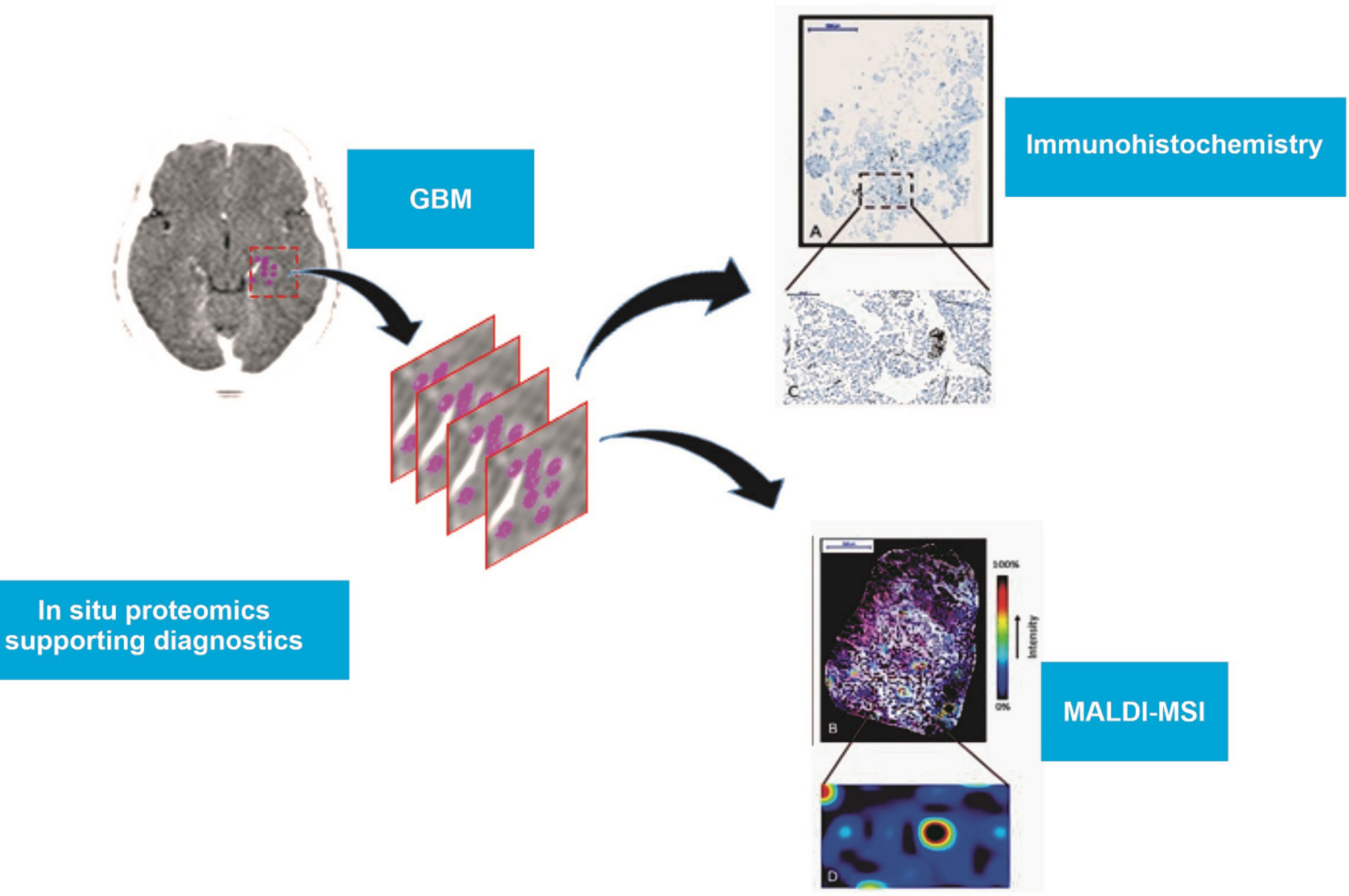
BrJAC

Brazilian Journal of Analytical Chemistry

an International Scientific Journal



In Situ Proteomic Analysis of Glioblastoma Multiforme

A Translational Approach to Improve Prognostic/Diagnostic Routines

Anally Ribeiro da Silva Menegasso, Marcel Pratavieira, Lucilene Delazari dos Santos, Flávio de Oliveira Lima, Marcelo Padovani T. Moraes, Marco Antonio Zanini, Mario Sergio Palma

October – December 2020 Volume 7 Number 29

IMSC 2021 will cover all aspects of mass spectrometry, from fundamentals to instrumentation and applications, with the most recent findings and innovations and a very selected group of renowned speakers but with a great addition: our "Brazillian seasoning".

The conference is organized with short courses, oral and poster sessions, workshops, awards, commercial exhibitions, as well as very exciting social programs and companies user meetings. All major MS companies are supporting Rio 2021 and will be in the exposition area presenting their latest products and innovations.

We are working very hard to make Rio 2021 the largest and greatest ever IMSC and for it to be an exceptional occasion for you to enjoy science, to grow your network and make new friends and colleagues, to get new ideas for your research, for establishing new collaborations, starting new projects, and more; all in the unforgettably warm, beautiful and charming city of Rio de Janeiro!

Please come, bring your students, invite friends and colleagues and contribute to make Rio 2021 a very successful, most amazing and unforgettable IMSC ever!

You are very welcome to RIO 2021!

Access WEBSITE

Welcome to the

23rd **International Mass Spectrometry Conference** 2021 in Rio de Janeiro

December 12 to December 16

IMSF international mass spectrometry foundation

Sociedade Brasileira de Espectrometria de Massas
BrMASS

Brazilian Journal of Analytical Chemistry

VISÃO FOKKA - COMUNICATION AGENCY

The *Braz. J. Anal. Chem.* (BrJAC) is a peer-reviewed, open access, scientific journal intended for professionals and institutions acting mainly in all branches of analytical chemistry.

ISSN 2179-3425 printed ISSN 2179-3433 eletronic

Scope

BrJAC is dedicated to the diffusion of significant and original knowledge in all branches of Analytical Chemistry and Bioanalytics. BrJAC is addressed to professionals involved in science, technology and innovation projects in Analytical Chemistry at universities, research centers and in industry.

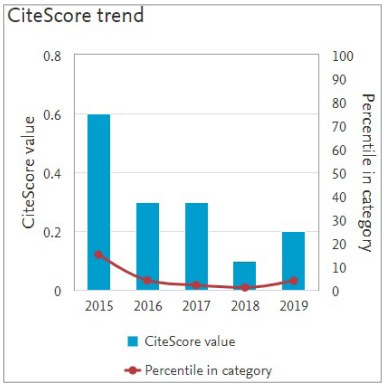
BrJAC is a quarterly journal that publishes original, unpublished scientific articles, reviews and technical notes that are peer reviewed in the double-blind way. In addition, it publishes interviews, points of view, letters, sponsor reports, and features related to analytical chemistry. Once published online a DOI number is assigned to the paper.

Manuscripts submitted for publication in BrJAC cannot have been previously published or be currently submitted for publication in another journal. For manuscript preparation and submission, please see the Guidelines for the Authors section at the end of this edition.

Indexing Sources

Scopus

CiteScore 2019: 0.2; SJR: 0.112; SNIP: 0.086. CiteScore Tracker 2020: 0.3 Access source details



CLARIVATE ANALYTICS EMERGING SOURCES CITATION INDEX INDEXED

Expectancy: to be in the SCIE in 2021

CAPES QUALIS

Expectancy: to be classified in the A3 stratum for the Quadrennial Evaluation in 2021

Google scholar

Production Editor
Silvana Odete Pisani

Publisher
Lilian Freitas
MTB: 0076693/ SP
lilian.freitas@visaofokka.com.br

Advertisement
Luciene Campos
luciene.campos@visaofokka.com.br

Art Director: Adriana Garcia

WebMaster: Daniel Letieri

BrJAC's website: www.brjac.com.br / Contact: brjac@brjac.com.br
Like BrJAC on Facebook: https://www.facebook.com/brjachem

ABEC BRASIL BrJAC is associated to the Brazilian Association of Scientific Editors

VISÃO FOKKA

BrJAC is published quarterly by:
Visão Fokka Communication Agency
Av. Washington Luiz, 4300 - Bloco G - 43
13042-105 — Campinas, SP, Brazil
contato@visaofokka.com.br
www.visaofokka.com.br

EDITORIAL BOARD

EDITORIAL ADVISORY BOARD

EDITORIAL ADVISORY BOARD (Continuation)

CONTENTS

CONTENTS (Continuation)

Releases

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 1-2
doi: 10.30744/brjac.2179-3425.editorial.mspalma

BrJAC

EDITORIAL

Proteomics: The Science of Functional and Structural Characterization of Proteins

Mario Sergio Palma
Guest Editor of this special edition of BrJAC dedicated to Omics Tools
Full Professor of Proteomics
Department of Basic and Applied Biology, Institute of Biosciences of Rio Claro
São Paulo State University (UNESP), Rio Claro, SP, Brazil

The science of proteins was born many years ago as a branch of biochemistry, initially associated with medical applications. At that time, the standard protocols required the initial isolation and purification of an individual protein followed by the use of classical approaches of protein chemistry for amino acid sequencing and structural characterization. The analytical strategies were based on the use of large and low-performance chromatography columns and many spot tests for the chemical characterization of individual amino acids. This usually required very large amounts (grams/kilograms) of biological material in the initial steps of protein isolation and purification. At the time, most sequencing methods were based on the application of controlled chemical degradation protocols followed by thin layer chromatography of protein/peptide hydrolysates. At first, all these analytical steps were operated manually; later, protein science adopted automation. However, the analytical strategies were focused on the individual investigation of each protein, which made progress on the structural characterization of each protein very slow (in general, 3–4 years of work were necessary to achieve the complete sequencing of a protein).

The initial application of mass spectrometry for peptide sequencing was made by using N-terminal-directed acid hydrolysis of tryptic peptides (obtained by the proteolysis of a purified protein), followed by acid esterification to make them sufficiently volatile to be analyzed by GC-MS. Many important proteins were sequenced using this strategy.

Another aspect that must be emphasized was the success of DNA sequencing projects abroad that gave rise to several genome projects; these initiatives resulted in many different DNA databanks. The use of the universal genetic code to translate DNA sequences virtually into protein sequences, together with the development of algorithms for carrying out this translation and sequence alignment, contributed to create novel protein databanks. To give an idea about the impact of these initiatives, by the end of the 1990s the most-consulted protein databanks contained some hundreds of protein sequences; these new protein sequences generated by the virtual translation of deposited DNA sequences generated many thousands of protein sequences in a few years, achieving currently around 200 million entries. This progress represented an important boost for the popularization of protein science.

The immediate consequence was the combination of 2D electrophoresis of proteins (for the separation of large numbers of individual proteins) with protocols for in-gel digestion, followed by mass spectrometry analysis of the proteolytic peptides; this strategy resulted in the first experimental approach for the simultaneous identification of large numbers of proteins (peptide mass fingerprinting). At this time, the term "**genomics**" associated with studies on DNA sequencing started to be used, this meaning the complete DNA sequencing of all chromosomes of an organism. As an analogy with this, in 1994 the term "**proteomics**" was coined to refer to the complement of proteins expressed by a cell / tissue / secretion / whole organism under determined experimental conditions. This initiative has been considered the official birth of proteomics and peptidomics.

Cite: Palma, M. S. **Proteomics: The Science of Functional and Structural Characterization of Proteins.** *Braz. J. Anal. Chem.,* 2020, *7* (29), pp 1-2. doi: http://dx.doi.org/10.30744/brjac.2179-3425.editorial.mspalma

The possibility of the identification of several proteins in a single analytical round created an atmosphere of excitement surrounding proteomics, which stimulated scientists, algorithm developers and manufacturers of analytical instruments to face the challenges for the development of novel mass spectrometers, software for data treatment and the interpretation of amino acid sequencing, and HPLC systems dedicated to proteomics analysis. Thus, proteomics became a very attractive multidisciplinary area of investigation in which analytical proteomics played a pivotal role. The Brazilian Journal of Analytical Chemistry (BrJAC) may become an interesting academic forum for discussions about the development of proteomics, especially for those aspects related to the translational applications of analytical chemistry.

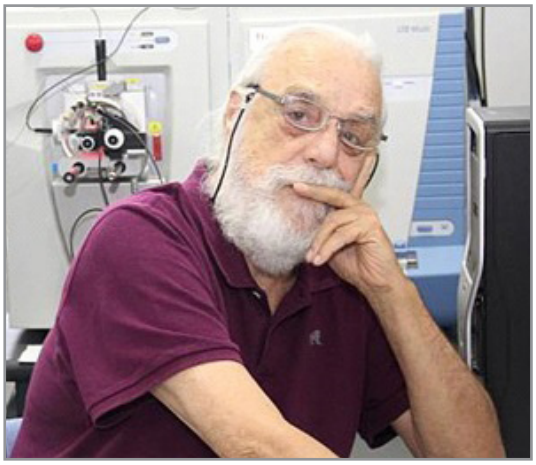
The Brazilian proteomics community is very active and productive, investigating a wide variety of subjects and reflecting the large biodiversity of Brazilian nature as well as themes related to agriculture and human health. To celebrate the opportunity presented by BrJAC, some contributions from leading Brazilian proteomics laboratories have been selected for this special issue. The issue presents an interview with Professor Gilberto Barbosa Domont (UFRJ) – a well-respected protein chemist based abroad and one of the pioneers of proteomics in Brazil. Professor Domont discusses aspects of his scientific career, talks about scientific politics in Brazil, and reveals his inspiration to become such an active and productive scientist. The section “Point of View” was written by Professor Marcelo Valle de Sousa (UnB) – also one of the pioneers of this field in Brazil. He presents some reflections about the development of proteomic analysis without the use of mass spectrometry. The “Letter” was written by Professor Daniel Martins de Souza (UNICAMP) – a brilliant young researcher in our community – in which he focuses on the understanding of schizophrenia by the use of MS-based proteomics. The present issue also presents two articles illustrating the state of the art of proteomics in Brazil: one is a contribution to analytical proteomics made by the study of human intestinal mucus and describes the interesting set-up conditions for the analysis of a very difficult biological matrix; the other contribution is an article describing the use of MALDI/MS and MS/MS to perform in-situ proteomic analysis accompanied by the generation of molecular images of marker proteins in slide sections of glioblastoma multiforme from the human brain for diagnostic purposes, and comparing this to the classical approach of immunohistochemistry. This issue also presents an interesting revision of the use of proteomic bioinformatics tools as an opportunity to access Systems Biology from proteomic data. We hope you enjoy reading the material in this special issue of BrJAC.

