REVIEW



The Importance and Challenges for Analytical Chemistry in Proteomics Analysis

Weliton Pedro Batiston^{1,2} , Emanuel Carrilho^{1,2*}

¹Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense, 400, 13566-590 São Carlos, SP, Brazil

²Instituto Nacional de Ciência e Tecnologia de Bioanalítica – INCTBio, 13083-970, Campinas, SP, Brazil



Although Linus Pauling had an exceptional scientific contribution to the study of chemical bonds, reported in his book *The Nature of Chemical Bond*, the lousy image he got for the X-ray diffraction drove him to an unstable structure with an unreal DNA triple helix publication. Oppositely, for the consecration of James Watson & Francis Crick, they had the opportunity to enter science history using the right image of X-ray to propose the famous DNA double

helix structure correctly. This chapter of science is an excellent example of how analytical chemistry performance affects horizons and scientific advances. Today the complexity of the systems is more significant and understanding how all proteins truly work into cells and organisms is the current challenge from proteomics. Comprehending how analysis is carried out and how instruments work could promote new insights to improve the analytical performance in proteomics. Here we described an overview based on our expertise on the analytical chemistry toolkit for proteomics analysis: *shotgun*, *bottom-up*, *middle-down*, *top-down*, and *native* proteomics, and their inherent instrumentation technologies. In addition, a detailed discussion of the analytical figures of merit in proteomics analysis is provided. We also address the limitations in multidimensional liquid chromatography and tandem mass spectrometry platforms. Furthermore, we present some perspectives in bioinformatics, mathematical modeling simulations, and chemometrics tools, as well.

Keywords: instrumentation, liquid chromatography, mass spectrometry, bioanalysis, biopharmaceuticals.

INTRODUCTION

The importance of studies on the proteins: implications in medicine and economics

The advances in proteomics have many implications for the economic and scientific fields because *i*) the modern pharmaceutical industry is developing biopharmaceutical, most of them based on proteins; *ii*) the current bet to the next advances in medicine is molecular diagnosis and the therapy linking proteome anomalies to a more in-depth characterization of diseases; *iii*) the final products from

Cite: Batiston, W. P.; Carrilho, E. The Importance and Challenges for Analytical Chemistry in Proteomics Analysis. *Braz. J. Anal. Chem.*, 2021, 8 (31) pp 51-73. doi: http://dx.doi.org/10.30744/brjac.2179-3425.RV-64-2020

Received 2 November 2020, Revised 25 March 2021, Accepted 6 April 2021.

the machinery of living beings are proteins. Therefore, the future will be even more exciting to molecular information from proteomics than currently from genomics [1,2].

In any of these examples, the chemical elucidation of proteomes and protein is critical to advancing the subject. However, when macromolecules are introduced in to an analytical equipment for characterization, the complexity of its chemical structure imposes various drawbacks that affect several analytical figures of merit. Giant molecules have many possibilities to change conformation due to the free rotation of any *sigma* bond. They show a complex nature driving to several intra- and intermolecular interactions possibilities and many structural conformations. For that, the next generation of analytical scientists needs to overcome the challenges of analyzing complex macromolecular samples that mean understanding how to drive the proteomics analysis and the instrumentation modes to its maximum. Although the theme is old and pertinent, the discussion in the proteomics community needs more insights [1,3,4].

Proteins provide essentials functions in all biological steps. They are catalyst agents in biochemistry reactions, carry and store other molecules, provide signal transduction, they are essential constituents in the structure of various biologics organisms, etc. Thus, reliably unraveling the biological process of proteins in the cell is the path to the solution and understanding of various diseases, suggesting advances in chemistry, pharmacy, biology, biochemistry, medicine, and economic development [1,2].

In practice, scientists in molecular medicine compare healthy samples with samples containing the questioned disease. Using analytical and bioinformatics techniques, they monitor the appearance of analogous proteins in the cell or monitor the variation in their quantity, or even map changes in the threedimensional structure that can represent the origin of the disease [3]. Some proteins have the function of controlling the normal condition of cells. Characterization of them and understanding the amount suitable to an individual's necessities opens the opportunity for creation of supplements or drugs that promote the normal functionality of cells, preventing the proliferation of disease into an organism. They are the basis of biologicals and biosimilar medicines, most of them made up of monoclonal antibodies. These biopharmaceuticals have revolutionized the treatment of many previously untreatable diseases. Making it one of the top products of the modern pharmaceutical industry [4].

In the 90s, BASF Pharma, in partnership with Cambridge Antibody Technology, developed the incorporated drug of the protein Adalimumab whose market name is Humira[®]. It is intended for the antiinflammatory treatment of several chronic diseases, such as *i*) arthritis, *ii*) ankylosing spondylitis, *iii*) Crohn's disease, *iv*) colitis, and *v*) psoriasis. The Institute for Applied Economic Research (IPEA) mentioned that this product generates US\$ 10 billion a year of revenue to the German chemical company [5]. Several Brazilian pharmaceutical companies have started the production of biosimilars in the last decade. In Brazil, despite the economic crisis, this sector presented annual growth of 10%. The current increase in the aging of the population in emerging countries drives the need for constant innovation, but the expiration of patents for biosimilar medicines drives the plans of the pharmaceuticals companies for expanding the market.

State of the art: multidimensional proteomics for cell biology

A proteome concerns the set of all proteins expressed in a cell, tissue, or organism, from the functional expression of a specific genome. Thus, the term proteomics encompasses strategies for the identification and quantification of all proteins from a specific proteome. The proteome is a dynamic system in which each protein has interconnections in the cell that can have many different properties (chemical, physical, and biological dimensions). When analyzed altogether, it describes the phenotype of the cell/organism. Although it is possible to obtain relevant information from proteomes, the diversities, and the dynamic nature of living systems infer several drawbacks to measure these properties, meaning that is a great challenge to analytical chemistry and bioinformatics fields to overcome them. Also, there are many different proteomic profiles for the same biological system that depends on proteins expressed at the moment of sample collection. This fact represents a true challenge to get a representative sample and to control the statistical variable in terms of random errors [6].

Thus, bioanalytical chemistry and bioinformatics seek strategies for a systemic understanding of the proteome, which includes different experiments to elucidate the different properties of proteins, such as *i*) total abundance, *ii*) expression of isoforms, *iii*) turnover number, *iv*) post-translational modifications, and *vi*) molecular interactions. The connection of the results on the different properties of proteins and proper biochemical interpretation is currently called multidimensional proteomics. That allows a global view of the proteome, suggesting several scientific advances, especially in drug development and understanding of currently incurable diseases. In Table I, Professors Mark Larance and Angus I. Lamond of the University of Dundee, described the main experiments used to measure the different properties (dimensions) of the proteome [6].

Techniques applied to proteomics

Proteomics experiments can be classified in two ways: *i*) analysis of peptides from final products of proteolytic digestion from specific enzymes, named bottom-up, shotgun, and middle-down proteomics, or *ii*) analysis of intact proteins, named top-down and native mass spectrometry. Analysis of peptides allows greater identification of species, and it is currently the most suitable technique for comprehensive characterization of a proteome and comparison between sample treatments. Analysis of intact proteins allows better structural characterization of macromolecules and, generally, they come from a less complex mixture [7–12]. The readers must have in mind that if there are various ways of analyzing a proteome, a single technique cannot solve alone completely all questions to understand them. Each technique is complementary to one another.

Dimension	Examples of techniques used*					
	Label-free quantitation					
	SILAC					
	¹⁵ N-labelling					
Abundance (absolute and relative)	NeuCode SILAC					
	Dimethyl-labelling					
	TMT					
	iTRAQ					
	Centrifugal elutriation					
Cell cycle regulation	Chemical inhibitors of cell cycle regulators					
	FACS (for DNA content or phase-specific markers)					
Tissue distribution	Dissection					
Tissue distribution	FACS (for cell-type-specific markers)					
	Affinity-enrichment (endogenous immune-precipitation or tagged fusion protein pull-down)					
Interactions	Protein correlation profiling					
	Proximity-labelling					
	Affinity enrichment: TiO ₂					
	Affinity enrichment: IMAC					
Post-translational	Modification-specific antibodies					
modifications	Liquid chromatography: IEX					
	Liquid chromatography: HILIC					
	Liquid chromatography: ERLIC					

Table I. Some of the main dimensions and complexities of a proteome. There are, however, other dimensions not described in this table. [Reprinted (adapted) with permission from [6] Larance, M.; Lamond, A. I. Nat. Rev. Mol. Cell Biol., 2015, 16 (5), pp 269-280 (https://doi.org/10.1038/nrm3970). Copyright© (2021), Springer Nature.]

Table I. Some of the main dimensions and complexities of a proteome. There are, however, other dimensions not described in this table. [Reprinted (adapted) with permission from [6] Larance, M.; Lamond, A. I. Nat. Rev. Mol. Cell Biol., 2015, 16 (5), pp 269-280 (https://doi.org/10.1038/nrm3970). Copyright© (2021), Springer Nature.] (Cont.)

