize identification and quantification of the component peptides by MS, as identification technique.

Implementation of liquid phase separations before MS analysis reduces the number of analytes entering the mass spectrometer at any given time, which minimizes ionization suppression where a nominally detectable species is not detected due to detector dynamic range limitations, and under-sampling in ion selection for MS/MS analysis in shotgun measurements. Furthermore, analytes can be focused within narrow zones (or peaks) during the liquid phase separation steps, which concentrate them and benefits MS detection sensitivity. Within liquid phase separation techniques, LC, especially in its capillary format, has significantly advanced over the past decade to make it a prevalent technique in modern-day proteomics analyses as the physicochemical properties (e.g., mass, charge, and hydrophobicity) of peptides make them amenable to efficient LC separation.

2.3 Metabolomics

Recently, the biology studies used to be discussed focusing on the gene and products of the gene coding, mainly related to transcriptomics and proteomics, whose technologies and tools are very well established and widely applied. However, because of the necessity to fully understand the phenotype diseases caused by DNA mutations, attack of pathogens, and environmental conditions, metabolomics have been an expanding field of research to develop and integrate transcriptomics, proteomics, and metabolomics in a fully systems biology approach [83].

In the post-genomics era, metabolomics became crucial due to its strong relation to phenotype, besides integration with comprehensive transcriptomics and/or proteomics systems to discover specific biomarkers and validate biomarkers significance [84]. In this context, metabolomics has been used to validate and verify the regulation of genes and/or proteins in a biological system and their corresponding variations (e.g., upregulation, downregulation, concentration, or intensity levels) under specific experimental conditions (e.g., different times, gene mutations, biotic or abiotic stress, phenotype alterations).

Moreover, it has been a relevant field to investigate functional integration of gene expression to transcriptomics and proteomics, as well as the emerging fields of "phenomics" and "fluxomics," contributing to the development of biological system networks, the identification of unknown gene/protein functions, abnormal gene-metabolite relationships due to knowns/unknowns gene mutations, and the analysis of metabolic pathways to explore biochemical activities [85]. Metabolite levels can reflect the closer integration of gene expression and protein synthesis, considering the influence of the environmental conditions and/or other organism's interactions as well as to control gene expression through allosteric interactions of transcriptions factors related to specific metabolites [86]. Therefore, untargeted metabolomics tries to measure all metabolites, which can be assessed by a multiplatform study (in a cell, tissue, or organism) within a specific design, reflecting a snapshot of all the physiological events as a response of gene expression and environmental conditions [85]. While metabolomics contributes to the obtention of a snapshot of the biological system under investigation, integration between transcriptomics, proteomics, and metabolomics led to visualization of a well-detailed picture or network with the respective biological mechanisms and their association to diseases from gene mutation or phenotypic alterations, constituting the so-called systems biology approach (Fig. 2).

Metabolomics plays an important role in the analysis of gene function or loss of function. It contributes with information about biological systems, as products from biochemical processes in living systems, which are influenced by abiotic (environment and stage of development) and biotic factors (transcription, mRNA degradation, posttranslational modification, protein dynamics, metabolite concentrations, and fluxes) [87, 88]. As an emerging and developing field, various concepts and definitions for the term "metabolomics" are found in the scientific literature. Drexler et al. define metabolomics as the "qualitative and quantitative (relative or absolute) analysis of the entire endogenous metabolome (metabolites with masses less than 1500 Da)" [89].

Metabolome analysis has recently been used in systems biology studies to quite comprehensively investigate the metabolic changes originated from genetic, environmental, and organism differences factors by comparing the basal levels of metabolites with those produced after alteration [90]. Autism is a genetic disease into intellectual disability (ID) diseases most commonly caused by fragile X syndrome (FXS). FXS is caused by the mutation of the X-linked fragile X mental retardation 1 (*fmr1)* gene resulting in their hypermethylation. The metabolic signature and biomarker identification associated with FXS due to *fmr1* gene inactivation results in brain metabolism alterations related to neurotransmitter levels, osmoregulation, energy metabolism, and oxidative stress response. Cardiovascular and metabolic disorders are also strongly related to gene mutations. Systems biology has been used to understand the impact of these genetic disorders in human metabolism through identification of variance in the promoter of specific genes as *fads1*, *elovl2*, *acads*, *acadm*, *acadl*, *sptlc3*, *etfdh*, and *slc16a9* that are responsible to enhanced change in the conversion rate of several metabolites associated with dyslipidemia, obesity, and diabetes. The *elovl2* and *slc16a9* genes have been associated to lipid concentration, as well as arachidonic acid, and cholesterol/triglyceride levels have been related to variant of the *fads1* gene that encodes a fatty acid desaturase enzyme [91, 92]. Therefore, information of the organism's phenotype can be explained as a consequence of the genome mutation and/or environment on the metabolome.