Mario Sergio Palma is a Full Professor of Proteomics at the Dept. of Basic and Applied Biology, Institute of Biosciences of UNESP Rio Claro. He got his PhD in Biochemistry at the Faculty of Medicine of São Paulo University, in Ribeirão Preto, SP, Brazil. He had internship at the Suntory Institute of Bioorganic Chemistry, in Osaka, Japan. He is a specialist in mass spectrometry of biological macromolecules and has experience in the area of Biochemistry of Macromolecules, Protein and Peptide Chemistry, as well as Chemistry of Natural Products of animal origin. He has worked in the study of the structural and functional characterization of venomous arthropod toxins, in the metabolomic and proteomic analysis of glandular secretions of insects, in the study of the structural dynamics of proteins and peptides with spectroscopic techniques. Currently, he is a member of the Core Manager of the National Institute of Science and Technology in Immunology (INCT iii) and of the Board of Directors of the Brazilian Society of Mass Spectrometry. He is currently Editor of Protein and Peptide Chemistry for the journal AMINO ACIDS (Vienna); Editor Reviewer of Frontiers in Plant Science - Section of Metabolism and Chemodiversity; member of the editorial board of the journals: Scientific Report, Peptides, Amino Acids, Toxins Reviews and Open Journal of Spectroscopy, and Toxins Reviews. He supervised 17 master’s dissertations, 19 doctoral theses, 16 postdoctoral internships, and has about 300 published works.

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 3-10
doi: 10.30744/brjac.2179-3425.interview.gdomont

BrJAC

INTERVIEW



Back to the Future with Proteins and Proteomics

Gilberto B. Domont
Emeritus Professor
Proteomics Unit, Institute of Chemistry
Universidade Federal do Rio de Janeiro, UFRJ
Rio de Janeiro, RJ, Brazil

Gilberto B. Domont has a BSc in Chemistry from the University of Brazil and a Ph.D. in Biochemistry from the Universidade Federal do Rio de Janeiro. He did graduate studies at the University of Southern California and Ohio State University and had internships at Technicon Corporation, Applied Biosystems, Institut Pasteur, Lyon, and the W.M. Keck Foundation Biotechnology Research Laboratory, Department of Biochemistry, Yale University. He holds the title of Professor Emeritus of UFRJ. As a member of the Brazilian Academy of Sciences, Dr. Domont is a founding father and ex-President of the Brazilian Society of Biochemistry and Molecular Biology and the Brazilian Society on Toxinology. He was Associate Editor of the Journal of Proteome Research and is currently on the Editorial Board of the Journal of Proteomics. Dr. Domont founded the Brazilian Proteomics Society. He is PI of the Chromosome 15 – Centric Human Proteome Project – Biology/ Diseases and a member of its Executive Committee. Heads the Proteomics Unit of the Institute of Chemistry, UFRJ. Is associated with the European Cancer Moonshot Project at the Dept. of Biomedical Engineering, Div. Clinical Protein Science & Imaging at the Biomedical Center, Lund University, dedicated to the study of the human proteome in cancer. His research interests are also centered on human viral infections, as well as plant, venom, and microorganism proteomics. He is devoted to the application of proteomics techniques to biological systems. His two favorite biological systems are Sonia and Solange, his daughter and wife.

How was your scientific initiation?

In 1956, I was a junior undergraduate enrolled in the Chemistry program at Faculdade Nacional de Filosofia, Universidade do Brasil. Of course, as often happens with so many, I was hooked by Dr. João C. Perrone's lectures on Biochemistry. In the middle of the year, he invited me to join his Protein Chemistry Laboratory at the Instituto Nacional de Tecnologia, granted with a Scientific Initiation scholarship from CNPq. There I lived for 13 years.

Which were your scientific influences to become a scientist? Did you have some teacher that gave you the necessary input to your scientific career?

The scientific atmosphere in Perrone's lab was creative, enthusiastic, critical, and highly scientific. Very limited money was available for research, so one had to have strong commitments to pursue basic science

Cite: Domont, G. B. **Back to the Future with Proteins and Proteomics.** *Braz. J. Anal. Chem.,* 2020, *7* (29), pp 3-10. doi: http://dx.doi.org/10.30744/brjac.2179-3425.interview.gdomont

and confront difficulties. Perrone's expertise helped us to survive building electrophoresis apparatuses and a gas chromatograph, potentiometer, colorimeter, fraction collector, and amino acid analyzer. Dr. Abrahão Iachan, an industrial chemist, a chemical engineer, and a staff lab senior researcher, contributed with the synthesis of enzyme substrates and protein chemistry reagents. From both of them I learned how to do science and chemistry as well as how to practice their most cherished ethical values.

Why did you decide to be a chemist? How Chemistry motivated you to become a proteomist?

At the age of 15, I was already doing simple chemistry experiments. I liked to use *picpoc* magic to scare my friends. *Picpoc* was made of two kinds of powders: all you had to do was drop a small amount of each powder, one on the thumb and the other on the indicator, and then pop both fingers to hear the sound of an explosion. Fantastic! Of course, as a teenager scientist, I wanted to increase the noise of the explosion to scare my friends even more. The obvious thing to do was to mix both powders; in my brain, having bigger amounts on each finger would provoke a louder explosion! However, what I did not figure out was that when mixing the contents of the two *picpoc* boxes that I had poured over a newspaper page the same explosion would result in friction during mixing. And so, it happened! The consequences were a burned face, eyes, and hair, two months of blindness, and eye surgery. We know a scientist never gives up. I had to understand what had gone on.

"By memory, I recall Fred and Bill chatting on Fenn's MS electron spray experiments. The discussion was about the consequences of Fenn's electrospray technique to scientific research. No one had any idea where it would lead, but everyone was sure that it would revolutionize research in the health sciences."

I was trained and did classical protein chemistry until 1989/1990 when I spent the winter months working at the Yale School of Medicine with Kathy Stone and Kenneth Williams, founder of the W.M. Keck Foundation Biotechnology Research Laboratory, Department of Biochemistry, now The Keck Biotechnology Resource Laboratory. I was there to work on modern protein sequencing techniques. In the department, Fred Richards and William Konigsberg, as well as John Fenn at the Dept of Chemical Engineering, Yale University, were also there, who would drop by sometimes. By memory, I recall Fred and Bill chatting in the large school corridor on Fenn's mass spectrometry electron spray experiments. The discussion was about the consequences of Fenn's electrospray technique to scientific research. No one had any idea where it would lead or what could happen with the science that was done at the time, but everyone was sure that it would revolutionize research in the health sciences. I found the electrospray technique to be my calling. Years after, dinning a hot soup in the relaxing hotel restaurant after the promenade of the II ESPRIT and I EuPA Congress, Valencia, Spain, 2007, I told John how I was hooked by his ESI at Yale.

Did you realize in the early stages of your career that you would be involved with protein science the way you currently are?

Absolutely. As soon as I began working with proteins, I was trapped. Proteins became my partner; love at first sight. No divorce on sight.

What are the recent analytical contributions you consider fundamental for the success of proteomics abroad? Can you comment on your contributions to this field?

It is a long list. Mass spectrometers, analytical techniques, robots, and MS applications are booming. A few examples chosen at random are the following:

- In analytical instruments, contributions were made in robotics and mass spectrometers (MSs), especially for sample preparation and to increase speed, resolution, and sensitivity. Hybrid instruments play an advanced role, especially those that use different analyzers and in ion storage separation devices, giving back a higher resolution and mass accuracy, as well as increased identification of the number of peptides and, hence, proteins. Ion mobility (IMS) has added value for the ion separation capability of MS, especially when coupled to time-of-flight analyzers. TIMS (trapped IMS), PASEF (parallel-accumulation

serial fragmentation), FAIMS (field asymmetric IMS), and SLIM (structures for lossless ion manipulations) have provided other IMS methods for the fractionation of complex mixtures of ions, increasing the number of mass spectra of peptides and more protein identifications. Improvements were seen in the characterization of intact proteins and complexes to verify their structures, proteoforms, and modifications. Two methods to break the protein amide bond which is more robust than the peptide, are electron transfer dissociation (ETD) and ultraviolet photodissociation (UVPD) and they have been used for more efficient fragmentations that can also be achieved using surface-induced dissociation in which a greater amount of kinetic energy of the collision is directed to the ion complex.

- In analytical methodology, immunoprecipitation has powered great advances in the measurements of protein interactions by pull-down, as in the spatial proteomics approach. Techniques for the identification of more proteins by match-between runs or protein quantitation such as parallel reaction monitoring (PRM) or data-independent analysis (DIA) have combined shotgun and targeted proteomics. Other areas of intense research and method development are labeled techniques for protein quantitation, post-translational modifications, especially glyco- and phospho-peptides/proteins, glycomics, and lipidomics, as well as new software for protein searches and statistics. Clinical proteomics is another fast-developing field for testing metabolites, peptides, and proteins for diagnostics, prognostics, and therapeutics.

A scientist has different ways of contributing to society. In my early years in Perrone's lab doing classical protein chemistry, my contributions were centered on the development of automatic analytical methodologies, protein purity and purification, primary structure determination, toxin isolation and sequencing, and plant proteins for feeding. In 1962, jointly with Perrone and Panek, I founded the first graduate program in biological sciences in Brazil, named the Graduate Program in Biochemistry at the Institute of Chemistry, UFRJ. In the program, my protein lab was the embryo for the Proteomics Unit of the last almost 20 years During the proteomics era, the contribution of the group ranged from bioinformatics tools and method development to contributions in understanding the physiological molecular mechanisms of cancer, neurodegenerative diseases, and virus infection diseases. I have also dedicated much of my time to spread the proteomics gospel in talks, conferences, seminars, and international and national courses. As a founding member of SBBqBM, I knew the importance of disseminating proteomics to the Brazilian scientific community, and in many of its annual Congresses I had the chance, with the help of my colleagues, to bring the best possible foreign proteomics scientists and research subjects to a gathering of undergraduates and graduate students, post-docs, and junior and senior scientists.

You are one of the pioneers of proteomics in Latin America. Could you point out the difficulties you faced to increase this area in Brazil and about the creation of the Brazilian Society of Proteomics (BrProt)?

The first big drawback, as always happens in science, is money for research. The second is common to all groups: purchasing reagents and small replacements pieces, maintenance of equipment, administration problems and personnel, accounting, scholarships, grant time validation, etc. Proteomics is expensive, is big money research. MS high-accuracy, high-resolution and high-sensitivity instruments worth US 1million make impossible to renew science labs and parks around the country. Depending on the instrument capability, some problems cannot be even approached. Worse, the consequences are dramatic in terms of data collection and physiological analysis because we cannot do a deeper analysis of any biological sample. In other words, we cannot obtain all the data that more sophisticated instruments provide. The proteomics potential of rare samples will not be explored, and the non-collected data are lost forever together with important discoveries that could have been made. Finally, I want to leave words of hope. As scientists, we deal with the real world and must be realistic in life as well. I am pretty much optimistic when evaluating the course of science in Brazil. My many years of experience mentoring students and dealing with Brazilian science assure me that we have the best people studying, who are highly committed to doing science. By providing more scholarships, research money, and the freedom to pursue ideas, the science

done in Brazil will quickly expand in quality and number!

I did not find it difficulty at all to set up the Brazilian Proteomics Society because the scientific proteomics community fully responded immediately to the call. In a meeting that lasted three days we were able to show great science, discuss relevant themes, create by-laws, and vote and inaugurate elected officials. This is the Brazilian Proteomics Society at it best, our BrProt, and these are the commitments of the members.

Which are your current interests in protein research? You have published interesting scientific papers during your career in both proteomics and in toxinology (and other biochemistry subjects). Could you comment on which ones you consider the most significant papers? Why?

"The most important paper in my career was my first, which dealt with the Electrophoretic Heterogeneity of Trypsin [Perrone JC, Disitzer LV, and Domont GB., Nature 183: 605 (1959)]."

I am now dedicated to advance the Human Proteome Project (HPP) and to study neurodegenerative diseases, cancer, and COVID-19. I am a founding father of the HPP, that involves two projects: the Chromosome-Centric Human Proteome Project (C-HPP) to demonstrate the existence of at least one protein per one of the 19.773 human genes, as accepted by HUPO, and the Biology Disease Human Proteome Project (B/D-HPP), which provides a framework for the study of biology and diseases. Scientifically, our group is responsible to giving life to chromosome 15 proteins, that is for finding its missing proteins, those that have never been identified by mass spectrometry. Administratively, I am a member of the C-HPP Executive Committee and attend Council Meetings.

The most important paper in my career was my first, which dealt with the ***Electrophoretic Heterogeneity of Trypsin [Perrone JC, Disitzer LV, and Domont GB., Nature 183: 605 (1959)]***. Crystallization was considered a criterion of protein purity, and in this work we demonstrated that the widely known and used crystallized trypsin was not a pure enzyme. It is unforgettable that at 25 years old I was the senior author of a Nature manuscript, which was followed by my third one five years later ***(Iachan, A, et al., Fractionations of trypsin by paper electrophoresis. Nature 203: 43, 1964)***.