Dimension	Examples of techniques used*
	Centrifugation
Localization	Protein correlation profiling
Localization	Proximity-labelling
	Detergent solubility
Turpovor	Metabolic pulse-labeling
Turnover	Cycloheximide treatment
Isoform expression	High sequence coverage to identify isoform-specific peptides Targeted mass spectrometry analysis may be used to detect isoform-specific peptides
Solubility	Thermal denaturation followed by differential centrifugation
Activity	Analog-sensitive kinases
Activity	Activity-dependent binding domains
Tortion Structure	Protease sensitivity
	Crosslinking

*ERLIC, electrostatic repulsion hydrophilic interaction chromatography; FACS, fluorescence- associated cell sorting; HILIC, hydrophilic interaction chromatography; IEX, ion-exchange chromatography; IMAC, immobilized metal affinity chromatography; iTRAQ, isobaric tags for relative and absolute quantification; LC-MS/MS, liquid chromatography followed by tandem mass spectrometry; SILAC, stable isotope labeling by amino acids in cell culture; TiO₂, titanium dioxide; TMT, tandem mass tag.

Bottom-up proteomics refers to the analysis of proteolytic digests made under different chromatography fractions of the whole proteome of interest. Shotgun proteomics is the proteolytic digestion performed under the entire proteome (without previous fractionation of the sample). Both bottom-up and shotgun seek complete cleavage of peptides bonds, therefore, products with low molecular weight. Analysis of small molecules such as peptides simplifies sample handling, separation, and identification performance in the analytical instruments. However, after analyzing the results (a deluge of peptides), it is imperative to use bioinformatics that provides the sequencing of the peptides and proposes the proteins identity. This subject refers to protein coverage; the ideal condition would be the sequencing of all peptides from all proteins, meaning 100% coverage of each protein from proteome. However, this fact has proved to be very challenging—nearly impossible—because of the limitations from shotgun proteomics, where most peptides are lost or not detected in the analytical method [13,14]. Thus, the software uses other metrics to indicate the accuracy of relative protein identification. Besides the coverage of the proteins, other issues are the *score*, the *number of peptides*, the *number of unique peptides*, and the *number of proteins* [15–17].

Top-down proteomics is the study of proteoforms, *i.e.*, the analysis of intact proteins, usually molecular masses of 30 kDa. Top-down proteomics tolerates structural changes arising from the conditions of the method or sample preparation. In native protein mass spectrometry, another class of intact protein analysis, non-covalent interactions (molecular interactions) between proteins, are kept. Thus, the technique uses analytical method conditions close to the biological middle of origin. Therefore, it is useful to study functional and structural dynamism from interactions that occur into proteins complexes, *i.e.*, in the quaternary state of proteins (protein-protein and protein-ligand complexes). Intact protein analysis shows many challenges in sampling, separation, and detection processes than peptides analysis [10,11].

Middle-down proteomics is the analysis of peptides of larger molecular weight than shotgun and bottomup (> 5 kDa) or limited peptide cleavage. This approach allows studies on the relative characterization of proteoforms with a good number of biomolecular species for identification. The technique is the middle ground between bottom-up/shotgun and top-down/native protein mass spectrometry. Figure 1 summarizes the different techniques applied to proteomics analysis [7].



Figure 1. Different types of proteomics analysis and strategies. Left (blue) and center (white): Peptide analysis (shotgun, bottom-up, and middle-down proteomics). Right (green): Protein analysis (native and top-down proteomics). In bottom-up proteomics it is necessary the chromatography fractionation before proteolytic digestion, in shotgun proteomics is not required, but after proteolytic digestion liquid chromatography is always used. In top-down proteomics is necessary a mild sample treatment, in native is not substantially required. Mass spectrometry and bioinformatics for peptides analysis is essentially to sequence peptides into proteins. Mass spectrometry and bioinformatics for proteins analysis is essential for characterization of proteoforms. Peptides analysis allows better characterization of the proteome, whereas protein analysis allows better characterization of proteoforms.

State of the art in analytical instrumentation for large scale identification for proteomics

Proteomic analysis is complex, time-consuming, and expensive. To solve the proteome's biomolecular constituents before detection and identification is essential to use high separation power techniques, preferably automated and reproducible. Currently, multidimensional liquid chromatography (LC×LC) has the most appropriate attributes for that. Also, a molecular detection technique for accurate characterization is essential. Thus, mass spectrometry is the crucial tool because of its versatility. It allows detection in MS^1 and characterization after fragmentation in MS^2 quickly and, most importantly, at the molecular ion level. Besides, this technique provides an incredible separation power, separating ions by a difference as low as 0.0001 *m*/*z* in high resolution instruments. Such versatility of MS adds to the separation power of the LC×LC. Therefore, in terms of high orthogonality of separation, the state of the art in proteomic studies is multidimensional liquid chromatography coupled to high-resolution tandem mass spectrometry [18–22].

Multidimensional or two-dimensional liquid chromatography (LC-LC and LC×LC)

Multidimensional chromatography is designed for techniques that combine two or more chromatographic columns, so for each column, an analogy of a separation dimension is made. Its applicability becomes necessary for samples with high complexity, usually > 400 compounds. Figure 2 summarizes when two-dimensional chromatography (2D-LC) becomes necessary. It is historically verified that the isocratic 1D mode allows the separation of up to 10 compounds in an average of 10 minutes of analysis indicated by blue area in Figure 2. In contrast, for samples with approximately 100 compounds, it is essential to use a gradient, as indicated by yellow area. For samples above 1000 compounds, it is necessary to use 2D LC, green area.



Figure 2. Representation of the evolution of different liquid chromatography strategies on the complexity of a sample. The different colors indicate the historical domain of the different modes of separation. [Reprinted (adapted) with permission from [22] Stoll, D. R.; LI, X.; Wang, X.; Carr, P. W.; Porter, S. E. G.; Rutan, S. C. J. Chromatogr. A, 2007, 1168 (1-2), pp 3-43 (https://doi.org/10.1016/j.chroma.2007.08.054). Copyright© (2021), Elsevier.]

Peak capacity definition means the maximum number of components that can theoretically have separation in the chromatographic column for a specific time interval of the gradient. The main advantage of 2D-LC is the high peak capacity (n), as indicated by Figure 3, where two peaks in 1D are separated into four peaks in 2D. This phenomenon occurs due to the differences in selectivity in the solute partition process between the stationary phases of the different dimensions [22–24].



Figure 3. Scheme of separation from two chromatographic columns (bidimensional or multidimensional) illustrating peak capacity. [Reprinted (adapted) with permission from [22] Stoll, D. R.; LI, X.; Wang, X.; Carr, P. W.; Porter, S. E. G.; Rutan, S. C. J. Chromatogr. A, 2007, 1168 (1-2), pp 3-43 (https://doi.org/10.1016/j.chroma.2007.08.054). Copyright© (2021), Elsevier.]

J.W. Dolan et al. discuss different ways of measuring peak capacity, where the purpose is to accurately measure all the factors inherent to the random errors of chromatographic analysis. However, Equation 1 and Equation 2 are the most used due to their simplicity, where, *n* is peak capacity, t_R and t_1 are the

retention time of the last and the first chromatographic peak, respectively, W_b is the average of the base width of all peaks, and t_a is the gradient time [24,25].

$$n = \frac{(t_R - t_1)}{W_b}$$
Equation 1
$$n = 1 + \frac{t_g}{W_b}$$
Equation 2

The expectation is to obtain high values of *n*, so the LC strategies to achieve that are: *i*) decrease the rate of the organic phase gradient, *ii*) use columns of greater length, with a proportional increase in the gradient time, and a consequent increase in pressure of the system, and *iii*) employ columns packed with smaller sorbent particles. Achieving better performance in multidimensional chromatography means that the separations in the different columns have orthogonality. It is expected that the physical-chemical mechanisms in the analyte partition process for each chromatographic dimension are independent; in other words, they provide different selectivities [24–26].

Figure 4 defines the theoretical meaning of orthogonality. We can quickly notice the origin of the word orthogonality, which comes from trigonometric concepts. That is, each chromatogram is projected at an angle of 90°. Thus, when the analytes are randomly distributed in the chromatographic area, they have little correlation and high orthogonality. Therefore, excellent peak capacity. If the analytes do not have randomness and high correlation, there is low orthogonality and inefficiency in peak capacity [24].



Figure 4. Theoretical representation of orthogonality in multidimensional liquid chromatography. A) low correlation and high orthogonality. B) high correlation and low orthogonality. [Reprinted (adapted) with permission from [24] François, I.; Sandra, K.; Sandra, P. Anal. Chim. Acta, 2009, 641 (1-2), pp 14-31 (https://doi.org/10.1016/j.aca.2009.03.041). Copyright© (2021), Elsevier.]

In 1995, J.C. Giddings proposed that the theoretical peak capacity of multidimensional chromatography is the multiple products of the peak capacities of each separation dimension. Thus, when it is combining the orthogonal LC modes of the peak capacity of $n_1 = 60$ and $n_2 = 100$, the result is an n_c of 6000, according to Equation 3 [27].

$$n_c = n_1 \times n_2 \times n_3 \times \dots n_n$$
 Equation 3

Where n_c is the total peak capacity, and n_1 , n_2 , and n_n are the peak capacities of each chromatographic dimension. Therefore, the peak capacity depends on the orthogonality of the separation dimensions. If the selectivity of the separation modes is not entirely orthogonal (different), the maximum attainable peak capacity is less than expected.