The two most commonly used strategies for metabolomics analysis are "targeted" and "untargeted." Targeted analyses are directed to certain classes of compounds related to known and specific metabolic pathways for their quantification by using specialized extraction protocols, separation methods, and detection techniques. On the other hand, the untargeted analysis is directed to establish the fingerprint of metabolites present in a sample, constituting a qualitative analysis of crude metabolite mixtures [89].

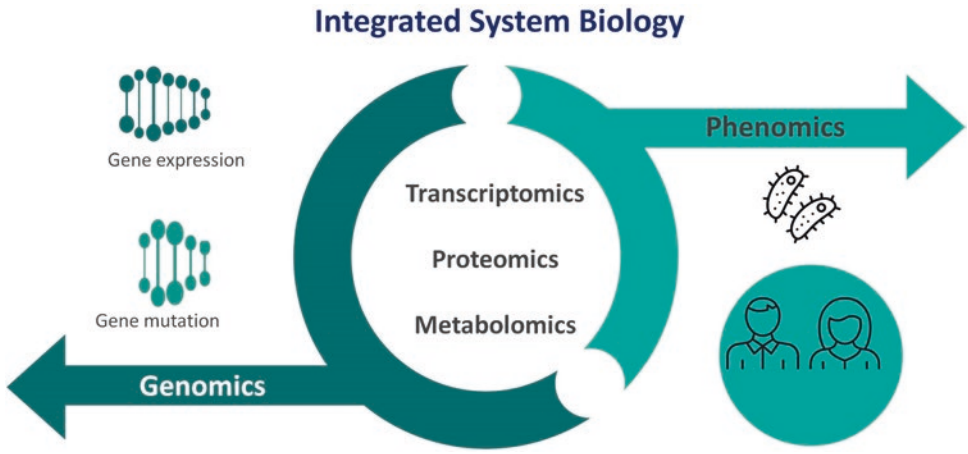

Fig. 2 Integrated systems biology from genomics to phenomics through transcriptoms, proteomics, and metabolomics sciences

A targeted approach is applied to detect few or several metabolites, if not all, according to analytical technique limitations, included in a targeted pathway. Primary metabolites including sugars, amino acids, and tricarboxylic acids that are involved in primary metabolic processes, such as respiration and photosynthesis, and secondary metabolites including alkaloids, phenolics, steroids, lignins, and tannins can be detected and quantified [93–95]. It requires optimization of the selected metabolite extraction step to reduce matrix effect and maximize the recovery and overall sensitivity for detection, generally using specific internal standards [96]. On the other hand, an untargeted approach focuses on the analysis of all metabolites (or most of them) of a biological system, not requiring a prior knowledge of the metabolites that will be altered under gene expression or mutation, causing phenotypical changes from transcriptomics, proteomics, or metabolomics alterations. Moreover, they can be detected using multivariate statistical analysis tools, such as principal component analysis (PCA), hierarchical cluster analysis (HCA), and discriminant analysis (DA) as partial least squares regression (PLS), *PLS* discriminant *analysis* (PLS-DA), Orthogonal-PLS-DA (O-PLS-DA) [97]. An untargeted analysis is usually applied prior to the targeted one to set a specific pathway to be studied, as well as the metabolites that vary under any genetic or phenotypic conditions [85, 93].