I should mention that another important work was not a published paper but a printed abstract ***(Silva, MH, et al., Studies on the amino acid sequence of crotamin, in IX International Congress of Biochemistry, Stockholm, in the Abstract Book, IUB, 1973, v.IX)***. We had worked out more than 95% of the amino acid sequence of crotamin, a neurotoxin isolated from *Crotalus durissus terrificus* venom, the first protein to be ***almost*** totally sequenced in Brazil. This work was a fantastic research experience because of the intellectual loneliness I lived in. I was alone; no one did protein sequencing at the time in Brazil, and I had no one to talk with and no one to exchange experiences with in terms of protein sequence techniques and rationales. My interlocutors were the journals; I talked to them every Wednesday at the National Institute of Technology Library, where I perused the recently arrived journal issues. I still remember inserting a hidden tiny pencil dot on the upper left corner to mark the ones I had read. We lost the publication race despite having chosen a niche; a small protein known only by Brazilians.

Natural inhibitors of snake venoms have always been a traditional field of research. Two groups pioneered these studies, setting the standards for understanding the mechanism of action of these inhibitors: Drs Haity Moussatché, Jonas Perales, Ana GC Neves-Ferreira, and Richard H Valente in the Laboratory of Toxinology, Fiocruz, RJ, and us in the Laboratory of Protein Chemistry, UFRJ. Some of these publications were summarized in three book chapters ***[Perales, J, et al., Are inhibitors of metalloproteinases, phospholipases A2 and myotoxins members of the innate immune system? In Andrè Menèz (Org) Perspectives on Toxinology, Wiley & Sons, Ltd, 2001; Neves-Ferreira AGC, et al., Natural Inhibitors: innate immunity to snake venoms. In Stephen P Mackessy. (Org.). Handbook of Venoms and Toxins of Reptiles, CRC Press, 2009; Neves-Ferreira, AGC, et al., Natural Inhibitors of Snake Venom Metallopeptidases, Springer Science, 2015]***.

Interesting results on the molecular mechanisms of neurodegenerative diseases were disclosed in recent years, such as Alzheimer's ***(Mendonça, CF, et al., Proteomic signatures of brain regions affected by tau pathology in early and late stages of Alzheimer's disease Neurobio. Dis. 130: 104509, 2019)*** and schizophrenia ***(Velásquez, E, et al., Synaptosomal proteome of the orbitofrontal cortex from schizophrenia patients using quantitative label-free and iTRAQ-based shotgun proteomics. J. Proteome Res. 16: 4481 2017*** **and** ***Velásquez, E, et al., Quantitative subcellular proteomics of the orbitofrontal cortex of schizophrenia patients. J. Proteome Res. 18: 4240, 2019)***. Through a memorandum of understanding with the University of Lund, the Proteomics Unit became part of the European Branch of the Cancer Moonshot Program, US. The first published manuscript was in melanoma ***[Sanchez, A, et al., Novel functional proteins coded by the human genome discovered in metastases of melanoma patients. Cell Biol Toxicol. 36:261–272 (2020)]***, an upgrade to reduce the number of missing proteins in the HPP.

A recently published multi-omic approach used by five groups to study congenital Zika syndrome ***(Aguiar, RS, et al., Molecular alterations in the extracellular matrix in the brain of newborns with congenital Zika syndrome, Sci. Signal. 13, eaay6736, 2020)*** reports proteomic data from postmortem brain samples of microcephalic stillborn, disclosing the depletion of collagen molecules and the molecular basis of ZIKV infection after vertical transmission.

Of course, we are beginning to work on proteomics and metabolomics of plasma of SARS-CoV-2 human COVID-19-infected subjects.

You are an active member of the Human Proteome Organization (HUPO), and you are very involved with some programs of this institution. Because of this, you know quite well the Brazilian community of chemists and biochemists involved with protein chemistry. Thus, how do you compare the level of qualification of Brazilian proteomists when compared to those from Europe and the USA?

Our students and researchers are the best; they are very well trained in proteomics techniques, including mass spectrometry. I have witnessed praises for the excellent scientific training of our MScs, PhDs, and post-docs who went to work abroad for short terms of for good as well as compliments from many top leading proteomics scientists. Our students and junior scientists have gained international respect and won national and international prizes and highly disputed international scholarships. Frequently, I am asked to recommend graduate students and post-docs for scholarships or staff positions in leading labs in the USA and Europe.

How do you compare the proteomics developed in Brazil with that developed in the USA, Europe, and Asia?

The difference resides in the research topics, which are more frontier-like outside the country, where modern techniques and instruments are used to deeper evaluate the molecular explanations of physiological phenomena as well as scientific structure and policy. This is the result of the social and governmental acceptance of the importance of science for health and social welfare. Brazilian officials and agencies continue to use decades-old science funding policies. The dichotomy of separately funding scholarships and underfunding research projects is untenable anymore. Projects should be the funding unit of grants. The budget of these grants must include all expenses needed to achieve the proposed aims, such as travel, other direct costs like lab supplies and services, equipment, scholarships, administration, etc.

"My entire scientific life was dedicated to having fun doing research with my students, implementing and developing protein chemistry and proteomics, science, teaching, mentoring, and helping students to begin and develop their national and international careers."

You paved the way for many young Brazilian proteomists to launch their careers. Could you comment about your relationship with young scientists?

This is a very kind question. My entire scientific life was dedicated to having fun doing research with my students, implementing and developing protein chemistry and proteomics, science, teaching, mentoring, and helping students to begin and develop their national and international careers. I am very much attached to the intellectual growth of the students. It is an extraordinary and gratifying experience to watch their dedication and to follow their struggle to accomplish and finish their work and become mature scientists. These are highlighted, for instance, in the creation of the Rio de Janeiro Proteomics Network in 2002, whose main operational ideas were to introduce and implement proteomics techniques as well as to congregate protein chemistry students and researchers around the new field.

I always keep my eyes level with my young colleagues; I consider undergraduate or graduate students my professional colleagues. All of us are on the same page, always. The friends I deal with daily in the Proteomics Unit, Department, and Institute are mostly under 40 years of age. It is a relationship based on respect, merit, acceptance of differences, and intellectual leadership. Strictly speaking, everyone is equally treated as a scientist; all are scientific fellows. My students teach and give me too much; I learn a lot from them and owe them a lot, including the privilege of intense collaborations, discussions, warm debates, and the exchange of ideas. Socially, we hang out for beers, barbecues, beaches, social gatherings, and, of course, for science, philosophy of science, and theory of knowledge discussions.

Please, comment on the relationship between BrProt and BrMass.

Initially, in a council meeting with Marcos Eberlin, BrMass leader, I proposed that we should combine the two societies, transforming them into a Mass Spectrometry and Proteomics Society. The idea was not accepted. Years later, in 2018, we met and agreed on organizing a joint meeting – the 4th BrProt / 7th BrMass Annual Congress. Because of the excellent experience and fantastic results, we joined ideals again to organize inside the IMSC2020, a new joint congress adventure.

The relationship between proteomists and mass spectrometrists is excellent because we all speak the same language and have a common basis, mutual human and scientific respect, and friendship. Ideas are freely discussed and adopted or rejected; lectures and symposia are suggested and accepted; and invited speakers are freely chosen. I hope this partnership will improve and will continue for a long time.

"I love science; I love my work; and I enjoy and love my colleagues and students, past and present. I am happy doing science; I cannot stop doing it. It is like breathing. It is a passion I cannot abandon and let expire."

You retired in 1995 and are probably more active today than at the time of your retirement; what is the secret to keeping yourself so motivated?

Absolutely, right! The secret is love. I love science; I love my work; and I enjoy and love my colleagues and students, past and present. I am happy doing science; I cannot stop doing it. It is like breathing. It is a passion I cannot abandon and let expire. Mentally, I never retired. It is easy to understand because I owe everything I have and have done in my life to the Universidade Federal do Rio de Janeiro, UFRJ, including my accomplishments, career, friends, happiness, kindness, patrimony, travels, intellectual activities, etc. During Cardoso's second presidential term, a social security reform was proposed and approved. Then, before its approval, I decided to close my official link as a public servant and changed my employment status. My colleagues proposed, and the Federal University of Rio de Janeiro was very kind in granting me, the title of Emeritus Professor, which, supposedly entitles me to stay and have access to labs and students. I have no plans to quit doing science and the only reason I can imagine doing so is if I was stealing someone else's position at UFRJ.

Could you comment on your impressions about the current crisis (COVID-19 pandemic, economic, political) in Brazilian science? What is your view for the future of Brazilian science? Is there also a window of opportunities created by the new challenges?

All themes – the COVID-19 pandemic chaos and death toll, the economic crisis, and political affairs – have a common explanation, at least for the debacle of the last two years: we are not a scientifically educated society. This, but not only this, explains chloroquine, vermifuges, and others. There is no better demonstration of the absence of reason, criticism, intelligence, or moral behavior than to have an army general such as the Ministry of Health replacing a physician, a Minister of Environment that praises the destruction of forests and rivers, or to have nominated the former incredible Minister of Education. This explains the leader. His ideas and acts tell us the whole story.

The demarcation criterium between science/knowledge (ἐπιστήμη *episteme*) and opinion (*δόξα, doxa*) is verification. To be scientific, it needs to be verified. If it is impossible to verify, then it is just an opinion. Never in my entire life have I witnessed a strange rational moment like this. Before, against facts, there were no arguments; now, against arguments, there are no facts. Shortly, this explains the COVID-19 chaos in the country, the sinking of the economy, and the political pandemonium in a pandemic.

Brazilian science is extraordinary and responded immediately to the Zika epidemics. The same is happening with COVID-19. Some examples are the improvement in health treatment adopted worldwide (heparin), the production of SARS-CoV-2 spike protein for serological tests, new methods for virus detection, and the development of vaccines. One actual issue in Brazil is the slow response of the government to problems. Scientists respond immediately, whereas the government response is pachydermic.

Brazilian science is at a crossroads. A plan for the sciences in the upcoming years is not on the agenda. We need a modern, technological science park. Challenges provide thousands of opportunities. Research opportunities in the application of mass spectrometry to problems in the health and environmental sciences are numerous and diverse, including viral diseases, mental health, environmental quality, immunization, health care, obesity, cardiovascular diseases, and so on. These challenges provide more and better positions for young scientists. We may face less funding for science, scholarships, and positions. These are the prospects for an economy that predates and lives in contingency with little or no money for science. The reaction of the scientific community to these processes is intense. The Brazilian Academy of Sciences, the Sociedade Brasileira para o Progresso da Ciência, other scientific societies, and individual scientists are fighting against and strongly objecting to the disassembly of Brazilian science and the discredit of leading scientists. History teaches us that science and scientists win the race.

Do you want to leave some scientific legacy? What is it?

I want to be acknowledged as a scientist who lived a professional life praising science and scientists with whom I exchanged ideas and ideals and discussed and shared scientific data and ethical values. Those were fantastic moments of joy. I was lucky to spend an entire life of happiness dedicated to science and scientists. It is worth it. I do recommend love and enthusiasm for science and stimuli to and from the students.

"To close, I recommend reflections on Bertold Brecht's words in his play Galileo Galilei: ***'I maintain that the only purpose of science is to ease the hardship of human existence'****. I fully and deeply agree!!!"*

Front row: Hugo Junqueira (Magno´s older son), Renata dos Santos (selphie). From left to right: Magno Junqueira, Solange Guimarães, Gilberto Domont, Rafael Melani, Gustavo Monnerat, Raquel de Farias, Isis Botelho, Vinicius Parracho, Domingos Melo, Mohab Andrade, Fabio Nogueira, Natalia Almeida. Center: Erika Velasquez, Yara Silva, Ana Jacob.
.

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 11-12
doi: 10.30744/brjac.2179-3425.point-of-view-mvsousa

BrJAC

POINT OF VIEW

Is Proteomics Possible Without Mass Spectrometry?

Marcelo Valle de Sousa
Laboratory of Protein Chemistry and Biochemistry, Department of Cell Biology, University of Brasilia 70910-900, Brasilia, DF, Brazil

Proteomics was already being performed for many years before the term was coined in 1994. For protein identification after separating proteomes by two-dimensional electrophoresis, we used to rely on N-terminal protein sequencing by chemical methods, since mass spectrometry was not practical for proteomics yet. However, successful identification of a few proteins could take weeks or even months. With the advent of mass spectrometers devoted to protein analysis in the ’90s, proteomics gained huge momentum. We continued to use two-dimensional electrophoresis to separate proteomes for a while but began using mass spectrometry to identify proteins through peptide mass fingerprinting. Later, liquid chromatography linked on-line to mass spectrometry became gradually more popular; therefore, two-dimensional electrophoresis is infrequently used nowadays, and mostly for specific applications. In current proteomic laboratories modern mass spectrometers with higher resolutions, accuracies, sensitivities, speeds, and throughputs allow for sequencing thousands of proteins in a few hours. We can categorize proteomics into two main modes, namely bottom-up and top-down, though there are currently several proteomic strategies and abundant protocols. Mass spectrometry is the central technique for all these approaches. However, one can be curious as to whether or not mass spectrometry is going to remain the primary technique for proteomics in the future. What else could substitute mass spectrometry as the dominant technique?

In 2002, I came across an interesting paper. Maybe that was not the first paper in the field, but it was the one that caught my attention immediately. Its title is “Macromolecular Architecture in Eukaryotic Cells Visualized by Cryoelectron Tomography” [1]. I could see beautiful images of the actin network in the cytoskeleton, membranes, and cytoplasmic macromolecular complexes obtained by electron microscopy of tomograms of *Dictyostelium* cells under cryogenic temperatures. The authors called the technique “cryoelectron tomography”. I then searched the literature for other papers and reviews using these keywords. Other insightful papers and reviews showed up. In the last 20 years, the development of high-resolution microscopy has been astounding. Cryoelectron microscopy has been increasingly used to resolve tertiary and quaternary protein structures. The Nobel Prize for Chemistry in 2017 was awarded to three scientists who developed cryoelectron microscopy for the high-resolution structure determination of biomolecules in solution (https://www.nobelprize.org/uploads/2018/06/advanced-chemistryprize2017-1.pdf). Cryoelectron microscopy now competes with crystallography and nuclear magnetic resonance as the most powerful and adequate technique for protein structural studies. Recently, proteins of the new SARS-CoV-2 coronavirus had their structures quickly resolved in record times by groups that have mastered cryoelectron microscopy.