Setup in multidimensional liquid chromatography

The technique can have two types of configurations: off-line or online. The off-line setup is based on collecting fractions from 1D and the injection of fractions into 2D. The online mode performs the direct transfer of the eluent fractions from the first dimension to the second dimension, without interruption in the flow of the mobile phase, usually using switching valves, that can be named column-switching (Figure 5A and B) [23,28].

The off-line setup offers greater instrumental simplicity and flexibility when the mobile phases between the dimensions are different since the fractions can be: diluted, concentrated, or dissolved in appropriate solvents. The disadvantages are a higher risk of loss and contamination of the sample due to manipulation and difficulty in automating the system, resulting in longer analysis times [19,28].

In online mode, sample handling and analysis time are reduced, which helps avoiding dead volumes, sample losses, and analyst errors. However, this technique has a strict requirement in the choice of solvents for the mobile phase because they must be compatible in both dimensions; otherwise, there is a need for a trap column to clean the fractions and change the eluent before the analytes are transferred to the second column. Consequently, it infers increasing in the time of analysis [23,28].

The proteomics community has divided online approaches into column-switching and multidimensional protein identification technology (MudPIT) systems (Figure 5C). MudPIT is a multidimensional chromatography setup proposed by Yates Group that uses a strong cationic exchange column (SCX) in 1D coupled directly to reversed-phase in 2D. Thus, the SCX column and the RP column are combined in one fused silica capillary with a glass needle directly connected to the mass spectrometer and act as an electrospray emitter due to voltage application in the bottom of the column (Figure 5C). For that, the sample is trapped first in the SCX phase. Different fractions of peptides are eluted from the SCX onto the RP phase by increasing salt steps. After the salt step, the peptides are eluted from the RP phase into the mass spectrometer using an acetonitrile gradient (Figure 5C and D). MudPIT has the limitation to clogged easily into glass needle tip from column interface to MS [29].



Figure 5. Different instrumental modes for two-dimensional liquid chromatography separations using SCX and RP as stationary phases. A) Off-line setup the sample is first separated by SCX, and fractions are collected by the handle or automatized. B) Online column switching setup. The sample is first loaded onto the SCX column and eluted stepwise onto the trap column. The sample is desalted and subsequently eluted onto the 2D, the analytical RP column followed by MS/MS analysis. C) MudPIT approach with SCX and RP stationary phases packed in one capillary also has the function of an electrospray tip for direct MS/MS analysis. In the triphasic setup, a different RP phase is packed before the SCX and functions as a trap for desalting the sample before SCX–RP-MS/MS. D) Typical multidimensional separation of a MudPIT-MS/MS analysis. Each color indicates the RP separation after one salt step. [Reprinted with permission from [29] Palma, S. D.; Hennrich, M. L.; Heck, A. J. R.; Mohammed, S. J. Proteomics, 2012, 75, pp 3791-3813 (https://doi.org/10.1016/j.jprot.2012.04.033). Copyright© (2021), Elsevier.]

The abbreviation LC-LC means that only relevant and properly chosen parts of the separation made in 1D are injected into 2D, and the rest of the eluate proceeds for disposal, which is linked to interferents or compounds that contribute to the matrix effect. Thus, LC-LC is often used when a greater resolution is required to examine a small segment of a particular peak from 1D, from a complex sample. In this case, the nomenclature of the technique is heart-cut, where only part(s) of interest(s) of separation in 1D is directed to 2D [23,28].

Another technique is named comprehensive multidimensional liquid chromatography. Here all fractions separated in 1D are submitted to 2D producing the analysis of all compounds in the sample, and consequently, all are of interest to the analyst. Generally, it has been used to mixtures with more than a thousand compounds where the goal is to characterize the sample completely. Thus, it has been used for studies on omics sciences, such as *i*) proteomics, *ii*) metabolomics, *iii*) lipidomics, *etc*.

The difficulties in the development of this technique are related to *i*) the compatibility of the mobile phases of the first dimension with the second dimension, *ii*) the resolution obtained in 1D should not be lost in the elution process for 2D, and *iii*) the size connections and columns should be planned to suitable elution, sensibility, separation in the 2D, mainly in online mode, due to the amount of fraction eluted be direct from 1D to 2D [23,24,26].

In terms of nomenclature/abbreviation when it refers to the heart-cut modality, the abbreviation is LC-LC. When it comes to the comprehensive modality, the abbreviation is LC×LC. The multiplication symbol means the products multiple of the peak capacity between 1D and 2D, which represents the separation area that has been usually represented as a rectangle [23,26,28].

Currently, several works have been related in the literature to LC×LC and LC-LC modes for proteomics, with most of them based on SCX columns as the first dimension and RP column as the second dimension. Other growing strategies include RP columns in both the dimensions, where the mobile phase in the first dimension has high pH and the mobile phase in the second dimension has low pH. A third approach refers to the use of HILIC (hydrophilic interaction chromatography) columns as the first dimension and RP column as the second chromatographic dimension [30,31].

The high orthogonality of these methods: *i*) SCX×RP or SCX-RP; *ii*) RP×RP or RP-RP, and *iii*) HILIC×RP or HILIC-RP comes from different selectivity that each dimension can offer over partition phenomena. For example, the method *i*) have electrostatic interaction *vs* hydrophobic interaction or nonpolar interaction; in *ii*) conformational structure of biomolecules change in different pH between chromatographic dimensions generating different retention times, and *iii*) have polar *vs* nonpolar interactions, a full complementary and independent physical-chemistry mechanism. Some influential researchers in the field have reviewed substantially 2D LC modes, Table II summarizes some applications and interfaces used in 2D LC [30,31].

Sample	Separation	Application	Hyphenation	Detection	Ref.		
Peptides	Affinity×RP	Glycated proteins	Online (RP trap)	Orbitrap-MS	[32]		
Peptides	SCX×RP	Proteins alterations in colon cancer	Offline	MS	[33]		
Peptides	RP×RP	Proteome analysis of the methane producing archaeon <i>Methanosarcina mazei</i> strain Go1 cytosolic	Offline	Q Exactive	[34]		
Peptides	RP×RP	Glycopeptides from Monoclonal Antibody	Offline	Orbitrap	[35]		
Peptides	RP×RP HILIC×RP SCX×RP Other 13 LC×LC methods	Yeast cell lysate	Offline	UV and Triple TOF	[36]		

Table II. Summary of some applications of 2D-LC to the proteomics analysis

Sample	Separation	Application	Hyphenation	Detection	Ref.
Peptides	SCX×RP RP×RP	Model of the brain of monkeys to study the Onlir characterization of molecular mechanisms of induced cerebral ischemia (SC) trap)		TripleTOF	[37]
Peptides	SCX×RP	Proteins in the jejunum tissues of enterotoxigenic <i>Escherichia coli</i> -infected piglets	Online	Q Exactive	[38]
Peptides	SCX×RP RP×RP	Characterization and quantitative proteomics mapping in cerebral infarcts	Online (SCX and RP trap)	Q-TOF and tripleTOF	[39]
Peptides	SCX×RP	Proteome analysis of S. cerevisiae	Online (RP trap)	FTICR-MS	[40]
Peptides	SCX × RP RP × RP HILIC × RP	Characterization of therapeutic monoclonal antibodies	Online (valve)	UV and QTOF- MS	[41]
Peptides	RP×RP	Detection and quantification of host-cell proteins in biotherapeutic drug substance	Online (valve)	lon mobility and QTOF-MS	[42]
Protein	RP×RP	<i>E. coli</i> cell lysate	Online (RP trap)	LTQ-Orbitrap	[43]
Protein	SEC-RP	Endogenous membrane proteins	Offline	ETD Q-TOF	[44]
Protein	Affinity×RP	Analysis of glycated albumin	Online (RP trap column)	LTQ-Orbitrap	[45]
Protein	SEC-RP	Determination of degradants and small molecules in antibody-drug conjugate (ADC) solutions	Online (loop trap)	UV and MS	[46]
Protein	Mixed mode – RP; IEX-RP	Characterization of polysorbate surfactants in monoclonal antibody drug formulation	Online (loop trap)	CAD and MS	[47]
Protein	SAX × RP	Quantification of Gly m 4 protein in soybean	Online (RP trap)	UV and MS	[48]
Protein	SEC – RP(µchip)	Characterization of monoclonal antibody degradation products	Offline	Orbitrap-MS	[49]
Protein	ein SEC – RP Characterization of monoclonal antibody variants		Online (RP trap and valves)	QTOF-MS	[50]

Table II. Summary of some applications of 2D-LC to the proteomics analysis (Continuation)

An overview over Table II infers that LC-LC-MS/MS method has been more used for monoclonal antibody or biomarker characterization, whereas LC×LC-MS/MS have been extensively applied for proteomes characterization. However, it does not seem to be a general rule for application choice, that can be inferred depending on the subject of study of the scientist. Besides, HILIC×RP has more challenges in the hyphenation because the mobile phase between dimensions is strictly incompatible. It is mean that loss of chromatography resolution between dimensions is easily achieved. For that, it has been used offline mode or sophisticated configuration using trap column, valves, or loops.