The main analytical techniques employed in recent metabolomics studies are nuclear magnetic resonance (NMR), one-dimensional and comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC-MS and GC × GC-MS, respectively), and liquid chromatography coupled to mass spectrometry (LC-MS), although other techniques such as capillary electrophoresis coupled to mass spectrometry (CE-MS) have also been employed. All these techniques and methods can be applied to the analyses of samples containing a wide variety of metabolites. However, each of them has different advantages and disadvantages considering detection and quantification limits, sensitivity, selectivity, interference, resolution, repeatability, reproducibility, and physicochemical characteristics of the compounds [89, 98]. In untargeted or targeted metabolomics, there is no single protocol capable of identifying and quantifying all possible metabolites in a single analysis, and due to a high degree of structural diversity, molecular mass, and main polarity differences between primary and secondary metabolites, there is no single protocol capable of identifying and quantifying all possible metabolites in a single analysis. Therefore, generally biphasic or triphasic extraction protocols with organic/aqueous solvents are performed to reduce sample com-

plexity, generating high polar, low polar, and protein fractions. Afterwards, sample fractions are analyzed by different separation techniques using suitable stationary phases and instrument platforms to obtain the best metabolite profile [99, 100].

Analytical instrumentations are commonly associated with metabolomics to measure numerous metabolites (from hundreds to thousands of compounds) in order to evaluate metabolic changes in response to external stimuli, such as attack of pathogens or changes of environmental conditions, and elucidate metabolic pathways of the gene expression or mutations. However, simultaneous separation and detection of all metabolites in a biological sample with appropriate analytical sensitivity and resolution in a single analysis has not been achieved yet, due to high sample complexity (numerous metabolites with different chemical classes in a wide range of concentration levels), pointing to the demand of analytical techniques improvement [101].

Ion mobility mass spectrometry (IMS) has gained popularity over the last few years due to high selectivity and resolution power for several isomeric compounds variety. Differential mobility spectrometry (DMS) has been utilized in targeted metabolomics to separate small molecules, although it has a great potential to be used in untargeted metabolomics as well [102, 103]. DMS addresses high selectivity through unique mass-to-charge and migration time combinations, allied to high accuracy. The DMS-MS analysis is typically four times faster than a typical LC-MS one, showing potential to the screening of numerous metabolites in order to address system biological tasks [103].

IMS resolution, selectivity, and accuracy improvement, mainly related to recent mass spectrometry advances, allowed detection of hundreds to thousands features in a single analysis, requiring advanced data analysis tools. Big data has become a fundamental aspect of systems biology to elucidate the complex networks by which gene expression, gene mutation, pathogen-host interaction, or environmental-host interaction is developed. The big data tools, such as machine learning and deep learning algorithms, and neural networks programmable in Python, Java, MATLAB, and R languages, allows mapping and modeling pathways to identify underappreciated gene, RNA, proteins, or metabolite functions and connections [104]. Diseases are driven by genetic and epigenetic factors and environmental factors such as attack of pathogens, leading to disturbance in immunological balance [105].

Systems biology and the main omics, approached in this book, provide new perspectives in science, rescuing the idea of how to understand biology as a system. Under molecular and cellular biology, during the last centuries, the studies were conducted by simplified methodologies, as the reductionism. Nevertheless, from the first classical sequencing techniques, the progress of the high-throughput techniques has moved forward the biology fields into the systems approach.

3 Perspectives

It is clear that the open questions in science have conducted the scientists into incredible knowledge fields. Classical science, even with the simplified methods (reductionism), has driven fields as genomics, proteomics, and metabolomics into systems biology progress. Looking back to the literature, the evolution of analytical techniques plays a key role in biology, especially in molecular biology areas (DNA/RNA, proteins, and metabolites), where the separation and characterization methods use analytical chemistry tools for comprehensive analyses of biological systems. However, as originated from classical science, these techniques allowed the progress and evolution of the theoretical and technological advances in nano-biotechnology, robotics, genetics, mathematics, and computational biology, among others, determining factors that allowed and facilitated integrative approaches, which constitute the main purpose of systems biology.