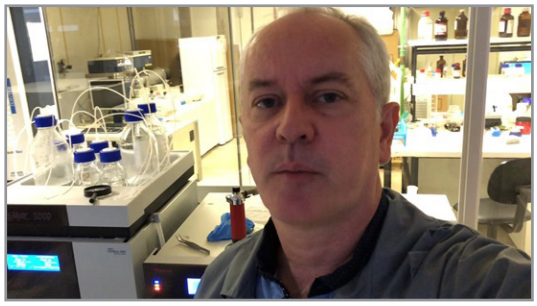
From my point of view, cryoelectron tomography will replace mass spectrometry as the principal technique for performing proteomics in the future. See yourself obtaining an image with all the proteome of a cell totally identified in a single shot. If you can have the 3D structures of proteins, you can also automatically get the protein identities. You will also have all the protein quantities, all the post-translational modifications, all the sequencing positions, all the protein subcellular locations, all the architecture of the complexes, all the 3D structures, all the molecular interactions. You will have everything!

Cite: Sousa, M. V. Is Proteomics Possible Without Mass Spectrometry? *Braz. J. Anal. Chem.,* 2020, *7* (29), pp 11-12. doi: https://dx.doi.org/10.30744/brjac.2179-3425.point-of-view-mvsousa

The major constraints against the rapid inclusion of cryoelectron tomography in modern proteomics are: 1) Cost of equipment. High-resolution electron microscopes are much more expensive than high-resolution mass spectrometers; 2) Technical difficulties in sample preparation. There are several accessory devices and technical details needed to prepare samples for cryoelectron microscopy experiments; 3) Lack of expertise. Proteomic scientists have mastered mass spectrometry over the years, but have not yet acquired intimacy with electron microscopy; 4) Bias against the unknown. Having not mastered cryoelectron microscopy yet, the proteomic community is impeded in going for it with greater confidence. However, needless to say that all the above obstacles are amenable to be surpassed, as mass spectrometry drawbacks were overcome by proteomic researchers in the past.

I am a molecular biologist (in the *lato sensu* of the term) who received training in protein chemistry and biochemistry from my first supervisor, Prof. Lauro Morhy. In the ’80s, he used to say that protein sciences would be extremely dependent on mass spectrometry, as he had used it for small molecules analysis in the ’60s. He was right. After completing my PhD in 1991 and coming back home, I quickly introduced mass spectrometry for the initial works in proteomics at the University of Brasilia in the early ’90s. Since then, I have been using mass spectrometry every day. I am a great fan of the fantastic mass spectrometry. However, I have also to admit that proteomics will be extremely dependent on electron microscopy in the future.

1. Medalia, O.; Weber, I.; Frangakis, A. S.; Nicastro, D.; Gerisch, G.; Baumeister, W. Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science*, **2002**, *298* (5596), pp 1209-1213 (http://dx.doi.org/10.1126/science.1076184).



Marcelo Valle de Sousa is a Full Professor at the Department of Cell Biology of the University of Brasilia, Brazil, where he leads the Laboratory of Protein Chemistry and Biochemistry and the Proteomics Center. Marcelo received his B.Sc. (1984) and his M.Sc. (1988) in Molecular Biology from the University of Brasilia, and his Ph.D. in Protein Biochemistry (1991) from the University of Durham, England. He was a Visiting Professor (2001-2002) at the Barnett Institute, Northeastern University, Boston, USA. His research interests are on proteomics applied to neurobiology using rats and bees as models and the biotechnology of amino acid-producing bacteria.

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 13-17
doi: 10.30744/brjac.2179-3425.letter-dmsouza

BrJAC

LETTER

Mass Spectrometry-Based Proteomics to Understand Schizophrenia

Daniel Martins-de-Souza[1,2,3,4]

[1]*Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas, Campinas, SP, Brazil*

[2]*Experimental Medicine Research Cluster (EMRC), University of Campinas, Campinas, 13083-862, SP, Brazil*

[3]*National Institute of Biomarkers in Neuropsychiatry, National Council for Scientific and Technological Development, São Paulo, SP, Brazil*

[4]*D'Or Institute for Research and Education (IDOR), São Paulo, SP, Brazil*

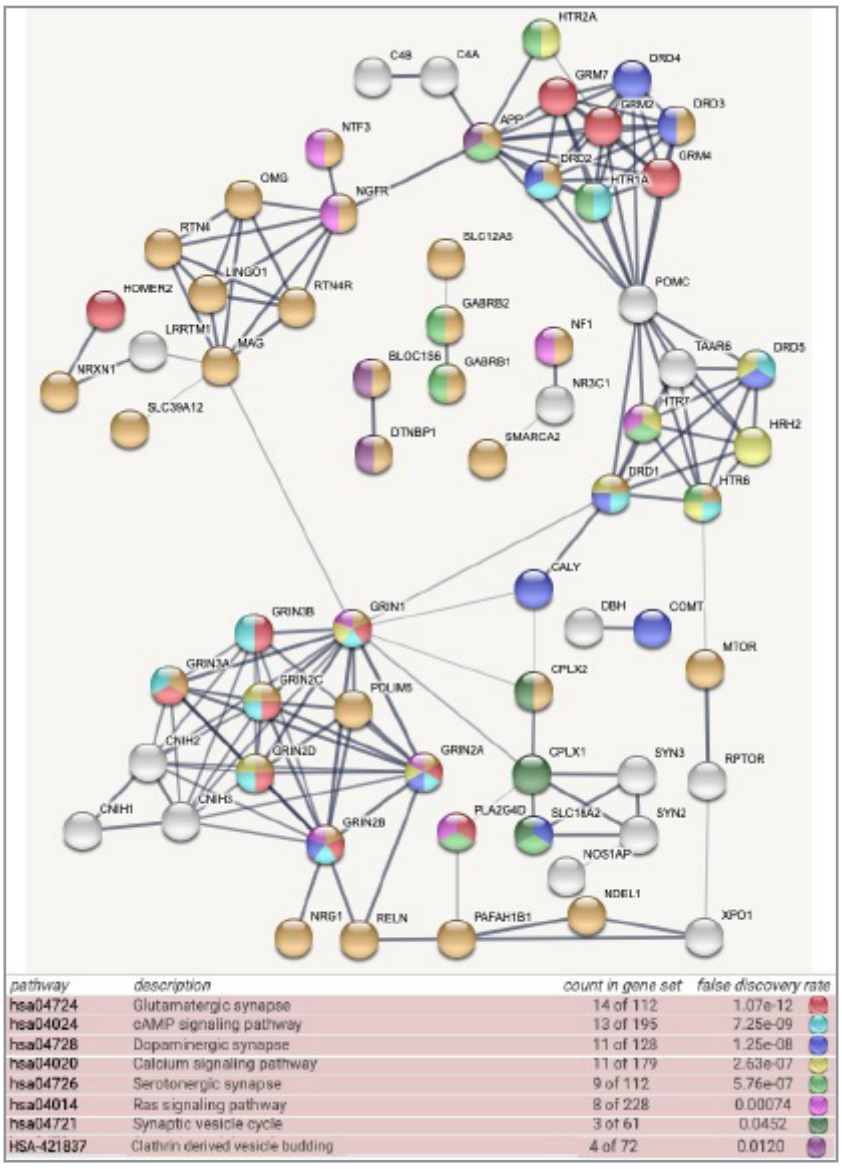
Psychiatric disorders are the most disabling disorders of humankind. From an economic and social point of view, a person affected by these diseases may lose about 30 years of their lives in an unproductive manner [1]. Although psychiatric patients have been diagnosed and treated for more than a century, these diseases are incurable, and the current medications only partially alleviate symptoms. These hurdles are mostly because we do not understand the biology behind these diseases. Understanding the most elementary molecular processes involved in the development and establishment of these multifactorial diseases is mandatory for better treating them, which will in turn improve patients' lives.

Proteomics emerged in the post genomic era as an alternative toolbox that, by definition, is appropriate to study multifactorial diseases such as psychiatric disorders. In the case of schizophrenia, we can easily see that proteomics has opened roads that have been traveled by scientists in the last decade.

Around 20 years ago, the first proteomic study approaching schizophrenia was published, combining two-dimensional gel electrophoresis and mass spectrometry [2]. From then on, proteomic methods have evolved towards automated liquid chromatography coupled to high-resolution tandem mass spectrometry (LC-MS/MS) and its derivations in qualitative and quantitative terms. The whole proteomic toolbox for LC-MS/MS-based shotgun quantitative proteomics has been employed to decipher the pathobiology of schizophrenia from the molecular point of view: 1DLC or 2DLC in terms of liquid chromatography (offline and online); data-dependent analysis (DDA) and data-independent analysis (DIA) — even including ion mobility — in terms of mass spectrometry [3]. In quantitative terms, different stable isotope labeling techniques - e.g., Isotope-Coded Protein Labeling (ICPL), Isobaric tag for relative and absolute quantitation (iTRAQ) - and label-free approaches (spectral counting and MS^E) have also been used [4]. Finally, targeted proteomics (selected reaction monitoring, SRM) have also been employed [5,6]. These technologies were employed mostly to study postmortem brains and blood plasma or serum. However, other human organs such as the skin and liver were also explored as well as several *in vivo* and *in vitro* models.

While studying the schizophrenia brains, the most consistent differences observed in the proteomes were those associated with energy metabolism, myelination, cytoskeleton assembly, alternative splicing (mRNA processing), and synaptic transmission. Differences associated with synaptic deficits have largely been documented, not only by proteomic analysis but also by large genomic studies. Some of the most commonly found genes or proteins associated to schizophrenia with synaptic function, according to the UniProt, are depicted in Figure 1.

Cite: Martins-de-Souza, D. Mass Spectrometry-Based Proteomics to Understand Schizophrenia. *Braz. J. Anal. Chem.,* 2020, *7* (29), pp 13-17. doi: http://dx.doi.org/10.30744/brjac.2179-3425.letter-dmsouza



pathway	description	count in gene set	false discovery rate
hsa04724	Glutamatergic synapse	14 of 112	1.07e-12
hsa04024	cAMP signaling pathway	13 of 195	7.25e-09
hsa04728	Dopaminergic synapse	11 of 128	1.25e-08
hsa04020	Calcium signaling pathway	11 of 179	2.63e-07
hsa04726	Serotonergic synapse	9 of 112	5.76e-07
hsa04014	Ras signaling pathway	8 of 228	0.00074
hsa04721	Synaptic vesicle cycle	3 of 61	0.0452
HSA-421837	Clathrin derived vesicle budding	4 of 72	0.0120

Figure 1. Genes and proteins involved in the synaptic dysfunction associated with schizophrenia, (color coded to the major biochemical pathways they are involved in). Dopamine, serotonin, and glutamate receptors are represented majorly. In silico analysis performed in STRING [14].

Interestingly, proteomic alterations observed in postmortem schizophrenia brains have also been observed in induced pluripotent stem cell (iPSC)-derived cerebral organoids generated from the cells of schizophrenia patients (not published). Since proteomics have highlighted the most altered biological processes in schizophrenia, these need to be studied more deeply.

Using *in vivo* and *in vitro* pre-clinical models, we were able to prove that the energy-metabolism-associated differences observed in schizophrenia brains are likely to happen in oligodendrocytes, which are the cells that produce myelin [7]. In Figure 2, based on data from the literature, we can see how the main energy metabolism proteins are associated with classical myelination markers. More specifically, our results indicate that glycolysis seems to be essential in this regard, which is also supported by the analysis conducted in Figure 2. Oligodendrocytes became one of the main topics of our studies since we believe that schizophrenia is not only a neuronal disease, as it has been treated so far, but it can also be a disease

centered on glia cells [8]. It is known that current antipsychotic medication mostly affects the function of neurons. We have also been investigating whether antipsychotics target oligodendrocytes [9]. By realizing that this does happen, we have also been investigating alternative treatments, such as those associated with the endocannabinoid system, as a means of better treating schizophrenia [10].

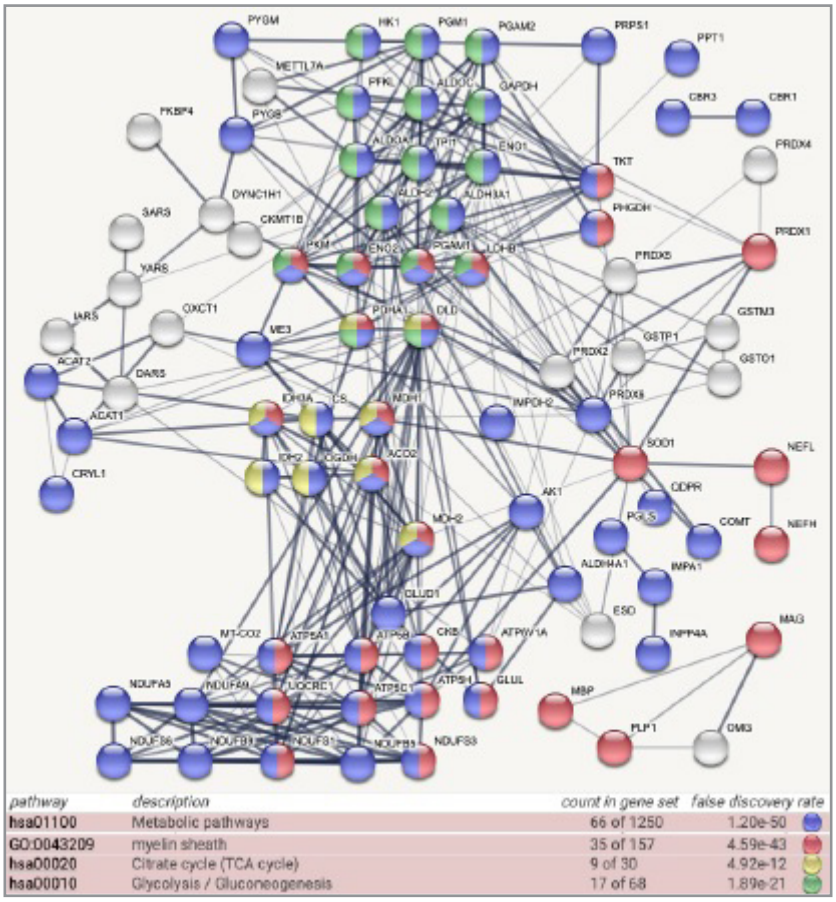


Figure 2. Differentially expressed proteins in schizophrenia samples associated with energy metabolism and myelination. Their strong connectivity shows their direct relation, as we demonstrated experimentally. The central role of glycolysis is also highlighted. In silico analysis perfomed in STRING [14].