Tandem mass spectrometry

According to the theories of electromagnetism, charged particles (ions) in motion (accelerated) generate an electromagnetic field, the action of an external field, the use of radiofrequency (RF), and Direct Current (DC) or Alternating Current (AC) over the ions, allows their control. Consequently, particles subjected to acceleration and deflection (force) from the external electromagnetic field enable the physical phenomenon to measure the *m*/*z* of the ions. That means the type of force applied and adequate equation obtain a solution in *m*/*z*. Figure 6 shows an introductory block diagram for a typical tandem mass spectrometry instrument for proteomics analysis. The hardware of the mass spectrometers have been changed continuously for the question of improvement of the instrument and probably by the strategy of the market of the companies.



Figure 6. Block diagram for tandem mass spectrometer in proteomics analysis. After multidimensional liquid chromatography (LC×LC), the sample is injected into the ion source of mass spectrometry by an electrospray ion source. The Mass Analyzer I filter/separate more intense peaks from a mass range previously selected. Collision cells activate the molecules for their fragmentation. The Mass Analyzer II filter/separate more intense peaks from precursor ions fragmented. After detection, the data is processed as several Analytical Information for protein characterization, where MS¹ is the masses from Analyzer I, and MS² are the masses from fragments of the precursor ion.

Proteomics has driven the advance of mass spectrometry analysis and instrumentation. The mass accuracy combined with the resolution are the most important figures of merit that infer greater detection of species due to the discrimination between masses (m) for values < 1 ppm (10^{-6} m/ Δ m for suitable IUPAC definition). Instrumentation have undergone considerable advances in sensitivity, speed of acquisition, and quality to biomolecule fragmentation in recent years. The researchers have focused on developing *i*) sources of micro and nanoflow ionization to increase the sensibility, *ii*) innovations in mass analyzer shapes to increase resolution, *iii*) including their different hybrid configurations. That means using two mass analyzers in sequence; *iv*) optical devices to control the focus or deflection of the ion beams in the vacuum, using a suitable electromagnetic field to increase the sensibility, and *v*) new peptide and protein fragmentation techniques to improve identification [11,19].

The separation properties to ions submitted by an electromagnetic field and, differentiated by their mass charge ratio (m/z), make mass spectrometry a high-speed technique compared to chromatography since separation is governed by physicochemical separation of partition. However, chromatography remains the most efficient and cheapest technique in terms of separation power for the solution of complex samples for > 1000 analytes [18,26].

In general, proteins and peptides are polar, non-volatile, and thermally unstable. Because of these properties, one of the most significant scientific advances in the area was the introduction of soft ionization methods, without the random and abrupt degradation of the biomolecule, such as electrospray ionization (ESI) [51], matrix-assisted laser desorption ionization (MALDI) [52] and, recently, ionization by desorption and electrospray (DESI) [53]. In contrast to MALDI and DESI, the ESI ionization source produces ions in solution, an aspect that makes it compatible with liquid chromatography. In terms of sensitivity for mass spectrometry, a substantial gain was creating nanoelectrospray (nano-ESI) [54].

The nano-ESI do not use gases (N_2 or He) for sample desolvation because it is a low flow rate technique (less than 20 nL min⁻¹). This low flow allows to shape of microdroplets with a larger surface area and a smaller drop radius, contributing to the Coulombic explosion and consequent increase of sensitivity due to the lower dilution of the analytes in the mobile phase.

The usual nanoflow of the coupling of the LC to nano-ESI is 250 nL min⁻¹. There are two factors that essentially need to work in together: *i*) chromatographic efficiency according to the parameters of the *van Deemter* equation in the nano chromatographic column (75 µm i.d.), and *ii*) efficient aerosol formation of the biomolecules into the ionized state (ionization), to reach the vacuum chamber of the mass spectrometer. It has been shown that in the ESI process non-covalent bonds (biologists' nomenclature) or molecular interactions (chemists' nomenclature) between biomolecules are kept. The versatility of the technique allows the ionization and infusion of protein complexes in the mass spectrometer. Thus, it is possible to analyze the quaternary structure of small proteins. That infers studies on the protein conformations in their native shapes [10,11].

In the mass analyzer, the physical phenomenon to measure the m/z of the ions occurs (Figure 6). Currently, the analyzers used in proteomics or other complex bioanalytical demands focus on couple hybrid systems between *i*) quadrupole (Q), *ii*) time of flight (TOF), *iii*) linear ion trap (LIT), *iv*) Orbitrap, and *v*) Fourier Transform Ion Cyclotron Resonance (FTICR). Hybrid systems in mass spectrometry obtain better analytical figures of merits for detecting ions from a complex mixture. The hybrid configuration increases the possibilities to control, filter, and separate the ions; it allows fragmentation of molecules, which increases information for the characterization of unknown chemical species.

Among all mass analyzers, FTICR offers the highest resolution. However, the Orbitrap analyzer has also a high resolution and delivers best results in sequential mass spectra (MS/MS), with lower maintenance costs. Orbitrap and FTICR use the Fourier transform signal processing to convert the time domain signal from the frequency of the ionic current into m/z spectra [8,11,19].

The hybrid configuration of the ion trap coupled to Orbitrap has been the tool of choice for biomolecular studies from samples of high complexity, mainly the studies driven by shotgun and bottom-up proteomics. The ion trap is considered a mass analyzer of high yield in proteomics. It has significant versatility because *i*) has high capacity to store ions, *ii*) provides enrichment of sensitivity, *iii*) shows excellent range of mass selection, *iv*) scans at high speed, *v*) features efficient service cycle (high-duty cycle), *vi*) moderate mass resolution (full width at half height - FWHH - 2000); *vii*) mass accuracy around 100 ppm (1×10⁻⁴ m/ Δ m), *viiii*) is robust, and *ix*) has simple architecture, what favors miniaturization [8,55].

The linear trap quadrupole (LTQ) is made of four hyperbolic rods positioned symmetrically on a 4 mm radius between them. Each rod is divided into three long axial sections 12, 37, and 12 mm (Figure 7A). In each section has a discrete DC level. The geometry was first simulated by the Simion technique, predicting the possibility of confining ions along the axis of the central section of the device, avoiding possible distortions of the electromagnetic field. The ability of this configuration to avoid distortion of the field is shown in the simulation of Figure 7B [55].

Although the routine operation of the LTQ requires further details in the explanation, the physical phenomenon to confinement of ion occurs due to the potential rotation applied in the central section, together with the application of discrete DC in each segment of the analyzer, as described by Figure 7C. Thus, when the positive ions are attracted to the opposing poles, in a short time, the negative poles are converted into positive poles causing the repulsion of the positive ions. The temporal and spatial change from positive to negative potential generates an oscillating, reproducible, and symmetrical potential surface (Figure 7D), causing the capture or storage of ions in vacuum.



Figure 7. Summary of the operation of the linear trap quadrupole mass analyzer. A) The basic characteristic of the analyzer; B) Representation of the Simion simulation of the excitation resonance of the electromagnetic field without distortion; C) The radiofrequency (RF) applied varies depending on the time and potential. In a symmetrical and synchronized way, the positive potential at a given moment is converted into a negative potential at a second moment. The final display resembles a circular signal/potential rotation. [Reprinted (adapted) with permission from reference [55] Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. J. Am. Soc. Mass Spectrom., 2002, 13, pp 659-669 (https://doi.org/10.1016/S1044-0305(02)00384-7). Copyright© (2021) American Chemical Society. Further permission related to the material excerpted should be directed to the ACS]. D) An instantaneous moment of the field generated by the quadrupole or surface of the potential, as a function of an applied RF. The orange circle represents an ion on the surface of the potential. [Reprinted (adapted) with permission from reference [56] March, R. E. Encyclopedia of Analytical Chemistry, 2012, p 4 (https://doi.org/10.1002/9780470027318.a6015.pub2). Copyright© (2021) John Wiley & Sons, Ltd.]

The origin of the Orbitrap analyzer began through the proof concept that a central spindle-like electrode could generate a purely harmonic potential in the direction of the z-axis. This was first devised by Kingdon in 1923, proven by Knight in 1981 by the Simion simulation, as shown in Figure 8A and applied by Makarov in 2000, as Kingdon's ideal trap, which would later be called Orbitrap [19,57].

The Orbitrap analyzer, also considered an ion trap, like the FTICR and the LIT, is configurated of a barrel-like anode electrode in a coaxial shape (external) that involves the cathode electrode in the shape of a spindle-like (central). The geometry confers a harmonic potential, ion oscillation stability, and purely governed by the electrostatic field (Figure 8B). The electrical voltage is applied axially between the central electrode, providing the stored ions with a stable orbital path and oscillation in the axial direction between the electrodes. Therefore, Orbitrap uses the oscillation frequency of the confined ions along the central electrode to determine its m/z ratio. These properties of the Orbitrap are the origin for the high resolution of the equipment. In contrast, the scanning speed of the m/z is considered moderate. As a rule, the use of high resolution affects the scanning speed and vice versa [19,57].



Figure 8. Summary of the operation of the Orbitrap mass analyzer. A) The Simion graphs of equipotential lines for ideal parameters of the Kingdon trap. The numbers refer to the potentials of the contour curves. Check that the central contours resemble the trajectory of the ions in Figure B. [Reprinted with permission from reference [19] Perry, R. H.; Cooks, R. G.; Noll, R. J. Mass Spectrom. Rev., 2008, 27, pp 661-699 (https://doi.org/10.1002/mas.20186). Copyright© (2021) John Wiley & Sons.] B) A sectional view of the Orbitrap analyzer. Through the C-trap the ions are injected into the Orbitrap at a point where z = 0 and perpendicular to the z-axis, where axial radiofrequency oscillations of the ions begin, without the need for additional excitation. The red cylinder is the area of the ions' trajectory. [Reprinted with permission from reference [58] Zubarev, R. A.; Makarov, A. Anal. Chem., 2013, 85, pp 5288-5296 (https://doi.org/10.1021/ac4001223). Copyright© (2013) American Chemical Society.]