References

1. Westerhoff HV, Palsson BO (2004) The evolution of molecular biology into systems biology. Nat Biotechnol 22(10):1249–1252
2. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) Molecular cell biology, 4th edn. National Center for Biotechnology Information, Bookshelf, 9
3. D'Argenio V (2018) The high-throughput analyses era: are we ready for the data struggle? High-Throughput 7(1):8
4. Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17(11):1519–1533
5. Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58(4):586–597
6. Van Regenmortel MHV (2019) Editorial: Biological Complexity Emerges from the Ashes of Genetic Reductionism. In: HIV/AIDS: Immunochemistry, Reductionism and Vaccine Design. Springer, Cham. https://doi.org/10.1007/978-3-030-32459-9_7
7. Brigandt I, Love AC (2008) Reductionism in biology. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2008 Edition). <http://plato.stanford.edu/archives/fall2008/entries/reduction-biology>
8. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. (2000) A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. Nature 403(6770):623–627
9. Ge H, Walhout AJ, Vidal M (2003) Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet 19(10):551–560
10. Palsson B (2002) In silico biology through "omics". Nat Biotechnol 20(7):649–650
11. Meng TC, Somani S, Dhar P (2004) Modeling and simulation of biological systems with stochasticity. In Silico Biol 4(3):293–309
12. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74(12):5463–5467
13. Gilbert W, Maxam A (1973) The nucleotide sequence of the lac operator. Proc Natl Acad Sci 70(12):3581–3584
14. Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94(3):441–448
15. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR et al (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321(6071):674–679
16. Cohen A, Najarian D, Paulus A, Guttman A, Smith JA, Karger B (1988) Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis. Proc Natl Acad Sci 85(24):9660–9663
17. Swerdlow H, Gesteland R (1990) Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Res 18(6):1415–1419
18. Huang XC, Quesada MA, Mathies RA (1992) DNA sequencing using capillary array electrophoresis. Anal Chem 64(18):2149–2154
19. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al (2001) The sequence of the human genome. Science 291(5507):1304–1351
20. Frazer KA (2012) Decoding the human genome. Genome Res 22(9):1599–1601
21. Consortium IHGS (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860
22. She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E et al (2004) Shotgun sequence assembly and recent segmental duplications within the human genome. Nature 431(7011):927–930
23. Nyrén P (1987) Enzymatic method for continuous monitoring of DNA polymerase activity. Anal Biochem 167(2):235–238
24. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P (1996) Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242(1):84–89
25. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380
26. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res 34(3):e22
27. Turcatti G, Romieu A, Fedurco M, Tairi A-P (2008) A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res 36(4):e25
28. Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci 100(7):3960–3964
29. Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW et al (2009) Virtual terminator nucleotides for next-generation DNA sequencing. Nat Methods 6(8):593
30. Uemura S, Aitken CE, Korlach J, Flusberg BA, Turner SW, Puglisi JD (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. Nature 464(7291):1012–1017
31. Deamer DW, Akeson M (2000) Nanopores and nucleic acids: prospects for ultrarapid sequencing. Trends Biotechnol 18(4):147–151
32. Schneider GF, Dekker C (2012) DNA sequencing with nanopores. Nat Biotechnol 30(4):326
33. Kaiser J (2008) A plan to capture human diversity in 1000 genomes. Science 319(5868):1336