In addition, more recently we have been searching for protein biomarkers that could predict an unsuccessful response to antipsychotics in the blood serum or plasma of schizophrenia patients. This is important because almost half of schizophrenia patients do not respond properly to the first round of medication. When medication does not function properly, the disease severity increases, and patients never recover their full brain performance. Moreover, given the side effects of antipsychotic medications, there has been a significant drop in medication usage by patients, which can only worsen the symptoms and, therefore, the mental health of the patients. Thus far, we were able to generate panels of lipid [11] and protein biomarkers [12], which may be implemented as biochemical tests for the prediction of a successful drug response. Proteomic signatures have also been used to build a diagnostic test for schizophrenia [13], which was even commercialized, but later discontinued.

In the last two decades, proteomics has added significant value to the understanding of schizophrenia, which is also true for the other psychiatric disorders. These investigations will end up improving the lives of patients since the next generation of medication may be based on the molecular underpinnings associated

with the disease. Eventually, we may also have biomarker tests in the future for better diagnosis and treatment outcomes from proteomic investigations.

Acknowledgments

We dedicate our work to psychiatric patients and their families. Our research was funded by FAPESP (São Paulo Research Foundation), grant 2017/25588-1, and CNPq (The Brazilian National Council for Scientific and Technological Development).

REFERENCES

1. Insel, T. **2013**. https://www.youtube.com/watch?v=u4m65sbqbhY
2. Johnston-Wilson, N. L.; Sims, C. D.; Hofmann, J. P.; Anderson, L.; Shore, A. D.; Torrey, E. F.; Yolken, R. H. *Mol Psychiatry*, **2000**, *5* (2), pp 142-149 (https://doi.org/10.1038/sj.mp.4000696).
3. Souza G. H.; Guest, P. C.; Martins-de-Souza, D. *Methods Mol Biol*., **2017**, *1546*, pp 57-73 (https://doi.org/10.1007/978-1-4939-6730-8_4).
4. Smith, B. J.; Martins-de-Souza, D.; Fioramonte, M. *Methods Mol Biol*., **2019**, *1916*, pp 3-39 (https://doi.org/10.1007/978-1-4939-8994-2_1).
5. Martins-de-Souza, D.; Alsaif, M.; Ernst, A.; Harris, L. W.; Aerts, N.; Lenaerts, I.; Peeters, P. J.; Amess, B.; Rahmoune, H.; Bahn, S.; Guest, P. C. *BMC Res Notes*, **2012**, *5* (146) (https://doi.org/10.1186/1756-0500-5-146).
6. Velásquez, E.; Martins-de-Souza, D.; Velásquez, I.; Carneiro, G. R. A.; Schmitt, A.; Falkai, P.; Domont, G. B.; Nogueira, F. C. S. *J Proteome Res*., **2019**, *18* (12) pp 4240-4253 (https://doi.org/10.1021/acs.jproteome.9b00398)
7. Guest, P. C.; Iwata, K.; Kato, T. A.; Steiner, J.; Schmitt, A.; Turck, C. W.; Martins-de-Souza, D. *Front Cell Neurosci*., **2015**, *9*, Article 180 (https://doi.org/10.3389/fncel.2015.00180).
8. Gouvêa-Junqueira, D.; Falvella, A. C. B.; Antunes, A. S. L. M.; Seabra, G.; Brandão-Teles, C.; Martins-de-Souza, D.; Crunfli, F. *Front. Psychiatry*, **2020**, *11*, Article 379 (https://doi.org/10.3389/fpsyt.2020.00379).
9. Brandão-Teles, C.; de Almeida, V.; Cassoli, J. S.; Martins-de-Souza, D. *Front. Pharmacol*. **2019**, *10*, Article 186 (https://doi.org/10.3389/fphar.2019.00186).
10. de Almeida, V.; Martins-de-Souza, D. *Eur Arch Psychiatry Clin Neurosci*., **2018**, *268* (7) pp 727-737 (https://doi.org/10.1007/s00406-018-0874-6).
11. Aquino, A.; Alexandrino, G. L.; Guest, P. C.; Augusto, F.; Gomes, A. F.; Murgu, M.; Steiner, J.; Martins-de-Souza, D. *Front. Psychiatry*., **2018**, *9*, Article 209 (https://doi.org/10.3389/fpsyt.2018.00209).
12. Garcia-Rosa, S.; Carvalho, B. S.; Guest, P. C.; Steiner, J.; Martins-de-Souza, D. *J Proteomics*, **2020**, *224*, 103813 (https://doi.org/10.1016/j.jprot.2020.103813).
13. Schwarz, E.; Guest, P. C.; Rahmoune, H.; Harris, L. W.; Wang, L.; Leweke, F. M.; Rothermundt, M.; Bogerts, B.; Koethe, D.; Kranaster, L.; et al. *Mol Psychiatry*., **2012**, *17* (5) pp 494-502 (https://doi.org/10.1038/mp.2011.42).
14. Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; et al. *Nucleic Acids Res*., **2019**, *47* (D1) D607-D613 (https://doi.org/10.1093/nar/gky1131).

Daniel Martins-de-Souza is Associate Professor of Biochemistry, Advisor to the Vice-President for Research at the University of Campinas (Unicamp), and one of the Coordinators of the Biology Area at FAPESP (São Paulo Research Foundation). Daniel founded the Laboratory of Neuroproteomics in 2014 right after having returned from Germany after a period of two years as Principal Investigator at LMU, Ludwig Maximilians Universität Muenchen (2012–2014). His expertise is mainly in mass spectrometry-based proteomics (shotgun proteomics) and systems biology *in silico*, employed in clinical and pre-clinical models for psychiatric disorders. Daniel is a Biologist (2003) with a PhD in Biochemistry (2008), which was obtained at Unicamp. His postdoctoral experience was at the Max Planck Institute of Psychiatry in Germany (2008–2010) and at the University of Cambridge in the United Kingdom (2010–2012), where he also led the Mass Spectrometry Unit at the Cambridge Center for Neuropsychiatric Research (CCNR). At the same time, Daniel was a consultant for Psynova Neurotech Ltd. Daniel is an Affiliated Member of the Brazilian Academy of Sciences (2017–2021) and an Affiliated Member of the Academy of Sciences of the State of São Paulo (2019–2023). He is also a founding member of BrProt (Brazilian Society of Proteomics) and a board member of BrMass (Brazilian Society of Mass Spectrometry). He was elected to the board of the Human Proteome Organization (2015–2017) and has been a member of the Human Brain Proteome Project Steering Committee since 2015. Daniel is one of the Associate Editors of npj Schizophrenia (Nature), a member of the editorial board of 7 other scientific journals, and a permanent editor of a book series entitled “Proteomics, Metabolomics, Interactomics and Systems Biology” by Springer-Nature. Thus far, he has given more than 60 oral presentations at international conferences.

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 18-44
doi: 10.30744/brjac.2179-3425.RV-03-2020

BrJAC

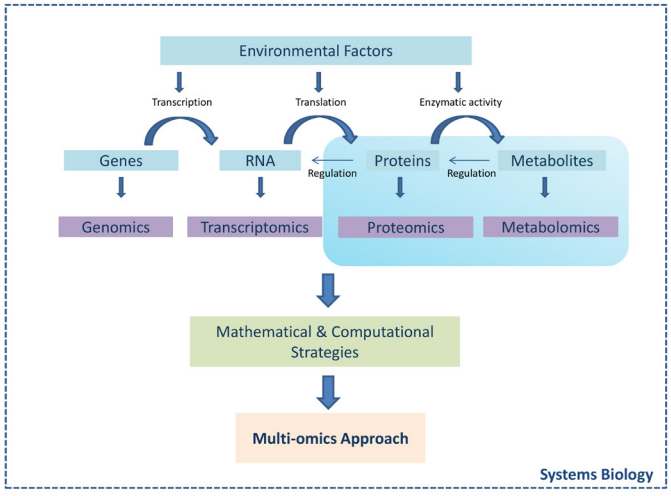
REVIEW

Multi-omics: An Opportunity to Dive into Systems Biology

Emerson Andrade Ferreira dos Santos[1] Elisa Castañeda Santa Cruz[1] Henrique Caracho Ribeiro[1] Luidy Darllan Barbosa[1] Flávia da Silva Zandonadi[1] Alessandra Sussulini[1,2,*]

[1]Laboratory of Bioanalytics and Integrated Omics (LaBIOmics), Department of Analytical Chemistry, Institute of Chemistry, University of Campinas (UNICAMP), P.O. Box 6154, 13083-970, Campinas, SP, Brazil

[2]National Institute of Science and Technology for Bioanalytics – INCTBio, Institute of Chemistry, University of Campinas (UNICAMP), P.O. Box 6154, 13083-970, Campinas, SP, Brazil



Omics data integration employing multi-omics approach is an outstanding opportunity to design a reliable picture of the biochemistry and dynamics of biological systems, as well as prioritize strategies for biomarker discovery. Biological functions are characterized by complex interaction networks, in which the dynamics of biomolecules manages physical and biochemical processes. However, since the emergence of omic sciences, researchers are still looking for the most accurate method to classify and determine the identity and function of biomarkers that describe a system and the ongoing biological processes. Thus, according to the strategies unveiled in the multi-omics literature, this is considered a challenging science field. Therefore, this review describes a workflow example regarding multi-omics data integration, indicating mathematical and computational tools in analysis pipelines that use various methods to perform a sequence of tasks, which would be able to describe biological processes within the systems biology context.

Keywords: multi-omics, data integration tools, omics

INTRODUCTION

The application of computational and mathematical modeling towards a deeper and broader understanding of biological systems is called systems biology. It proposes the study of biology through the use of holistic approaches, in opposition to reductionism, which focuses on the study of subsystems. Systems biology is highly dependent on the biological information acquired by molecular biology and/or omic strategies, which often provide a hypothesis that demand confirmation using a complementary reductionist approach. One important strategy to gain new insights and also assist in the experimental design is data integration [1]. Thus, computational biology can be used to accomplish two main tasks: (1)

Cite: dos Santos, E. A. F.; Santa Cruz, E. C.; Ribeiro, H. C.; Barbosa, L. D.; Zandonadi, F. S.; Sussulini, A. **Multi-omics: An Opportunity to Dive into Systems Biology**. *Braz. J. Anal. Chem.*, 2020, 7 (29), pp 18-44. doi: http://dx.doi.org/10.30744/brjac.2179-3425.RV-03-2020

Submitted 11 February 2020, Resubmitted 22 May 2020, 2nd time Resubmitted 26 June 2020, Accepted 29 June 2020, Available online 22 July 2020.

knowledge discovery, which is performed by the analysis of large amounts of experimental data in order to reveal unknown patterns that usually result in a hypothesis, and (2) simulation-based analysis, with the application of *in silico* experiments affording predictions to be further confirmed by experimental assays [2]. Within this context, the term multi-omics was proposed as a combination of methods to integrate data obtained from different omic approaches, aiming at gaining insight on how the different biomolecules (*e.g.*, proteins, RNAs, metabolites) are interconnected and how the flow of biological information occurs [3].

The term omics comes from the Latin suffix *ome*, which means mass or many [4]. The difference between omics and molecular biology approaches is, therefore, that the first one englobes a larger number of measurements per endpoint rather than one or a few. Despite the number of parameters measured per analysis is increased in omics, the number of replicates is decreased. In part, it happens because of the costs and time necessary for the experiments, and also due to the super estimation of methods, since there is a belief that more measurements would compensate a small number of samples [5]. Frequently, single-omic studies attempt to address specific biological issues without requiring a prior understanding of the biological bases involved [6]. However, experimental limitations, such as sample size (*e.g.*, rare samples), imperfect sequence identifications (*e.g.*, proteins) in databases, or representation of kinetic models from “static” data (*e.g.*, biochemical interactions) may generate gaps in the response of biological questions, which can be filled by multi-omics analysis.

After the advent of genomics, the scientific community has been trying to establish a correlation between the genotype and the phenotype in cells and living organisms [7]. Even with the development of strategies that provide information closer to the phenotype description, like transcriptomics [8], proteomics [9], and metabolomics [10], the individualized data provided by each one of these omics alone do not answer how the different biological processes are correlated [11] and how to explain this complexity. With the purpose of revealing these connections, an alternative is to integrate all (or most of) the available omics data. Through these connections, it is possible to provide complementary information from each omics strategy by the observation and understanding of how these relationships behave in a biological system, *i.e.* the study of genes and their products (RNAs, proteins, and metabolites) could provide a broader view of the modulations at genotype and phenotype levels of a system undergoing a specific biological process, such as a disease. Therefore, data integration from different omic sciences is a promising tool for the early detection of illnesses, as well as to study different treatments and their effects on patients, helping to choose the right medication, within the personalized medicine context [6].

Recently, the strategies in data integration were conducted from proteomic and metabolomic datasets, providing promising and significant results. In 2015, Del Boccio *et al.* [12] integrated these two omics to assess the differential pathways and networks between protein and metabolites in multiple sclerosis. In 2018, Cambiaghi *et al.* [13] employed a different approach integrating targeted metabolomics and proteomics data by using a correlation algorithm in order to observe the importance of circulating lipids and coagulation cascade in septic shock patients, evaluating the progression of the disease. Furthermore, in 2018, Gui *et al.* [14] described a disturbance in the phospholipid metabolism pathway in major depressive disorder and reported 74 differential proteins and 28 metabolites related to this specific metabolic pathway.