The challenges to mass spectrometry analyzers concern intact protein analysis. So, scientists have invested in developing new FTICR cell harmonization, high-field Orbitrap geometries, and advances in signal processing. For that, the base unit resolution has recently been extended to proteins with masses of 150 kDa, using the FTICR or Orbitrap analyzers [59]. Due to the higher resolution of FTICR, it has been the analyzer of choice for top-down proteomics. However, the hybrid system LTQ-Orbitrap has proved to be a successful strategy.

In general terms, the quality of ion fragmentation determines the successful identification of peptides and proteins. Fragmentation in mass spectrometry activates the molecule, with the purpose to produce biomolecular "fingerprints" of a diverse set of molecules. The goal of all activation methods is essentially the same: to deposit energy in the ions to cause cleavages in chemical bonds reproducibly and to produce interpretable ion fragments, which means to reveal structural information or the sequence of the molecule of interest [60].

The fragmentation techniques used in proteomics are based on *i*) collision energy transfer (collisioninduced dissociation - CID, high energy collision dissociation - HCD, electron capture dissociation – ECD, and electron transfer dissociation - ETD, surface-induced dissociation - SID; ion-ion reactions) or *ii*) photonbased (infrared multiphoton dissociation - IRMPD and Ultraviolet photodissociation - UVPD) [7,60].

Fragmentation methods by collision energy transfer are more useful for peptides with lower molecular weight and less efficient for peptides with high molecular weight and intact proteins or species containing labile PTMs, such as phosphopeptides.

The electron-based fragmentation methods generate fragmentation patterns of larger molecules that favor the interpretation of a species with labile PTMs, such as phosphorylation and glycosylation. Thus, the fragmentations type ECD and ETD have efficient results in the characterization of intact proteins [11,59].

SID is the alternative that provides the highest energy to gas-phase collision methods, in which the ions are activated and fragmented after colliding with a surface (which works how a large target). In addition to its ability to generate rich fragmentation patterns, for many classes of ions, SID has also been used to characterize the structure of protein complexes due to its high energy transfer. In photodissociation techniques, ions accumulate energy through the absorption of one or more photons. Consequently, the fragmentation of the ionic species is obtained [60].

In particular, two fragmentation methods have contributed considerably to the advances in topdown proteomics, ETD, and ultraviolet photodissociation UVPD. Currently, they are the most efficient fragmentation methods for intact proteins, especially when used in combination [10,59].

CID/CAD collision-based methods are the most common fragmentation techniques in proteomics. These generally provide fragmentation at the amide bond of the peptide/protein. Thus, ions type *b* and *y*-are obtained. The HCD method provides fragmentations that result in the formation of ions of type *a*- and *x*-. Electron-based methods, such as ECD and ETD, generally result in the cleavage of the N-C bond and produce fragments of ions type *c*- and *z*- [7,11,59].

The analytical challenges of proteomics

Although there have been considerable advances in techniques for the separation and identification of biomolecules, the analysis of intact proteins from a complex mixture remains a challenge to analytical chemistry because of the extreme complexity of large-sized proteins, post-translational modifications, and physicochemical properties. Proteins have all possibilities of chemical interactions, in terms of intramolecular and intermolecular. That infers analyzes with low reproducibility in the analytical methods and techniques. Besides, proteins have a high dynamic range of protein expression, spanning many orders of magnitude (>10¹⁰). Additionally, proteins have conformational heterogeneity resulting from post-translational modifications (PTMs), besides expressed in a diversity of species in qualitative and quantitative terms. Therefore, our ability to understand a proteome through the analysis of intact proteins is currently limited [59].

Analytical chemistry has advanced significantly for the analysis of small molecules. Because of the less complexity, the chemical behavior of small molecules in analytical separation and detection systems is more conclusive for successful analysis. Certainly, shotgun and bottom-up proteomics have gained knowledge of analytical techniques for the analysis of small molecules. Consequently, the shotgun has provided greater characterization of the proteome emerging to expressive numbers in identifying proteins, never reported before. However, the analysis of a proteome from a complex mixture of several proteins digested into peptides increases the dependence on the accuracy of bioinformatics tools. The bioinformatics strategy alone does not have great chances to succeed in terms of the typical analytical figures of merits. That infers that the sequence of peptides into proteins in some cases brings certain bias. The main drawback is that peptides can assume a math property of combinatory analysis in the sequencing step. That infers some peptides can be sequenced the wrong way to some proteins, consequentially increase the number of identification of proteins that, in reality, does not exist. Thus, shotgun proteomics is a relative technique for protein identification, as well as for quantification. Also, protein coverage is desired to reach 100%, a challenging aspect for liquid chromatography and high-resolution mass spectrometry because of the misunderstood loss of peptides in the analysis process [13,61].

The high orthogonality and selectivity of LC×LC have an essential influence on the proteomics results, mainly due to the decrease in sample complexity in the separation, providing a better detection and biomolecular characterization in MS/MS. However, LC×LC is not able to separate all compounds from proteomes. Therefore, two or more compounds with similar retention time and molar mass (isobarics) can be considered the same precursor ion in the m/z range, and they are co-sequenced together, mainly in shotgun proteomics. This drawback is named "chimera" spectra. Thus, the results from the isolation and simultaneous fragmentation of two or more different molecular ions reduce the score values and the number of identifications of bioinformatics tools due to unidentifiable fragment ions derived from contaminating parent ions.

Bioinformatics scientists have developed alternatives to lessen the effect chimera through the statistical treatment of mass spectra. Additionally, it becomes essential to develop high-resolution mass spectrometers or strategies such as ion mobility-MS/MS, to increase the rate of ion isolation and consequent improvement in the detection of MS.

Another common problem in mass spectrometer applications is ionic suppression because some peptides or proteins can ionize more efficiently than others, causing that other biomolecules have an inefficient ionization. Thus, the first one can make ionic suppression phenomena over the second one that affects the sensibility of analysis implicating in not identifying bioinformatics tools.

Yet, the main challenges continue to be top-down and native mass spectrometry proteomics because the detection limit and sensitivity of the mass spectrometer for intact proteins are much lower than for peptides. In any type of MS instrument, the sensitivity decreases drastically with the increase of mass weight. Today, it is still relatively difficult to analyze intact proteins more massive than 70 kDa. Moreover, as the molecular mass of the protein increases, the tertiary structure of proteins becomes more difficult to disrupt, which thereby limits the MS/MS fragmentation efficiency of intact proteins. Thus, most of the top-down applications focused on proteins less than 50 kDa, and there are very few applications to date on larger proteins (>100 kDa) [56].

In any proteomic technique (bottom-up, middle-down, or top-down), the ideal condition to increase sensitivity and to achieve the lowest limit detection or quantification for 2D-LC are basically: *i*) the highest orthogonality of separation; *ii*) gradient using a nano-flow scale for less amount of mobile phase (that infers lower dilution of each chromatography peak and leads to great Coulombic explosion in the MS); *iii*) capillary chromatography columns, usually 75 μ m of inner diameter and 3 μ m of size particle; *iv*) reduced plate height following *van* Deemter equation, usually 250 nL min⁻¹. Likewise, in mass spectrometry provides: *i*) efficient ionization for the most suitable gas-phase state of all molecular ions to enter in the electromagnet field driving to the vacuum chamber of MS; *ii*) a suitable interval of time storage ions in the analyzers with high velocity to analysis and fragmentation in the sequent analyzer; *iii*) fragmentation quality that depends on technique and type of biomolecule analyzed; *iv*) high resolution, and *v*) high sensitivity. Usually, these two last parameters are the intrinsic conditions of the mass analyzer type of choice, which means how much the analyzer differentiates the background noise from the analytical signal. Usually, they are linked to an electromagnetic theory from physics, such as summarized in Figures 7 and 8.

Mathematical modeling and chemometrics tools to improve proteomics analysis

Mathematical modeling is based on the mathematical model concept to representing a complex phenomenon or incomprehensible subject. Therefore, it brings ways simpler and more palpable to understand complex data using simulations. The construction of a model requires multidisciplinary knowledge and the ability to make sophisticated guesses when getting information, testing, etc. The steps of mathematical modeling and troubleshooting can be summarized in Figure 9 [62].

Once the mathematical model has been built, it can be applied for studying the modeled phenomenon. The most important is to check if the model is coherent to represent genuinely the phenomena that occur in the problem investigated, for that some adjust in the equation(s), proving tests, and validations must be evaluated. If the mathematical model is robust, it can be used to investigated similar problems. In chemistry, simulation can bring faster and cheaper solutions, which means that it is unnecessary to do many experiments to observe phenomena because simulation can be used as the first step to refine the experiments [62].



Figure 9. The process of mathematical modeling. First, we must identify the problem that is trying to solve by modeling. Next, it is necessary to understand which building blocks (equations) have to be included in the model. At this step, we define the most critical variables and quantities of them, and we also think of any relevant background assumptions, simplifications, and so forth.