34. Mardis ER (2017) DNA sequencing technologies: 2006–2016. Nat Protoc 12(2):213
35. Woychik RP, Klebig M, Justice M, Magnuson T, Avrer E (1998) Functional genomics in the post-genome era. Mutat Res Fundam Mol Mech Mutagen 400(1–2):3–14
36. Greenbaum NL, Mundoma C, Peterman DR (2001) Probing of metal-binding domains of RNA hairpin loops by laser-induced lanthanide (III) luminescence. Biochemistry 40(4):1124–1134
37. Attur M, Dave M, Tsunoyama K, Akamatsu M, Kobori M, Miki J et al (2002) "A system biology" approach to bioinformatics and functional genomics in complex human diseases: arthritis. Curr Issues Mol Biol 4:129–146
38. Bunnik EM, Le Roch KG (2013) An introduction to functional genomics and systems biology. Adv Wound Care 2(9):490–498
39. Jansen R, Greenbaum D, Gerstein M (2002) Relating whole-genome expression data with protein-protein interactions. Genome Res 12(1):37–46
40. Shoemaker DD, Linsley PS (2002) Recent developments in DNA microarrays. Curr Opin Microbiol 5(3):334–337
41. LaBaer J, Nicolini C (2019) Functional proteomics and nanotechnology-based microarrays. CRC Press
42. Kuruvilla FG, Shamji AF, Sternson SM, Hergenrother PJ, Schreiber SL (2002) Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. Nature 416(6881):653–657
43. Peng B, Yu C, Du S, Liew SS, Mao X, Yuan P et al (2018) MSN-on-a-chip: cell-based screenings made possible on a small-molecule microarray of native natural products. Chembiochem 19(9):986–996
44. Wilkins MR, Sanchez J-C, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF et al (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 13(1):19–50
45. Avise JC (2010) Footprints of nonsentient design inside the human genome. Proc Natl Acad Sci 107(Supplement 2):8969–8976
46. Bock C, Farlik M, Sheffield NC (2016) Multi-omics of single cells: strategies and applications. Trends Biotechnol 34(8):605–608
47. Hasin Y, Seldin M, Lusis A (2017) Multi-omics approaches to disease. Genome Biol 18(1):83
48. Bensimon A, Heck AJ, Aebersold R (2012) Mass spectrometry–based proteomics and network biology. Annu Rev Biochem 81:379–405
49. Göth M, Pagel K (2017) Ion mobility–mass spectrometry as a tool to investigate protein–ligand interactions. Anal Bioanal Chem 409(18):4305–4310
50. Lamond AI, Uhlen M, Horning S, Makarov A, Robinson CV, Serrano L et al (2012) Advancing cell biology through proteomics in space and time (PROSPECTS). Mol Cell Proteomics 11(3):O112. 017731
51. Altelaar AM, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14(1):35–48
52. Varusai TM, Nguyen LK (2018) Dynamic modelling of the mTOR signalling network reveals complex emergent behaviours conferred by DEPTOR. Sci Rep 8(1):643
53. Walker JM (2005) The proteomics protocols handbook. Humana press
54. Von Hagen J (2008) Proteomics sample preparation. John Wiley & Sons
55. Andrecht S, von Hagen J (2008) The need for standards in proteomics sample preparation. In: Proteomics sample preparation. John Wiley & Sons. 5–20
56. Westermeier R, Naven T, Höpker H-R (2008) Proteomics in practice: a guide to successful experimental design. John Wiley & Sons
57. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC (2000) The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis 21(6):1104–1115
58. Willard KE, Giometti CS, Anderson NL, O'Connor TE, Anderson NG (1979) Analytical techniques for cell fractions: XXVI. A two-dimensional electrophoretic analysis of basic proteins using phosphatidyl choline/urea solubilization. Anal Biochem 100(2):289–298
59. Santoni V, Rabilloud T, Doumas P, Rouquié D, Mansion M, Kieffer S et al (1999) Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. Electrophoresis 20(4–5):705–711
60. Shevchenko A, Tomas H, Havli J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1(6):2856
61. Görg A, Drews O, Lück C, Weiland F, Weiss W (2009) 2-DE with IPGs. Electrophoresis 30(S1):S122–SS32
62. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250(10):4007–4021
63. Görg A, Obermaier C, Boguth G, Weiss W (1999) Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. Electrophoresis 20(4–5):712–717
64. Görg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R et al (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21(6):1037–1053
65. Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4(12):3665–3685
66. Øye OK, Jørgensen KM, Hjelle SM, Sulen A, Ulvang DM, Gjertsen BT (2013) Gel2DE-A software tool for correlation analysis of 2D gel electrophoresis data. BMC Bioinf 14(1):215