Although biology has always been a science of complex properties, in order to perform multi-omics data integration and ensure the data quality, some parameters (*e.g.*, list of genes, proteins, lipids, metabolites) need to be defined for subsequent data integration. For the study of a biological system, single-omic analyses are initially performed in order to identify sets of biomolecules, such as proteins and metabolites, which discriminate the evaluated conditions (*e.g.*, depression patients *vs.* healthy controls). Subsequently, these lists are submitted to multi-omics integration, which aims to reveal how the different types of biomolecules interact and are related to the phenotype. Multi-omics acquired the status of a new scientific area, partly because of the computational mathematics development for high-throughput data integration, while bioinformatics was previously used to treat data from genomics, transcriptomics, metabolomics, lipidomics, and other omic approaches separately. The complete workflow involved in a multi-omics study can be reached in two stages. The first is the acquisition of omics data and their subsequent treatment

by bioinformatics tools and the second is the integration of the parameters previously obtained from the isolated omic approaches by computational mathematics models.

It is clear how systems biology is strongly influenced by the data obtained from omic strategies and by how they are chemometric treated (bioinformatics), and mathematically modeled (multi-omics). If scientists want to improve their models and get more accurate results in systems biology, they first need to look at the difficulties found in the individual omic approaches, since they determine the parameters used in multi-omics analysis and carry possible errors in their results when integrating data. There is a myriad of different platforms for omics integration. However, some of them offer little support or do not possess a clear example dataset as a guide for new users. Moreover, review articles in this area focus on presenting the different applications for each platform without deepening in analyzing the interface and assessing their pros and cons. In this context, this review aims to describe different omics integration platforms for scientists who plan to start working on multi-omics with no experience in the area, focusing on the platform interfaces and their particularities, especially considering proteomics and metabolomics (other relevant platforms used in different multi-omic approaches were previously discussed by Pinu *et al*. [15], Misra *et al.* [16], and Fondi *et al.* [17]). Furthermore, the primordial steps of data preprocessing and pretreatment are addressed in this review, as well as the software used for this purpose and for chemometric analysis and current problems and pitfalls found in omic approaches, since good quality of omics data is indispensable for valuable results when integrated and correlated to the biology of the system.

SAMPLE COLLECTION AND PREPARATION

For omics clinical analysis, serum, saliva, urine and cerebrospinal fluid samples are generally used. Because omic approaches aim to have an insight in the entire ome, it is extremely important to warrant sample stability. There may be active enzymes in the biological specimens, which continue to present activity, thus degrading part of the biological content [18]. Hence, after collection, samples should be correctly stored until analysis. Blood needs to be left to coagulate, following centrifugation and separation of serum, and must be kept at -80 ºC [19]. In the case of urine, phosphate buffer should be added and then it can be stably stored at -25 ºC. For storage at 4 ºC, sodium azide should be added [20]. Saliva can be collected through expectoration in collecting tubes, with the help of a cotton swab or placing a cup over a particular salivary duct, followed by storage at -80 ºC [21,22]. Cerebrospinal fluid represents the most invasive sample type. It is obtained by lumbar puncture, and needs to be centrifuged for removing blood content before freezing at -80 ºC [23].

Samples should be analyzed as soon as possible. If it is not the case, they need to be properly stored and the cycles of freezing-thawing must be controlled, since biological samples are sensitive to environmental changes [24]. The samples may be used for multiple analyses. Biofluids such as blood serum and cerebrospinal fluid can be fractionated by liquid-liquid extraction in an organic and an aqueous layer and also in a protein pellet [25-27]. In saliva and urine (which lack an organic layer), the protein content may be separated from the aqueous layer by different techniques, such as protein precipitation with acetonitrile or acetone [28,29]. The extractions should be executed according to the protocol chosen amongst the multiple possibilities described in the literature, in which the amount of sample necessary varies from quantities such as 30 µL up to 20-50 mL, depending also on the sample type [20,22,25-27,29-31].

PREPROCESSING, PRETREATMENT AND STATISTICAL PLATFORMS

From proteomic and metabolomic approaches, a list of differentiating proteins and metabolites accompanied by their signal intensities are generated, respectively. These lists compose the information required to accomplish integration by characterizing the compounds that present different relative concentrations when different biological conditions are compared, indicating which metabolic pathways have divergent activity. Thereby, for relative quantification, the respective software calculates the fold changes for each biomolecule. For metabolomics, quantification is derived from signal intensities, and for proteomics, spectral counting. Consequently, protein and metabolite absolute quantification is

not a necessary step for multi-omic approaches. There is also neither a mandatory minimum number of biomolecules for the integration, nor the selection of specific metabolites and proteins (the important biomolecules are the ones that discriminate the compared biological conditions). However, the higher the number of identifications, the richer is the biological information obtained from multi-omics analysis.

In lists generated by proteomics, the influence of post-translational modifications (PTM) in proteins depends mainly on the steps involving specific sample preparation techniques and setting parameters for data processing and analysis, though it is not compulsory in a multi-omics approach. Obviously, for researchers interested in PTM analysis, it must be kept in mind that this type of information will be almost completely dissolved when observing the total proteome. In this sense, the necessity for PTM investigation will depend on the biological relevance and the techniques previously applied, as PTM enrichment and purification methods (for examples, see Supplemental Table I) added to the steps of data acquisition and processing downstream information.

Before starting data integration, the first step is to perform processing and treatment of the obtained information, and the second one is to determine the best approach for data integration, in order to guarantee a proper data quality analysis. There is a pipeline, in multiple stages, that is essential to follow when performing the data processing (Figure 1) [32]. In this review, some tools, approaches, and statistical methods from these two essential stages are discussed.

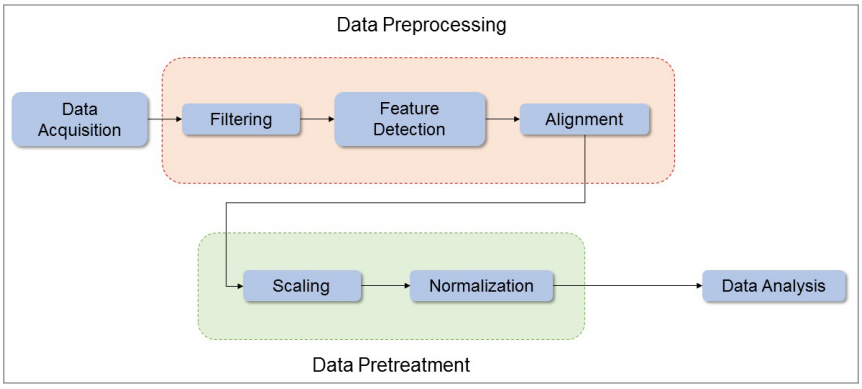


Figure 1. Workflow for the data analysis in multiple stages.

Preprocessing

In mass spectrometry analysis, usually, a specific software (*e.g.*, LabSolutions LCMS, Chromeleon, MassLynx, Analyst, etc.) is employed to extract and export the necessary information from the mass spectrometer to the computer. After exported, the information is mostly presented as files with an extension known as “raw file”. Table I presents the different mass spectrometry vendors’ file formats.

Table I. Raw file extensions from different mass spectrometry vendors

Company	Extension
Agilent/Bruker	.d, .YEP
Bruker	.BAF, .FID, .TDF
ABI/Sciex	.WIFF, .t2d
Thermo Scientific	.RAW
Waters	.PKL, .RAW
Shimadzu	.LCD

Depending on their formats, the files may be provided with different extensions. This is one of the bottlenecks on data integration: data are provided with different extensions or may appear as folders or even as files. In this way, some data types are unable to be opened by any other program than the one which exported them. In order to solve this issue, and redefine the extensions, free tools are available to convert them into the desired format, *e.g.,* MSConvert from ProteoWizard [33].

The MSConvert (http://proteowizard.sourceforge.net) is a simple tool to transform different data formats. Figure 2 presents one of the initial configurations to set up. When converting data, it is possible to apply different filters. Some of these filters and their descriptions are listed in Table II.

Table II. Description of some filters available on MSConvert. More information can be found on http://proteowizard.sourceforge.net/tools/filters.html

Filters	Description
Polarity	Keeps only the spectra with the selected polarity: positive or negative
msLevel	Keeps only spectra with the selected msLevel
scanNumber	Select spectra by scan number
Analyzer	Keeps only spectra with the selected analyzer
Analyzer Type	Filter by mass analyzer type

For Mascot users, the final data conversion format should be .cms1, .cms2, .ms1, .ms2, .text, .mz5, .mgf; while for XCMS (discussed later), they should be .mzML or .mzXML.

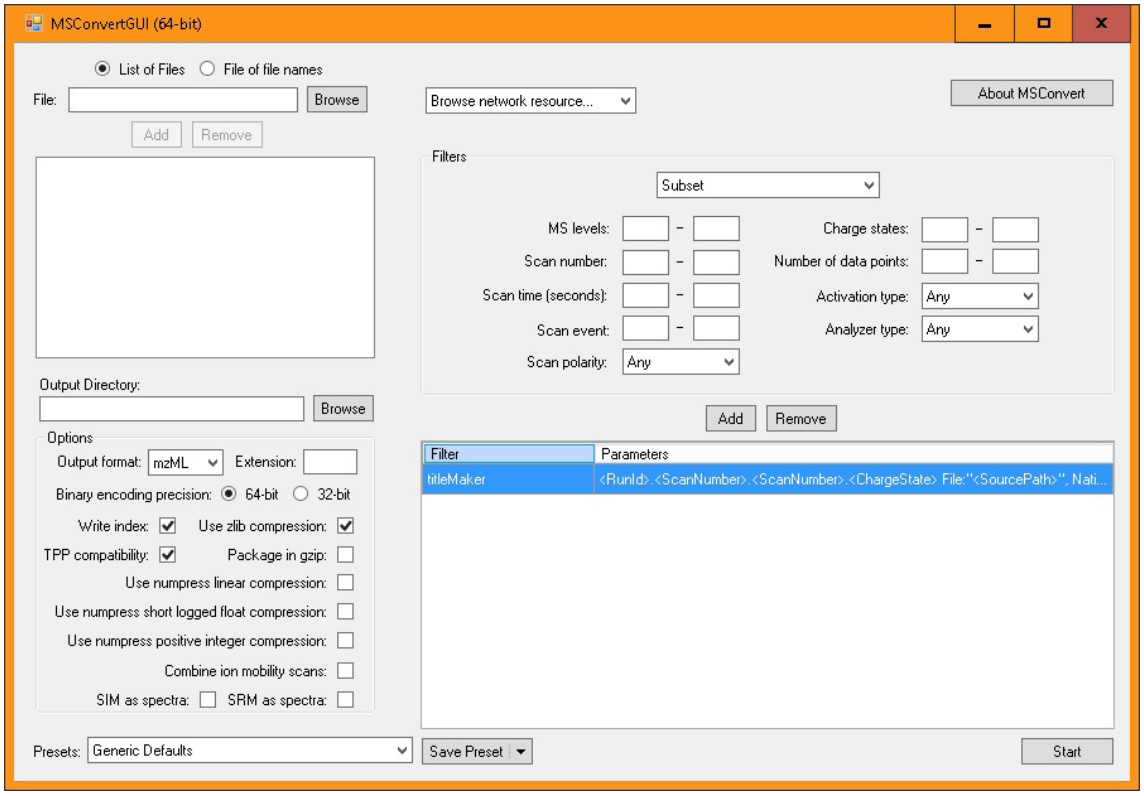


Figure 2. MSConvert interface for Windows.

Another necessary step is to remove the noise generated by uncontrollable variables (*e.g.*, instrumental fluctuations, analyst-related errors, or method inaccuracies), which can happen in any chemical analysis,

through application of filtering methods, which improve the signal-to-noise ratio (*i.e.*, the ratio between the intensity of the measured signal and that of interferences) [34]. Filtering is one of the most important stages of data preprocessing. However, inadequate use of filters can affect subsequent statistical analyses, generating false positives and results without biological value or end up discarding peaks lacking good quality, but which contains reliable biological information [35]. Therefore, it is common to use software to make it automatic and faster. One of them is called XCMS.

XCMS Online (https://xcmsonline.scripps.edu) is a software developed by the Siuzdak Lab at Scripps Research Institute (La Jolla, California, USA). It is used to perform statistical and identification analyses from liquid or gas chromatography coupled to mass spectrometry data for untargeted metabolomics - or targeted analyses by XCMS-MRM. This software performs, besides statistical analysis, chromatographic peak alignment, feature detection, and feature matching in the METLIN database. The software was initially developed for use in R software as a package called 'xcms' [36]. There are two additional help packages: IPO [37], which performs XCMS parameters optimization (*e.g.,* retention time - alignment of the deviation of the time that same compounds present in different samples when eluting from the chromatographic column -, and bandwidth - correction of the standard deviation of the Gaussian model for the shapes of peaks) [38] and CAMERA [39], available only for liquid chromatography coupled to mass spectrometry data, which performs extraction of compound spectra, annotate isotope and adduct peaks and propose the accurate compound mass in highly complex data. All packages are free to download at https://www.bioconductor.org [40].

For beginners, XCMS is used on the online version, providing some default filtering parameters based on the equipment used for analysis. On the other hand, in R software, an Integrated Development Environment (IDE) (discussed later), these parameters can be obtained and optimized by the IPO package [37]. This choice is beneficial for the ones who have programming language knowledge and join the preprocessing steps along the stages. As an additional step procedure, for the R software approach, there are filtering peaks and instrumental (*e.g.*, mass resolution - ability to distinguish two peaks of slightly different *m/z* ratios*,* which depends on the mass analyzer used -, and polarity) and statistical (*e.g.*, *p*-value) parameter settings. For data preprocessing, in this way, this software uses data from quality control (QC) samples. QC is a pool composed of the same amount from each sample used in the assay. Since the QC is analyzed together with the samples, the idea is that all the data from QC analyses do not present significant differences among the measurements, being reproducible [41]. Therefore, if that information does not vary, it means that regardless of the sequence and number of times the same sample was analyzed, the conclusion is "the method is robust and the dataset is ready for analysis".