Chemometrics is a science-based tool that uses mathematical concepts to simplify complex chemical data or optimizations. For example, using the Design of Experiments (DoE), an input data analysis, the most used chemometrics tool, it is possible to evaluate several parameters to find two or three parameters most significant on an analytical method to be optimized. Although DoE is very used in the analytical chemistry field in proteomics it is very incipient with few papers published [63,64].

Another chemometric tool used, in this case, in output data, to proteomics is principal components analysis (PCA). It is useful to simplify a complex data analysis to bring relevant information to interpretation. PCA uses the concept of dimensionality reduction by projecting of new axis (PCs). For example, in a proteomics analysis with more than 1,000 proteins for each sample analyzed, PCA allows us to plot 2D or 3D graphs that separate or put together sample groups. Thus, separate groups have different chemical information, and groups in the set have similar chemical information. Without this strategy, it is practically impossible to understand the most important chemical information of different samples analyzed. Another very well used chemometric tool used in proteomics is Hierarchical Cluster Analysis (HCA). Here different samples form clusters until all of them in the dataset are linked together in a hierarchical tree [65–67].

Other useful chemometrics tools can be used in proteomics analysis include multivariate calibration, Kth Nearest Neighbor (k-NN), Soft Independent Modeling of Class Analogy (SIMCA), Linear Discriminant Analysis (LDA), Partial Least Squares – Discriminant Analysis (PLS-DA), Partial Least Squares – Regression (PLS-R), Orthogonal Partial Least Squares – Discriminant Analysis (OPLS-DA), Parallel Factor Analysis (PARAFAC), PARAFAC 2, and Artificial Neural Networks (ANN). Some of these tools have been applied currently in proteomics analysis in food and biological questions. However, application of chemometrics tools in proteomics is still incipient. Although chemometrics is an excellent strategy to obtain relevant information from complex samples, it must be used with caution because, before its application, a study on the mathematical and statistical treatment must be considered. In general terms, the consensus on the previous information of the sample must have a link to the chemical information the chemometrics

has provided. The detailed description of all of these useful tools would demand a complementary review, which is not the goal on this one. However, Table III summarizes the main applications of chemometrics tools in proteomics data analysis [67–69].

Field	Sample -	Chemometrics tools used						Def		
Fleid		DoE	PCA	HCA	PLS-DA	LDA	PLS-R	OPLS-DA	ANN	Ret.
Food Chemistry	Meat		\checkmark							[70]
Food Chemistry	Avocados		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		[71]
Food Chemistry	Meat		\checkmark	\checkmark	\checkmark					[72]
Food Chemistry	Cucumber		\checkmark	\checkmark				\checkmark		[73]
Food Chemistry	Meat		\checkmark					\checkmark		[74]
Food Chemistry	Shrimp		\checkmark					\checkmark		[75]
Food Chemistry	Biactive peptides							\checkmark		[76]
Analytical Chemistry	Mix of proteins				\checkmark					[77]
Analytical Chemistry	Mix of proteins		\checkmark	\checkmark						[78]
Analytical Chemistry	Saccharomyces cerevisiae	\checkmark								[64]
Analytical Chemistry	Saccharomyces cerevisiae	\checkmark								[63]
Health	Cutaneous Leishmaniasis		\checkmark							[79]
Health	Ovarian cancer								\checkmark	[80]

In the last years, a chemometric strategy based on DoE has been introduced, named as Chemical Mathematical Model to maximize protein sequence coverage for shotgun proteomics. Thus, two significant parameters were statistically evaluated from the whole LC×LC-MS/MS platform [81,82]. These parameters show more representative to increase shotgun proteomics coverage. Besides, the evidence shows that performance is strongly linked to the chemical nature of the peptides, for example, hydrophobicity. Thus, the origin of the chemical approach of the model is revelated. Figure 10 shows the behavior of the surface of response in different methods to improve identification of proteoforms in shotgun and bottom-up proteomics analysis, once it allows greater coverage of proteins. Note that different instrumental conditions are necessary to detect different hydrophobicity of peptides. Lastly, an optimization condition from a complex system (LC×LC-MS/MS) can be amenable to simplification into mathematical models due to *Pareto Principle* [81,82].



Figure 10. Two Chemical Mathematical Model to maximize protein sequence coverage for shotgun proteomics. A) Digested bovine serum albumin (BSA) sample to evaluate the optimal conditions of SCX-RP-Q-ToF. B) Digested yeast proteins to evaluate the optimal conditions of MudPIT-LTQ-Orbitrap. The x-axis is the significant parameter from 2D-LC and the y-axis is the significant parameter from ESI-MS/MS. Areas in red means larger peptides identifications through evaluation of three responses i) Coverage of Proteins ii) Score, and iii) the number of peptides. [Reprinted (adapted) with permission from reference [82] Batiston, W. P. Exploring the fundamentals of liquid chromatography and mass spectrometry for integration between proteomics, microfluidics, and chemometrics. Doctoral thesis, 2020. Sao Carlos Institute of Chemistry, University of Sao Paulo, Sao Carlos, SP, Brazil (https://doi.org/10.11606/T.75.2020.tde-27082020-142027).]

Finally, bioinformatics has emerged as an imperative tool for proteomics analysis and is essential in any proteomics technique (bottom-up, middle-down, or top-down). The most important subject of study of bioinformatics is the algorithm. An algorithm is a finite sequence from executable actions that gets a solution for a specific problem. As the algorithm are precise, unambiguous, mechanical, efficient, and correct procedures it is important to a deep understanding of the figures of merit in analytical chemistry for proteomics. Among the main outcomes from bioinformatics, we can benefit from reproducible analysis, precision, accuracy, chemical behaviors of the biomolecule and its fragmentation pattern, as well as understanding the random or systematic errors. Because an algorithm only can interpret a pattern or well-defined behavior, a proteomic analysis with significant random errors could infer some bias in the results from algorithm interpretation.

CONCLUSIONS

In terms of theoretical perspectives on the top-down and native proteomics, they would have conditions suitable to identify with accuracy proteins. However, there are tremendous analytical challenges to overcome the direct identification of proteins with >70 kDa to mass spectrometry. One of the main drawbacks sounds inevitable; the structural diversity of proteins when submitted to analytical methods drive different chemical structures under the same condition of analysis, what substantially decreases the reproducibility of the methods. Besides the expressive number of species to the identification and the challenges to get a good fragmentation in mass spectrometry for biomolecules above 50 kDa. Although of the several limitations of peptides analysis, consequently driven to relative identification of proteins. We can infer that shotgun and bottom-up proteomics must persist in studying the massive scale of analysis from proteomes using LC×LC-MS/MS, because of great simplicity and very well-established analysis ways. A light at the end of the tunnel for overcoming the drawbacks to separations limitations in the intact proteins could be the exploration of new stationary phases for high resolution and sensitivity in nanoflow liquid chromatography, such as monolithic columns or open tubular columns. Besides, in mass spectrometry, the understanding of new strategies and sophisticated ion control, increases resolution, and new fragmentation modes could provide more evidence to improve proteomics analysis. However, the development of analytical instruments demands long time, elevated costs, and skilled personnel. Such bottlenecks are not interesting to most of the scientific community, which can put the field in a stagnation, even if it is of high interest to analytical instrument companies. Smart strategies can improve analytical figures of merit in proteomics, like mathematical modeling, chemometrics tools, and continuous advances in bioinformatics.

Conflicts of interest

The authors declared having no conflicts of interest.

Acknowledgements

The São Paulo Research Foundation (FAPESP), [grant number 2015/16025-8] to WPB is gratefully acknowledged. Moreover, INCTBio is kindly acknowledged for the grants from FAPESP [grant number 2014/50867-3] and Council for Scientific and Technological Development (CNPq) [grant number 465389/2014-7].