67. Ünlü M, Morgan ME, Minden JS (1997) Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. Electrophoresis 18(11):2071–2077
68. Larbi NB, Jefferies C (2009) 2D-DIGE: comparative proteomics of cellular signalling pathways. In: Toll-like receptors. Springer, pp 105–132
69. Koomen JM, Haura EB, Bepler G, Sutphen R, Remily-Wood ER, Benson K et al (2008) Proteomic contributions to personalized cancer care. Mol Cell Proteomics 7(10):1780–1794
70. Yates JR III (1998) Mass spectrometry and the age of the proteome. J Mass Spectrom 33(1):1–19
71. Yates JR III (2013) The revolution and evolution of shotgun proteomics for large-scale proteome analysis. J Am Chem Soc 135(5):1629–1640
72. Wolters DA, Washburn MP, Yates JR (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73(23):5683–5690
73. Larance M, Lamond AI (2015) Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol 16(5):269–280
74. Shi Y, Xiang R, Horváth C, Wilkins JA (2004) The role of liquid chromatography in proteomics. J Chromatogr A 1053(1–2):27–36
75. Essader AS, Cargile BJ, Bundy JL, Stephenson JL Jr (2005) A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. Proteomics 5(1):24–34
76. Shen Y, Zhao R, Belov ME, Conrads TP, Anderson GA, Tang K et al (2001) Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. Anal Chem 73(8):1766–1775
77. Xu P, Duong DM, Peng J (2009) Systematical optimization of reverse-phase chromatography for shotgun proteomics. J Proteome Res 8(8):3944–3950
78. Boersema PJ, Mohammed S, Heck AJ (2008) Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem 391(1):151–159
79. Lecchi P, Gupte AR, Perez RE, Stockert LV, Abramson FP (2003) Size-exclusion chromatography in multidimensional separation schemes for proteome analysis. J Biochem Biophys Methods 56(1–3):141–152
80. Horvatovich P, Hoekman B, Govorukhina N, Bischoff R (2010) Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples. J Sep Sci 33(10):1421–1437
81. Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 11:49–79
82. Walther TC, Mann M (2010) Mass spectrometry–based proteomics in cell biology. J Cell Biol 190(4):491–500
83. Kuo T-C, Tian T-F, Tseng YJ (2013) 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. BMC Syst Biol 7(1):64
84. Nam H, Chung BC, Kim Y, Lee K, Lee D (2009) Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. Bioinformatics 25(23):3151–3157
85. Kumar M, Kuzhiumparambil U, Pernice M, Jiang ZJ, Ralph PJ (2016) Metabolomics: an emerging frontier of systems biology in marine macrophytes. Algal Res 16:76–92
86. Lempp M, Farke N, Kuntz M, Freibert SA, Lill R, Link H (2019) Systematic identification of metabolites controlling gene expression in E. coli. Nat Commun 10(1):1–9
87. Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48(1–2):155–171
88. Goodacre R (2005) Metabolomics – the way forward. Metabolomics 1(1):1–2
89. Drexler DM, Reily MD, Shipkova PA (2011) Advances in mass spectrometry applied to pharmaceutical metabolomics. Anal Bioanal Chem 399(8):2645–2653
90. Davidovic L, Navratil V, Bonaccorso CM, Catania MV, Bardoni B, Dumas M-E (2011) A metabolomic and systems biology perspective on the brain of the fragile X syndrome mouse model. Genome Res 21(12):2190–2202
91. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C et al (2010) A genome-wide perspective of genetic variation in human metabolism. Nat Genet 42(2):137
92. Adamski J (2012) Genome-wide association studies with metabolomics. Genome Med 4(4):34
93. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR et al (2018) Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 14(6):17
94. Koal T, Deigner H-P (2010) Challenges in mass spectrometry based targeted metabolomics. Curr Mol Med 10(2):216–226
95. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner HP (2010) Targeted metabolomics for biomarker discovery. Angew Chem Int Ed 49(32):5426–5445
96. Kueger S, Steinhauser D, Willmitzer L, Giavalisco P (2012) High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. Plant J 70(1):39–50

97. Shalhoub J, Sikkel MB, Davies KJ, Vorkas PA, Want EJ, Davies AH (2014) Systems biology of human atherosclerosis. Vasc Endovascular Surg 48(1):5–17
98. Almstetter MF, Oefner PJ, Dettmer K (2012) Comprehensive two-dimensional gas chromatography in metabolomics. Anal Bioanal Chem 402(6):1993–2013
99. Cajka T, Smilowitz JT, Fiehn O (2017) Validating quantitative untargeted lipidomics across nine liquid chromatography-high-resolution mass spectrometry platforms. Anal Chem 89(22):12360–12368
100. Cajka T, Fiehn O (2016) Increasing lipidomic coverage by selecting optimal mobile-phase modifiers in LC-MS of blood plasma. Metabolomics 12(2):1–11
101. Gonzalez-Dominguez A, Duran-Guerrero E, Fernandez-Recamales A, Lechuga-Sancho AM, Sayago A, Schwarz M et al (2017) An overview on the importance of combining complementary analytical platforms in metabolomic research. Curr Top Med Chem 17(30):3289–3295
102. Nichols CM, Dodds JN, Rose BS, Picache JA, Morris CB, Codreanu SG et al (2018) Untargeted molecular discovery in primary metabolism: collision cross section as a molecular descriptor in ion mobility-mass spectrometry. Anal Chem 90(24):14484–14492
103. Wernisch S, Pennathur S (2019) Application of differential mobility-mass spectrometry for untargeted human plasma metabolomic analysis. Anal Bioanal Chem 411(24):6297–6308
104. Kuenzi BM, Ideker T (2020) A census of pathway maps in cancer systems biology. Nat Rev Cancer 20(4):233–246
105. Tang HH, Sly PD, Holt PG, Holt KE, Inouye M (2020) Systems biology and big data in asthma and allergy: recent discoveries and emerging challenges. Eur Respir J 55(1):1900844