After preprocessing, data is saved in tabular format files, which contain information on detected features (*e.g.*, retention time, *m/z*, number of peaks detected, etc.) and their intensity for each sample. Then, the data need to go through pretreatment steps so that multivariate chemometric methods can be applied to extract biological information.

Pretreatment

From the tabular format files, two types of information are obtained from previous treatment. The first is composed of desirable (*e.g.*, an important metabolite detected) and undesirable (*e.g.*, a feature with no biological information, such as solvent signal) information, and the second, comprising stochastic contributions. The latter is called noise, which are all uncontrollable variables. Therefore, the pretreatment step must be applied to samples or matrix variables to scaling, normalizing, and treating the missing values imputations.

Scaling

Scaling methods are methods applied to variables and are procedures that precede chemometric analyses. Scaling means that, depending on the method applied, some variables have higher or the same importance as others (*i.e.*, all variables are relevant for the study) [42]. Following are described some

available scaling methods for use according to the type of assay.

i. *Mean center scaling*. This method increases relevance for the most intense peaks rather than the smaller (closer to the noise) in order to diminish the error in multivariate analysis. When using this scale, there is a translation of the data towards the average value of each one.
ii. *Autoscaling*. This method is applied when the data matrix has different dimensions and make all variables contribute with the same importance. This approach makes the data dimensionless - independent of the unit.
iii. *Pareto scaling*. This method is very similar to the previous one. The difference is that instead of using the standard deviation, the square root of the standard deviation is used. This makes the scaling less impactful than the previous one.

The mean center scaling approach is commonly used in spectroscopy. Even though autoscaling is well used for metabolomics and proteomics, an alternative is Pareto scaling, which causes a decrease in the influence of noise and is sensitive to small concentrations of sample constituents [43].

Normalization

At this stage, the values of each variable in each sample are divided by a normalization factor, which can be the average or median of all samples for this variable, putting all samples on a predetermined scale and maintaining the qualitative information. The main objective of this step is to remove the systematic bias, which can be caused by the degradation of the sample components, variations in the amount of sample injected, measurement errors, among others, and this is verified by the QC sample [44,45].

Missing value input

When processing any data, the analyst must be prepared to face some missing data in his data matrix. These missing values can be the result of equipment malfunction, stochastic variations, or even a rigid preprocessing process that excludes the data. Another reason for this effect may be the concentration of compounds lower than the limit of quantification of the equipment. For dealing with this, some methods can be applied [46]:

i. *Feature exclusion*. The first and simplest approach is to remove the entire feature, which presents more than 20 % of its values as NaN, following the “80 % rule” [47].
ii. *LOQ Filling*. The second approach is to fill the missing values with the limit of quantification (LOQ) of the equipment [48]. Missing values, in this approach, can be replaced by 0, LOQ, LOQ/2, or LOQ/$\sqrt{2}$, the most commonly used [49].
iii. *kNN (k-Nearest Neighbors Imputation)*. This approach estimates the value of missing values when finding k samples that do not present missing values and are closest to that not determined. It is calculated based on a simple average or a weighted average [50].
iv. *SVD*. This makes use of principal components analysis or the decomposition by singular values (SVD). In general, this method assigns a value to the missing data and principal component analysis (PCA) is applied. With this, the main components that are most significant are selected and a new value is attributed to the missing data. After that, consecutive PCAs are performed until the value is small enough [51].
v. *Mean/Median*. Used to replace the missing values with the average of the non-missing values of other samples, or with the median of those values.

The choice of an appropriate pretreatment is essential for a successful chemometric analysis. For untargeted metabolomics, for example, it is recommended to use kNN for missing value imputation, according to Do *et al.*, as well described in a metanalysis study [52].

Statistical Platforms

After data preprocessing and pretreatment, statistical analyses are performed. Some of the most used platforms in metabolomics and proteomics will be further discussed.

MetaboAnalyst

One user-friendly free online platform for metabolomics, and nowadays for proteomics, is MetaboAnalyst (https://www.metaboanalyst.ca) [53], which contains different tools for analysis, interpretation, and integration of omics data. Figure 3 presents the start screen with all MetaboAnalyst tools.

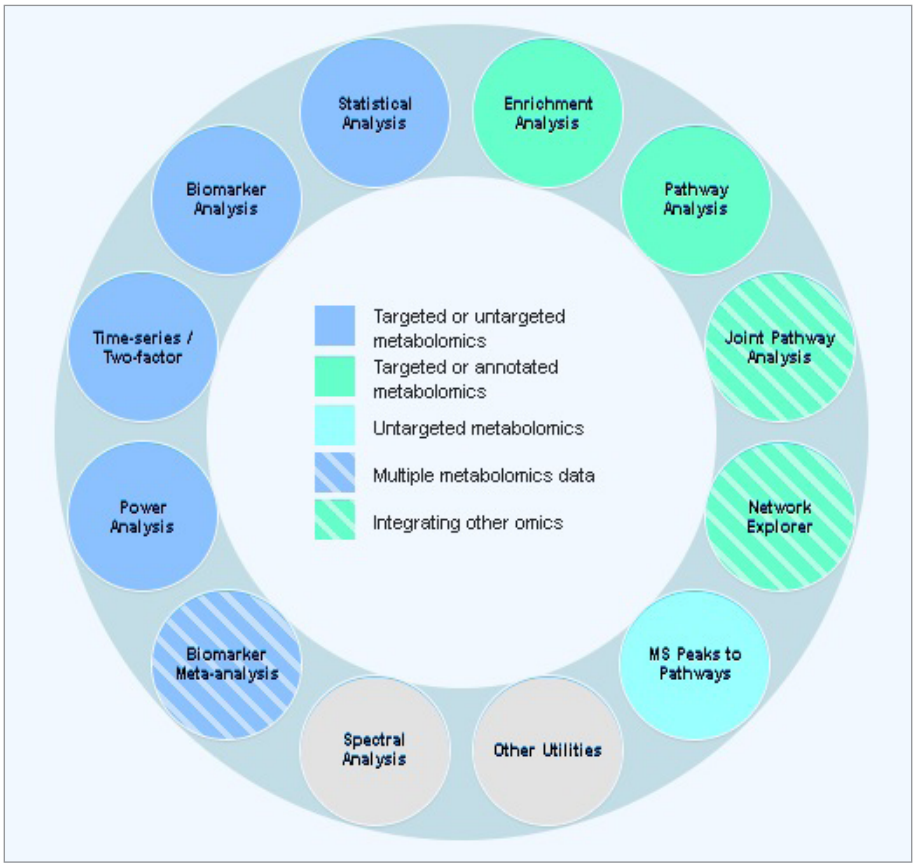


Figure 3. MetaboAnalyst modules screen showing the available modules and the main application of each one.

Following, some of these modules are detailed:

i. *Statistical Analysis*. This module has options for pretreatment such as normalization, scaling, missing value estimation, data filtering and data transformation. Besides, after the pretreatment step, this module can perform univariate (e.g., fold change, t-test, ANOVA), cluster (e.g., dendrogram, heatmaps) and chemometrics (e.g., principal component analysis - PCA -, and partial least squares discriminant analysis - PLS-DA) analysis, among others.

ii. *MS Peaks to Pathways*. This module performs a metabolic pathway enrichment analysis and visual exploration based on mummichog algorithm, with LC-MS spectral peak data. In the visual exploration, it is possible to know which metabolic pathways and which metabolites are significant. Human, mouse and zebrafish are some of the 21 organisms supported in this module.

iii. *Pathway Analysis*. This module performs a pathway analysis (enrichment and pathway topology analysis) for targeted metabolomics analysis. Here, 21 organism models are available (the same organisms from MS Peaks to Pathways module).

iv. *Network Explorer*. This module performs integration between metabolites (metabolomics) and genes (transcriptomics) or metagenomics data. Network Explorer allows the user to upload the data in a list format that is transformed in a visual network based on mummichog, showing their interconnections.

Figure 4. R command history screen of MetaboAnalyst showing some commands of the *Statistical Analysis* module.

For learning and exploring how these modules work, MetaboAnalyst offers sample datasets. Information on other modules can be found on the same page as the module selection panel. In addition to the online version, the R package (MetaboAnalystR) is also available [53]. For each task performed on the online platform, a box (on the right corner of the browser containing coding lines) is displayed, as demonstrated in Figure 4. These codes can be used on the MetaboAnalyst package in R software and describe what was done on the online platform.

Despite having similar functions to XCMS, the big difference between these two is that XCMS uses data in formats as .mzXML, .mzData, .mzData.XML, .netCDF, .cdf, .wiff, .wiff.scan, while MetaboAnalyst works with files already filtered and preprocessed in tabular formats (.txt or .csv). Although XCMS performs data preprocessing, it does not provide options for database queries and some pretreatment methods, while MetaboAnalyst offers options for pretreatment methods and database options. Hence, the idea is to complement the information using the best tools from each software. This can be done directly in RStudio, which has packages with the functions of XCMS and MetaboAnalyst in script format. For the use of all these methods, it is common to use statistical tools and software. The most common is the Integrated Development Environment (IDE). IDEs are software designed to gather in one place everything the developer needs, and often contain syntax checkers, task automakers, prompts, and more [54]. In this context, for omic sciences and statistics, two IDEs stand out: MATLAB [55] and RStudio [56].

MATLAB

MATLAB (MATrix LABoratory) is a powerful mathematical tool and IDE for developing mathematical methods. Created by Clever Moler, it is one of the most known worldwide and uses a C / C ++ derived language. MATLAB offers some expansions known as ToolBoxes that allow for higher expertise in some areas of method development. Although this software requires user knowledge of programming, it presents itself as a user-friendly platform, with clean layouts and functions named in an easy-to-remember way. More information can be found on the MathWorks website (https://www.mathworks.com).

RStudio

RStudio (https://rstudio.com) is a free IDE for R programming language and statistical graphics. Being an open-source IDE, it allows the community to participate in the development of new packages and methods. Although it has a straightforward layout, it does not present menus and buttons, requiring the development of lines of code to use functions. Thus, R users need a little more experience with coding. One of the significant advantages, besides its open-source, is that it is free. In general, this IDE has built-in console, syntax proofing and graph plotting tools, among other functions.

When starting RStudio for the first time, it presents itself with four main windows, in addition to the menus on the top bar. Figure 5 shows the RStudio initial screen. In the upper left window, it is possible to write the script of the functions to be used, depicting where the main work is carried out.

In the upper right window, one can see three tabs: environment, history and connections. The first

shows information about variables, functions, imported data frames and everything that need to be stored. The second shows the history of used functions, imported packages, etc. The third is where one can manage connections to data sources. In the lower left window, there are three tabs: the console, showing the results of processing and executing codes; the terminal, which is used to set up access to the system shell from within RStudio IDE; and Jobs, which manages the jobs RStudio is running. In the lower right window, one can see five tabs: Files, Plots, Packages, Help and Viewer. The 'Files' tab shows computer paths and can be used to import files and to set the Working Directory. The 'Plots' tab displays all plots made. 'Packages' shows installed packages and allows to import them and to install new ones. 'Help' shows RStudio documentation contents for packages and datasets when the help function (signalized as "?") is called followed by the package or the dataset's name (for example, ?ggplot2). 'Viewer' is used to view local web content.

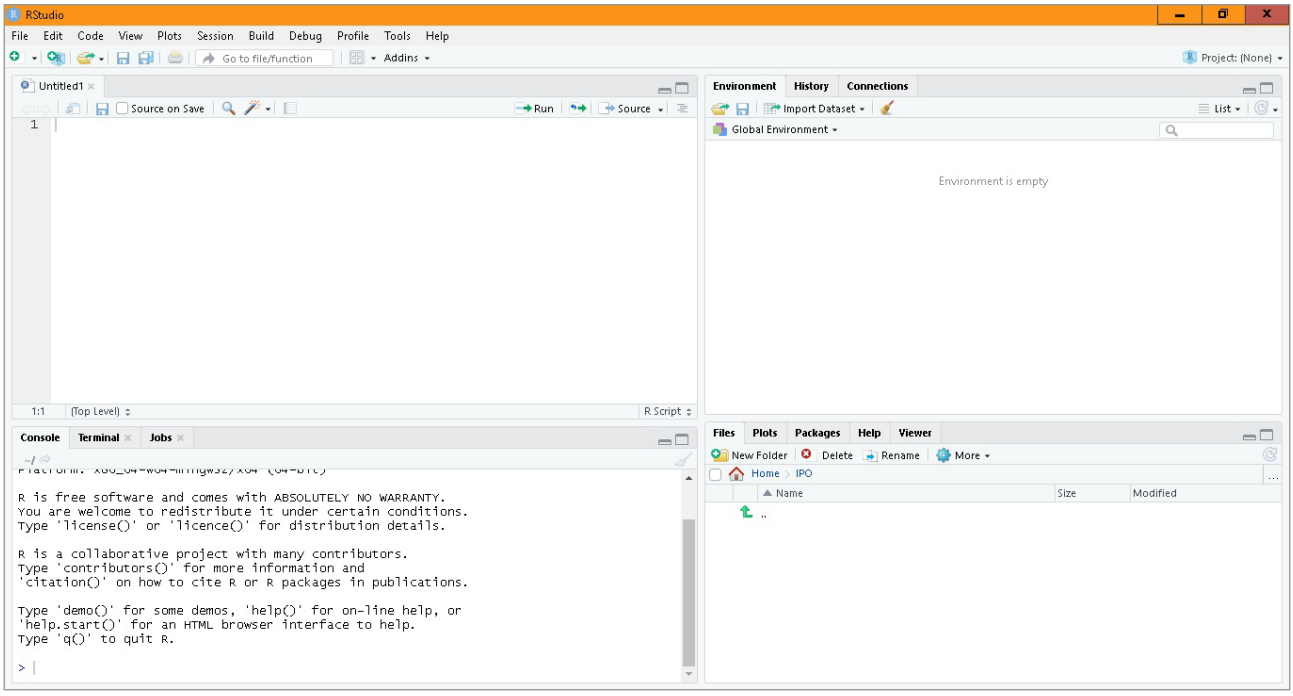


Figure 5. RStudio (version 1.2.5001) initial screen. RStudio layout can be customized according to the user's preferences. The 'View' menu assists in this customization.