REFERENCES

- 1. Nelson, D. L.; Cox, M. M. *Princípios de Bioquímica de Lehninger*. Artmed, Porto Alegre, **2018**, 7th ed., Chapter 5, pp 89-182.
- 2. Pauling, L. Protein Science, 1993, 2 (6), pp 1060-1063 (https://doi.org/10.1002/pro.5560020620).
- 3. Pankow, S.; Bamberger, C.; Yates, J. R. III. *Sci. Signaling*, **2019**, *12* (562), pp 1-14 (https://doi. org/10.1126/scisignal.aan7984).
- 4. Háda, V.; Bagdi, A.; Bihari, Z.; Timári, S. B.; Fizil, A.; Szántay Jr., C. *J. Pharm. Biomed. Anal.*, **2018**, *161*, pp 214-238 (https://doi.org/10.1016/j.jpba.2018.08.024).
- Salermo, M. S.; Matsumoto, C.; Ferraz, M. S. Instituto de Pesquisa Econômica Aplicada (Ipea). Ministério do Planejamento, Desenvolvimento e Gestão. *Biofármacos no Brasil: características, importância e delineamento de políticas públicas para seu desenvolvimento*, **2018**. Available from: http://www.ipea.gov.br/portal/publicacoes [Accessed 13 April 2020].
- 6. Larance, M.; Lamond, A. I. *Nat. Rev. Mol. Cell Biol.*, **2015**, *16* (5), pp 269-280 (https://doi.org/10.1038/ nrm3970).
- 7. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R. *Chem. Rev.*, **2013**, *113* (4), pp 2343-2394 (https://doi.org/10.1021/cr3003533).
- 8. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. *Annual Reviews of Biomedical Engineering*, **2009**, *11*, pp 49-79 (https://doi.org/10.1146/annurev-bioeng-061008-124934).
- Breuker, K.; McLafferty, F. W. Proc. Natl. Acad. Sci. U. S. A., 2008, 105 (47), pp 18145-18152 (https:// doi.org/10.1073/pnas.0807005105).
- 10. Van Den Heuvel, R. H. H.; Heck, A. J. R. *Curr. Opin. Chem. Biol.*, **2004**, 8 (5), pp 519-526 (https://doi. org/10.1016/j.cbpa.2004.08.006).
- 11. Toby, T. K.; Fornelli, L.; Kelleher, N. L. *Annu. Rev. Anal. Chem.*, **2016**, 9 (1), pp 499-519 (https://doi. org/10.1146/annurev-anchem-071015-041550).
- 12. Chait, B. T. Science, 2006, 314 (5796), pp 65-66 (https://doi.org/10.1126/science.1133987).
- 13. Michalski, A.; Cox, J.; Mann, M. *J. Proteome Res.*, **2011**, *10* (4), pp 1785-1793 (https://doi.org/10.1021/ pr101060v).
- 14. Meyer, B; Papasotiriou, D. G.; Karas, M. *Amino Acids*, **2011**, *41*, pp 291-310 (https://doi.org/10.1007/ s00726-010-0680-6).
- 15. Han, X.; Aslanian, A.; Yates, J. R. *Curr. Opin. Chem. Biol.*, **2008**, *12* (5), pp 483-490 (https://doi. org/10.1016/j.cbpa.2008.07.024).
- 16. Fonslow, B. R.; Carvalho, P. C.; Academia, K.; Freeby, S.; Xu, T.; Nakorchevsky, A.; Paulus, A.; Yates, J. R. *J. Proteome Res.*, **2011**, *10* (8), pp 3690-3700 (https://doi.org/10.1021/pr200304u).
- Xu, T.; Park, S. K.; Venable, J. D.; Wohlschlegel, J. A.; Diedrich, J. K.; Cociorva, D.; Lu, B.; Liao, L; Hewel, J.; Han, X.; Wong, C. C. L.; Fonslow, B.; Delahunty, C.; Gao, Y.; Shah, H.; Yates, J. R. *J. Proteomics*, **2015**, *129* (3), pp 16-24 (https://doi.org/10.1016/j.jprot.2015.07.001).

- McLafferty, F. W.; Lory, E. R. J. Chromatogr. A, **1981**, 203 (9), pp 109-116 (https://doi.org/10.1016/ S0021-9673(00)80285-X).
- 19. Perry, R. H.; Cooks, R. G.; Noll, R. J. *Mass Spectrom. Rev.*, **2008**, 27, pp 661-699 (https://doi. org/10.1002/mas.20186).
- 20. Valeja, S. G.; Xiu, L.; Gregorich, Z. R.; Guner, H.; Jin, S.; Ge, Y. *Anal. Chem.*, **2015**, 87 (10), pp 5363-5371 (https://doi.org/10.1021/acs.analchem.5b00657).
- 21. Stoll, D. R.; Carr, P. W. Anal. Chem., **2017**, *89* (1), pp 519-531 (https://doi.org/10.1021/acs. analchem.6b03506).
- 22. Stoll, D. R.; LI, X.; Wang, X.; Carr, P. W.; Porter, S. E. G.; Rutan, S. C. *J. Chromatogr. A*, **2007**, *1168* (1-2), pp 3-43 (https://doi.org/10.1016/j.chroma.2007.08.054).
- 23. Cassiano, N. M.; Barreiro, J. C.; Oliveira, R. V.; Cass, Q. B. *Bioanalysis*, **2012**, *4* (22), pp 2737-2756 (https://doi.org/10.4155/bio.12.226).
- 24. François, I.; Sandra, K.; Sandra, P. *Anal. Chim. Acta*, **2009**, *641* (1-2), pp 14-31 (https://doi.org/10.1016/j. aca.2009.03.041).
- 25. Dolan, J. W.; Snyder, L. R.; Djordjevic, N. M.; Hill, D. W.; Waeghe, T. J. *J. Chromatogr. A*, **1999**, *857* (1-2), pp 21-39 (https://doi.org/10.1016/S0021-9673(99)00767-0).
- 26. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Anal. Chem.*, **2005**, 77 (19), pp 6426-6434 (https://doi. org/10.1021/ac050923i).
- 27. Giddings, J. C. *J. Chromatogr. A*, **1995**, 703 (1-2), pp 3-15 (https://doi.org/10.1016/0021-9673(95)00249-M).
- 28. Mogollón, N. G. S.; De Lima, P. F.; Gama, M. R.; Furlan, M. F.; Braga, S. C. G. N.; Prata, P. S.; Jardim, I. C. S. F.; Augusto, F. *Quim. Nova*, **2014**, *37* (10), pp 1680-1691 (https://doi.org/10.5935/0100-4042.20140261).
- Palma, S. D.; Hennrich, M. L.; Heck, A. J. R.; Mohammed, S. J. Proteomics, 2012, 75, pp 3791-3813 (https://doi.org/10.1016/j.jprot.2012.04.033).
- Cacciola, F; Mangraviti, D; Mondello, L; Dugo, P. Hyphenations of 2D capillary-based LC with mass spectrometry, Elsevier Inc, Netherlands, 2020, Chapter 3.4, pp 369-412 (https://doi.org/10.1016/ C2015-0-05862-2).
- 31. Stoll, D. R. *Bioanalysis*, **2015**, 7 (24), pp 3125–3142 (https://doi.org/10.4155/bio.15.223).
- 32. Zhang, L; Liu C. W.; Zhang, Q. *Anal. Chem.*, **2018**, *90* (2), pp 1081–1086 (https://doi.org/10.1021/acs. analchem.7b03342).
- 33. Li, G.; Li, M.; Liang, X.; Xiao, Z.; Zhang, P.; Shao, M.; Peng, F.; Chen, Y.; Li, Y.; Chen, Z. *J. Cancer*, **2017**, *8* (3), pp 479–489 (https://doi.org/10.7150/jca.17192).
- 34. Cassidy, L.; Prasse, D.; Linke, D.; Schmitz, R. A.; Tholey, A. *J. Proteome Res.*, **2016**, *15* (10), pp 3773–3783 (https://doi.org/10.1021/acs.jproteome.6b00569).
- 35. Dong, Q.; Yan, X.; Liang, Y.; Stein, S. E. *J. Proteome Res.*, **2016**, *15* (5), pp 1472–1486 (https://doi. org/10.1021/acs.jproteome.5b01046).
- 36. Yeung, D.; Mizero, B.; Gussakovsky, D.; Klaassen, N.; Lao, Y.; Spicer, V.; Krokhin, O. V. *Anal. Chem.*, **2020**, *92* (5), pp 3904–3912 (https://doi.org/10.1021/acs.analchem.9b05407).
- Law, H. C. H.; Szeto, S. S. W.; Quan, Q.; Zhao, Y.; Zhang, Z.; Krakovska, O.; Lui, L. T.; Zheng, C.; Lee, S. M. Y.; Siu, K. W. M.; Wang, Y.; Chu, I. K. *J. Proteome Res.*, **2017**, *16* (3), pp 1150–1166 (https://doi. org/10.1021/acs.jproteome.6b00651).
- 38. Ren, W; Yin J, Chen S, Duan J.; Liu G.; Li T.; Li N.; Peng Y.; Tan B.; Yin Y. *Sci Rep*, **2016**, (6), pp 1–9. (https://doi.org/10.1038/srep25640).
- 39. Quan, Q.; Szeto, S. S. W.; Law, H. C. H.; Zhang, Z.; Wang, Y.; Chu, I. K. *Anal. Chem.*, **2015**, *87* (19), pp 10015–10024 (https://doi.org/10.1021/acs.analchem.5b02619).
- 40. Vonk, R. J.; Gargano, A. F. G.; Davydova, E.; Dekker, H. L.; Eeltink, S.; Koning, L. J.; Schoenmakers, P. J. *Anal Chem*, **2015**, *87* (10), pp 5387–5394 (https://doi.org/10.1021/acs.analchem.5b00708).
- 41. Vanhoenacker, G.; Vandenheede, I.; David, F.; Sandra, P.; Sandra, K. *Anal. Bioanal. Chem.*, **2015**, *407*, pp 355–366 (https://doi.org/10.1007/s00216-014-8299-1).