Biological Applications for LC-MS-Based Proteomics

Bradley J. Smith and Daniel Martins-de-Souza

Abstract

Since its inception, liquid chromatography-mass spectrometry (LC-MS) has been continuously improved upon in many aspects, including instrument capabilities, sensitivity, and resolution. Moreover, the costs to purchase and operate mass spectrometers and liquid chromatography systems have decreased, thus increasing affordability and availability in sectors outside of academic and industrial research. Processing power has also grown immensely, cutting the time required to analyze samples, allowing more data to be feasibly processed, and allowing for standardized processing pipelines. As a result, proteomics via LC-MS has become popular in many areas of biological sciences, forging an important seat for itself in targeted and untargeted assays, pure and applied science, the laboratory, and the clinic. In this chapter, many of these applications of LC-MS-based proteomics and an outline of how they can be executed will be covered. Since the field of personalized medicine has matured alongside proteomics, it has also come to rely on various mass spectrometry methods and will be elaborated upon as well. As time goes on and mass spectrometry evolves, there is no doubt that its presence in these areas, and others, will only continue to grow.

Keywords

Proteome · Proteomics · LC-MS-based proteomics · Personalized medicine · Liquid chromatography-mass spectrometry

B. J. Smith
Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, Brazil

D. Martins-de-Souza (✉)
Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, Brazil

Instituto Nacional de Biomarcadores em Neuropsiquiatria (INBION), Conselho Nacional de Desenvolvimento Científico e Tecnológico, São Paulo, Brazil

Experimental Medicine Research Cluster (EMRC), University of Campinas, Campinas, SP, Brazil

D'Or Institute for Research and Education (IDOR), São Paulo, Brazil
e-mail: dmsouza@unicamp.br

1 Introduction

To study the many aspects of cell biology, a dizzying amount of equipment and tool sets are available or even required for different applica-

A. V. Colnaghi Simionato (ed.), *Separation Techniques Applied to Omics Sciences*, Advances in Experimental Medicine and Biology 1336, https://doi.org/10.1007/978-3-030-77252-9_2

tions. Liquid chromatography is one such tool that has many individual uses; however, for many assays, including proteomics, it is better combined with another methodology, such as tandem mass spectrometry (MS/MS). One reason for this is due to the sheer complexity of protein-containing samples [1] and the wide dynamic range of proteins in living organisms [2]. Attempting to separate proteins by chromatographic methods alone would not be sufficient for the identification and quantitation of the thousands of different proteins and their isoforms and modified states that are present in a cell or tissue.

This problem is recognized – and greatly reduced – by coupling LC with mass spectrometry (MS) or tandem mass spectrometry (MS/MS), which can perform global identifications as well as other assays like the extremely sensitive quantitation of predetermined proteins. For these applications, separation by LC also plays a vital role in improving mass spectrometry methods by reducing the complexity of the sample at any given time point, additionally improving reproducibility from online coupling [3]. Over the past few decades, advances in LC-MS technology by way of sensitivity, resolution, reproducibility, and automation, as well as reductions in the costs to purchase and operate the systems [4], have caused their role in a clinical setting to steadily increase [5–7].

In this chapter, different methods of carrying out several LC-MS- and LC-MS/MS-based proteomic assays will be briefly discussed before their various overall applications are covered, which can range from laboratorial research to personalized medicine, as well as some specific examples of their uses. Since the field is expanding quickly, some applications that are still under research or going through testing phases will also be mentioned, many of which intend to bring more applications to the clinic in personalized medicine.

2 Proteomics via LC-MS

Mass spectrometry coupled with liquid chromatography (LC-MS) has become a widely used platform for many types of research projects. With the advent of online coupling between chromatographic separation and identification by mass spectrometry, many samples can now be set up in a queue for separation and any identification with reduced experimenter-based errors. Quantitation steps and fine tuning the various possible configurations also allow for its use in a widening range of experimental designs.

One of the first mass spectrometer designs required "eyeball" identification of molecules via obtained spectra, limiting both sample complexity and number [8]. However, over 25 years ago, advancements in computer systems and informatics allowed for the genesis of computer-based identification [9] and the eventual introduction of more varied and complex samples. When proteomics was first paired with mass spectrometry, one groundbreaking method to obtain large-scale proteomic data was to run the sample in a two-dimensional gel, whereupon individual spots with lower complexity could be excised, digested, and injected into the MS system [10].

Even after such a separation or in other types of purified samples, a single injection could still contain many proteins and even more peptides after digestion. Liquid chromatography was found to be extremely useful in such cases since a sample could be fractionated before being analyzed, reducing its complexity. In an attempt to reduce experimenter error, increase reproducibility, and decrease the time necessary to complete an analysis, a liquid chromatography system was coupled online to a mass spectrometer, generating a continuous flow of data across the sample's elution time [11]. Further improving upon the LC-MS design, multidimensional liquid chromatography systems have been since incorporated and optimized, utilizing multiple columns to quickly trap

and separate peptides with reproducibility and high-resolution peak separation [12–14].

When brought together, liquid chromatography and mass spectrometry now play a key role at the center of many types of omic studies [15], including proteomics, due to its high resolution and sensitivity and the reproducibility of modern equipment. A proteomic profile changes both with cell type and with a wide range of stimuli and regulatory processes, and understanding a proteomic snapshot of a cell, as well as how the proteome changes in response to a stimulus or condition, provides unique insight into the inner workings of cells. Since there are 24 standard amino acids with dozens of possible modifications, mass spectrometers must go beyond an initial mass/charge (*m/z*) reading of a peptide, which would be insufficient to identify a peptide, and subsequently a protein, especially in a complex mixture.

In assays that require such a level of detail, the peptides that are ionized at the mass spectrometer ionization source are focused and optionally filtered before peptide fragmentation, with collision-induced dissociation (CID) being the most widely used method for proteomics [16]. The newly formed fragments (or more specifically the transitions of the precursor ions) can then be optionally filtered again before their *m/z* values are registered. Reading multiple combinations of fragments of a single peptide allows a mass spectrometer to determine, at least partially, the sequence of amino acids present, which becomes crucial when performing assays to identify which proteins are present in a sample [17].

Proteomic assays can be divided into two fundamental groups: targeted and untargeted studies. In a targeted assay, a specific list of known precursors and transitions is focused upon to detect the presence of those peptides and quantify them. When performing targeted assays, dynamic range can exceed four orders of magnitude, and sensitivity can be extremely high; sub-fmol/mg-of-sample sensitivity has already been obtained [18]. Quantitation is also exceptionally accurate in targeted studies [19], though the actual accuracy also depends on experimenter technique and the capabilities of the spectrometer. The limit of detection has also rapidly decreased, moving past the attomole level [20] into the zeptomole range [21]. Such sensitivity has also received attention for clinical applications that use extremely small sample volumes. In contrast with targeted studies, untargeted studies trade some of the aforementioned sensitivity for the ability to identify the peptides in the sample. By doing so, thousands of proteins [22, 23] and proteoforms [24–26] can be identified and quantified in a single injected sample.

Though different experiments can call for modifications to these suggestions, Fig. 1 depicts a flow chart to assist in determining which methods may be compatible with a project that plans to involve proteomics. Each of these categories will be discussed further in the following sections. It is always important to keep in mind what equipment is available for an experiment as well. Due to inherent equipment capabilities and limitations, not every LC-MS setup is able to perform both targeted and untargeted experiments [16], and not every spectrometer can perform both data-dependent and data-independent acquisition, for example.

3 Targeted Proteomics

Targeted proteomics is especially useful for the detection and quantitation of proteins/polypeptides with low abundance, the visualization of rare isoforms or posttranslational modifications (PTMs), and the validation of other identification or quantitation methods. A targeted assay has two main steps: a selection step and an acquisition step. In the selection step, theoretical mass data or data acquired from discovery studies is used to create a method for the spectrometer that determines what precursor ions will be filtered at the MS1 level and optionally fragmented at the MS2 level. In MS2 methods, the resulting transitions (peptide fragments) can also be preselected so they can reach the mass analyzer for identification and quantitation, referred to as single reaction monitoring (SRM) or multiple reaction monitoring (MRM). If all transitions are allowed to reach the mass analyzer for a given precursor, this method is called parallel reaction monitoring (PRM) and requires a high-resolution mass