- 42. Stoll, D. R.; Harmes, D. C.; Danforth, J.; Wagner-Rousset, E.; Guillarme, D.; Fekete, S.; Beck A. *Anal. Chem*, **2015**, *87*, pp 8307–8315 (https://doi.org/10.1021/acs.analchem.5b01578).
- 43. Wang, Z.; Yu, D.; Cupp-Sutton, K. A.; Liu, X.; Smith, K.; Wu, Si. *Anal. Chem.*, **2020**, *92*, pp 12774–12777 (https://doi.org/10.1021/acs.analchem.0c03395).
- 44. Brown, K. A.; Tucholski, T.; Alpert, A. J.; Eken, C.; Wesemann, L.; Kyrvasilis, A.; Jin, S.; Ge, Y. *Anal. Chem.*, **2020**, 92, pp 15726–15735 (https://doi.org/10.1021/acs.analchem.0c02533).
- 45. Pepaj, M.; Thorsby, P. M. *J. Liq. Chromatogr. Relat. Technol.*, **2015**, *38*, pp 20–28 (https://doi.org/10.1 080/10826076.2013.864980).
- 46. Li, Y.; Gu, C.; Gruenhagen, J.; Zhang, K.; Yehl, P.; Chetwyn, N. P.; Medley, C. D. *J. Chromatogr. A*, **2015**, *1393*, pp 81–88 (https://doi.org/10.1016/j.chroma.2015.03.027).
- 47. Li, Y.; Hewitt, D.; Lentz, Y. K.; Ji, J. A.; Zhang, T. Y.; Zhang, K. *Anal. Chem.*, **2014**, *86*, pp 5150–5157 (https://doi.org/10.1021/ac5009628).
- 48. Julka, S.; Kuppannan, K.; Karnoup, A.; Dielman, D.; Schafer, B.; Young, S. A. *Anal. Chem.*, **2012**, *84*, pp 10019–10030 (https://doi.org/10.1021/ac3024685).
- 49. Mazur, M. T.; Seipert, R. S.; Mahon, D.; Zhou, Q.; Liu, T. *AAPS J.*, **2012**, *14*, pp 530–541 (https://doi. org/10.1208/s12248-012-9361-6).
- 50. Alvarez, M.; Tremintin, G.; Wang, J.; Eng, M.; Kao, Y-H.; Jeong, J.; Ling, V. T.; Borisov, O. V. *Anal. Biochem.*, **2011**, *419*, pp 17–25 (https://doi.org/10.1016/j.ab.2011.07.033).
- 51. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science*, **1989**, *246* (4926), pp 64-71 (https://doi.org/10.1126/science.2675315).
- 52. Tanaka, K. Angew. Chem., Int. Ed., **2003**, 42 (33), pp 3860-3870 (https://doi.org/10.1002/ anie.200300585).
- 53. Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science*, **2004**, *306* (5695), pp 471-473 (https://doi.org/10.1126/science.1104404).
- 54. Wilm, M.; Mann, M. Anal. Chem., 1996, 68 (1), pp 1-8 (https://doi.org/10.1021/ac9509519).
- 55. Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.*, **2002**, *13*, pp 659-669 (https://doi.org/10.1016/S1044-0305(02)00384-7).
- 56. March, R. E. Quadrupole ion trap mass Spectrometer. In: *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Ltd. **2012**, p 4 (https://doi.org/10.1002/9780470027318.a6015.pub2).
- 57. Makarov, A. Anal. Chem., 2000, 72 (6), pp 1156-1162 (https://doi.org/10.1021/ac991131p).
- 58. Zubarev, A. R.; Makarov, A. Anal. Chem., 2013, 85, pp 5288-5296 (https://doi.org/10.1021/ac4001223).
- 59. Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L. *Annu. Rev. Anal. Chem.*, **2014**, 7 (1), pp 427-454 (https://doi.org/10.1146/annurev-anchem-071213-020216).
- 60. Brodbelt, J. S. Chem. Soc. Rev., 2014, 43 (8), pp 2757-2783 (https://doi.org/10.1039/C3CS60444F).
- 61. Shen, Y.; Zhao, R.; Berger, S. J.; Anderson, G. A.; Rodriguez, N.; Smith, R. D. *Anal. Chem.*, **2002**, 74 (16), pp 4235–4249 (https://doi.org/10.1021/ac0202280).
- 62. Pohjolainen, S. *Mathematical Modeling*. Springer, Switzerland, **2016**, Chapter 1, p 241 (https://doi. org/10.1007/978-3-319-27836-0).
- 63. Andrews, G. L.; Dean, R. A.; Hawkridge, A. M.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.*, **2011**, 22, pp 773–783 (https://doi.org/10.1007/s13361-011-0075-2).
- 64. Randall, S. M.; Cardasis, H. L.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.*, **2013**, *24*, pp 1501–1512 (https://doi.org/10.1007/s13361-013-0693-y).
- 65. Barros Neto, B; Scarminio, I. S.; Bruns, R. E. *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria.* Bookman, Porto Alegre, **2010**, p 401.
- 66. Pereira Filho, E. R. *Planejamento fatorial em química: maximizando a obtenção de resultados.* EdUFSCar, São Carlos, **2015**, p 88.
- 67. Ferreira, M. M. C. Quimiometria: conceitos, métodos e aplicações. Unicamp, Campinas, 2015, p 407.
- 68. Bro R. *Chemom. Intell. Lab. Syst.*, **1997**, *38* (2), pp 149–171 (https://doi.org/10.1016/S0169-7439(97)00032-4).

- 69. Bro, R.; Smilde, A. K. Anal. Methods, 2014, 6, pp 2812–2831 (https://doi.org/10.1039/c3ay41907j).
- 70. Yuswan, M. H.; Aizat, W. M.; Desa, M. N. M.; Hashim, A. M.; Rahim, N. A.; Mustafa, S.; Mohamed, R.; Lamasudin, D. U. *Chemom. Intell. Lab. Syst.*, **2019**, *192*, pp 1-7 (https://doi.org/10.1016/j. chemolab.2019.103825).
- Varrota, V. G.; Fuentealba, C.; Hernández, I.; Defilippi-Bruzzone, B.; Meneses, C.; Campos-Vargas, R.; Lurie, S.; Hertog, M.; Carpentier, S.; Poblete-Echeverría, C.; Pedreschi, R. *Food Chem.*, **2019**, *289*, pp 512–521 (https://doi.org/10.1016/j.foodchem.2019.03.090).
- 72. Mi, S.; Li, X.; Zhang, C. H.; Liu, J-Q.; Huang, D-Q. *Food Res. Int.*, **2019**, *119*, pp 426–435 (https://doi. org/10.1016/j.foodres.2019.02.016).
- 73. Zhang, H.; Zhang, X.; Zhao, X.; Xu, J.; Lin, C.; Jing, P.; Hu, L.; Zhao, S.; Wang, X.; Li, B. *Food Chem.*, **2019**, *274*, pp 592–602 (https://doi.org/10.1016/j.foodchem.2018.08.082).
- 74. Yuswan, M. H.; Aizat, W. M.; Lokman, A. A.; Desa, M. N. M.; Mustafa, S.; Junoh, N. M.; Yusof, Z. N. B.; Mohamed, R.; Mohmad, Z.; Lamasudin, D. U. *Food Anal. Methods*, **2018**, *11*, pp 3505–3515 (https:// doi.org/10.1007/s12161-018-1327-6).
- 75. Hu, L.; Zhang, H.; Zhang, X.; Zhang, T.; Chang, Y.; Xue, Z.; Xu, J.; Xue, Y.; Li, Z.; Wang, Y.; Xue, C. *J. Agric. Food Chem.*, **2018**, 66, pp 10567–10574 (https://doi.org/10.1021/acs.jafc.8b04375).
- 76. Li, G.; Zhang, Z.; Quan, Q.; Jiang, R-W.; Szeto, S. S. W.; Yuan, S.; Wong, W-T.; Lam, H. H. C.; Lee, S. M. Y.; Chu, I. K. *J Proteome Res.*, **2016**, *15*, pp 2595–2606 (https://doi.org/10.1021/acs. jproteome.6b00152).
- 77. Michaud, F. T.; Garnier, A.; Lemieux, L.; Duchesne, C. *Proteomics*, **2009**, 9, pp 512–520 (https://doi. org/10.1002/pmic.200800300).
- 78. Bäckström, D.; Moberg, M.; Sjöberg, P. J. R.; Bergquist, J.; Danielsson, R. *J. Chromatogr. A*, **2007**, *1171*, pp 69–79 (https://doi.org/10.1016/j.chroma.2007.09.042).
- 79. Negrão, F.; Diedrich, J. K.; Giorgio, S.; Eberlin, M. N.; Yates, R. J. ACS Infect. Dis., **2019**, *5*, pp 2136–2147 (https://doi.org/10.1021/acsinfecdis.9b00275).
- 80. Swiatly, A.; Horala, A.; Hajduk, J.; Matysiak, J.; Nowak-Markwitz, E.; Kokot, A. J. *BMC Cancer*, **2017**, *17*, pp 1–9 (https://doi.org/10.1186/s12885-017-3467-2).
- Batiston, W. P. Chemometric approach to capillary multidimensional liquid chromatography coupled to tandem mass spectrometry for shotgun proteomics. Master dissertation, **2015**, Sao Carlos Institute of Chemistry, University of Sao Paulo, Sao Carlos, SP, Brazil (https://doi.org/10.11606/D.75.2015.tde-07052015-110233).
- 82. Batiston, W. P. *Exploring the fundamentals of liquid chromatography and mass spectrometry for integration between proteomics, microfluidics, and chemometrics.* Doctoral thesis, **2020**. Sao Carlos Institute of Chemistry, University of Sao Paulo, Sao Carlos, SP, Brazil (https://doi.org/10.11606/T.75.2020. tde-27082020-142027).