The main differences between these two IDEs are the syntax of each program and that one is free (RStudio), while the other is paid (MATLAB).

Similar to MATLAB, which has toolboxes, RStudio has packages made by developers, which can be installed. These packages have plots, data handling, filters, escalation functions, among others. MATLAB also has packages for data preprocessing, accepting files of different formats such as .mzCDF, .mzXML, .JCAMP, among others. One of these packages can be accessed at the main address (https://www.mathworks.com/help/bioinfo/mass-spectrometry-and-bioanalytics.html) or at the most accessible data analysis host in the tool package: https://www.mathworks.com/help/bioinfo/ug/mass-spectrometry-data-analysis.html. Hence, the IDE chosen to perform the described procedures are the user's preference. The main points to be considered for this are the investment to acquire MATLAB and the flexibility of developing the methods in RStudio.

Unlike metabolomics, proteomics data preprocessing and pretreatment are performed using vendor software, which employ their algorithms during data acquisition. Since data processing algorithms are not fully documented and usually are restricted to one instrument platform, it limits the portability to other data

processing tools and comparison of results [57]. Furthermore, most software are available commercially. Following, open and free software for proteomics data analysis are described.

MaxQuant & Perseus

MaxQuant (https://www.maxquant.org) is a free software that analyzes shotgun proteomics data sets obtained by liquid chromatography coupled to mass spectrometry to identify and quantify peptides or proteins. It is well recognized due to its availability to improve the mass accuracy of peptide features through computational techniques [58]. Beyond analyzing data from labeling and label-free strategies, it has a default setup compatible with leading brands of equipment on the market such as Thermo Fisher Scientific, Bruker Daltonics, AB Sciex, and Agilent Technologies [59]. Perseus (https://www.maxquant.org/perseus/) is a platform that helps interpret the data obtained from MaxQuant. There are a variety of statistical tools, such as covering normalization, pattern recognition, multiple-hypothesis testing. Like MaxQuant, Perseus comes in a user-friendly format and is also free [60].

For beginners, it is recommended starting data analysis of large mass-spectrometric data sets following MaxQuant website recommendations (http://coxdocs.org/doku.php?id=maxquant:start). The software includes the Andromeda (peptide search engine based on probabilistic scoring), as well as the Viewer application for inspection of raw data, identification and quantification results. For statistical analysis of its output data, the Perseus platform can be used. Additionally to the installation support, there is important information on the webpage concerning to guiding users on how to process the data from input raw files (MaxQuant & Andromeda) to output processing data (Viewer & Perseus), as well as a link to a forum in Google Groups that discusses questions not found in the software documentation.

DATA INTEGRATION

The platforms discussed in this section apply different strategies to integrate proteomics and metabolomics datasets. Omics integration platforms present two different approaches: (1) chemometric analysis and (2) pathway integration plus visualization. One platform for each approach will be discussed: MixOmics [61] and OmicsNet [62], respectively.

OmicsNet

OmicsNet is a new omics integration platform built by the same group responsible for MetaboAnalyst [53] and has a well-explained and straightforward interface. The tutorials explain the functionalities step-by-step and are very useful for beginners (it can be accessed in the option of the tutorial at https://www.omicsnet.ca/). The input data for proteomics must be structured as a list of differential proteins accompanied by their log-transformed fold change (logFC), where the fold change is defined as how many times the expression has increased or decreased (expression level in condition 1 divided by expression level in condition 2, also known as a ratio). The logFC input is not mandatory; however, it can offer new insights about pathway regulation after the data is integrated. Data with zero in a specific condition (logFC = 0) mean that the fold change is equal to 1. When a specific protein/metabolite shows this value, we can affirm that this specific variable presented no difference (increase or decrease) between the two analyzed groups.

The platform accepts several data formats (Entrez, Ensembl, Uniprot, Official gene symbol ID), as well as ten different organisms. The same scheme applies to metabolomics, although using different databases (such as KEGG [63], PubChem, and HMDB [64]). Figure 6 presents the initial screen, with options to select a proteomics-metabolomics integration and the input options.

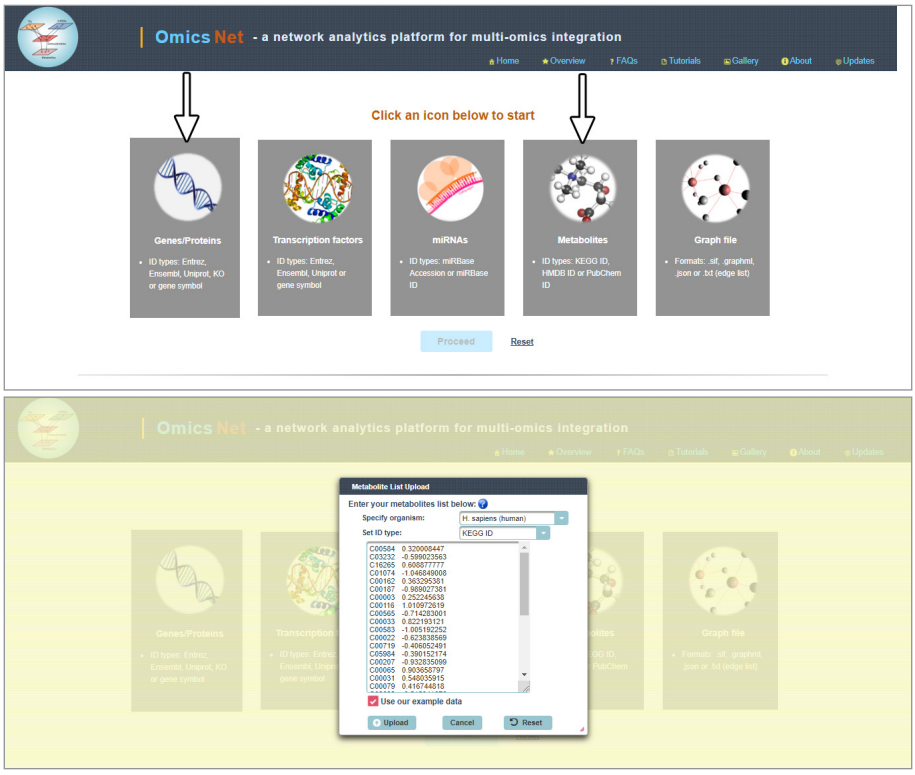


Figure 6. OmicsNet initial screen with the selections for a proteomics-metabolomics integration and input options [62].

After data inputting, the platform leads the user into a second screen, where the network details can be selected before its construction. Here, it is possible to select which interaction will be prioritized over the network construction. For proteomics and metabolomics data, there are two interaction options available: protein-protein interactions (PPI) and metabolite-protein interactions. OmicsNet tutorials suggest using metabolite-protein as primary interaction, as this procedure identifies which enzymes are interacting with the metabolites and the respective PPI [62], as demonstrated in Figure 7.

Depending on the input, the output network can be extensive, and sometimes confusing. The complexity can be reduced in the second screen by controlling the degree (number of connections among nodes) and the betweenness (the measure of the centrality in the graph based on the shortest paths going through the node). Besides, by reducing the network to a “minimum network”, the seeds and other essential non-seeds maintain the network working at minimum. Seeds are the inputted data on the first screen (list of differential proteins/metabolites), while non-seeds are all the other nodes that are used to establish connections between the seeds and, then, build a network with the results. This option is useful to analyze critical connections between both datasets but might reduce the identification of potential pathways in the “network viewer” (next step). After adjusting the settings and submitting the data, it proceeds forwards.

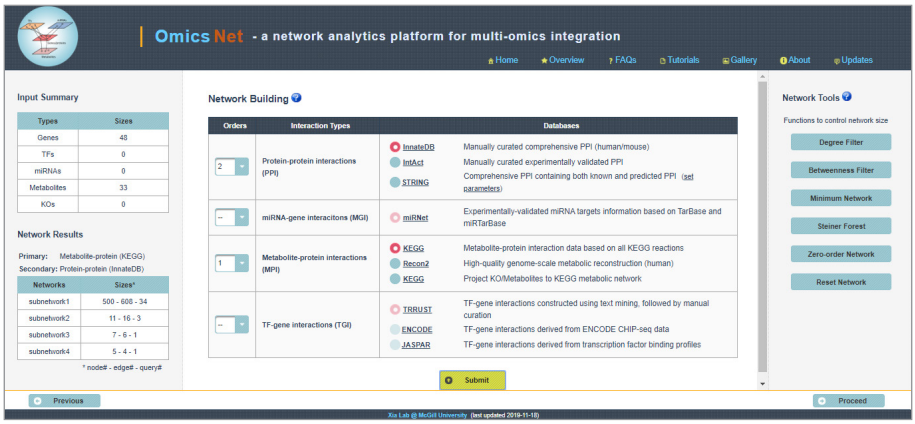


Figure 7. OmicsNet network building screen, with the building results on the left side, the building options in the center, and the network reducing tools on the right side [62].

Network viewer

In this stage, the visualization of the network previously built is possible and manageable. There are some aesthetic options, such as changing the color of the nodes, view options, layout, etc. Sometimes, after a network generation, it creates smaller subnetworks that are accessible in the upper left corner. In the left panel, the platform shows a table with all the current nodes, ranking them by their degree and expression levels using only the logFC input of the first step. Degree is the number of connections between nodes. Nodes with higher degree act as important hubs in the built network. Expression levels showed on the network viewer is a general term to address all the possible alterations of metabolites/proteins between groups (this value is only viewable if the user inputted the logFC in the first screen). The authors of the platform also address this optional second column on the input screen as abundance levels [65].

In the right panel, there are explorer modules and enrichment analysis, which shows the differential pathways in the data. Finally, the save option for the pathway enrichment analysis in a .csv format for further evaluations is possible, as well as downloading the built network for a presentation or an article. The network viewer screen is reproduced in Figure 8.

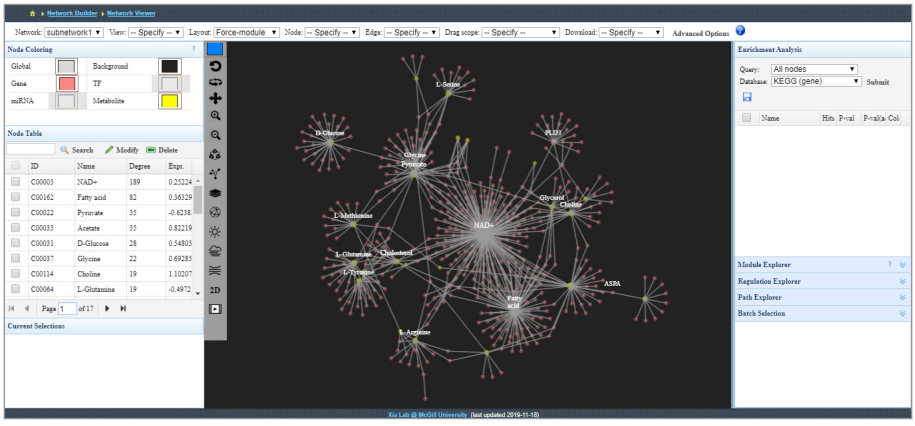


Figure 8. Network viewer screen showing the different modules, node table, and aesthetic options [62].

Regarding many journals that usually demand high-resolution images, one pitfall is the low-resolution images produced in comparison to MetaboAnalyst, which offers image exportation options (such as 300 dpi and 600 dpi in tiff format). Thus, researchers have to use other options to get an appropriate image. Providentially, the platform can export the network in .json format, which is accepted by other visual

systems biology platforms like Cytoscape [66]. Nevertheless, this format loses its expression information in the case of inputted logFC and has to be re-inputted manually.

MixOmics (R package)

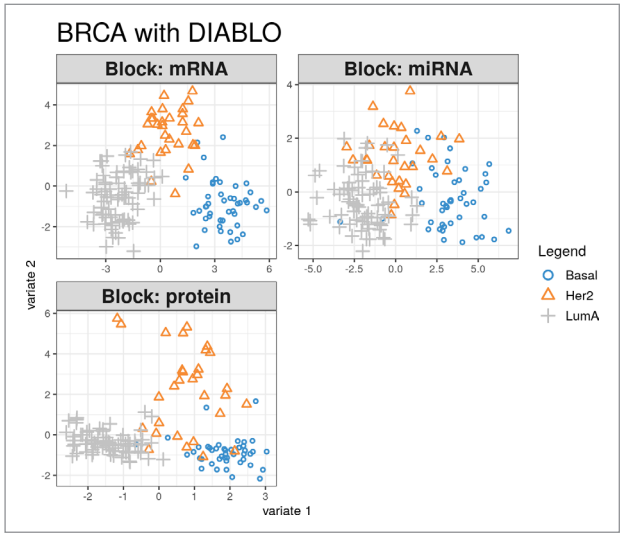
MixOmics [61] is an R package with various statistical and chemometric tools for omics data, which presents a focus on variable selection. The package is composed of nineteen different multivariate methodologies, such as PCA, PLS, PLS-DA, and sparse partial least squares discriminant analysis (sPLS-DA), among others. In this section, we will focus on a specific integration methodology from this package, named DIABLO [67].

DIABLO is an acronym for **D**ata **I**ntegration **A**nalysis for **B**iomarker discovery using **L**atent c**O**mponents. DIABLO method generalizes PLS for multiple matching datasets, and it is named as a N-Integration method by the authors who suggested the primordial analysis of the datasets to be integrated through sPLS-DA and PLS in order to evaluate the major sources of variation and guide the user through the integration processes. For the proper performance of DIABLO, the proteomics and metabolomics datasets must be in .csv or .tsv format to be uploaded into R. In this review, we used RStudio version 1.1.423 (R version 3.5.2), with MixOmics package version 6.10.6.

The datasets must be precisely composed of the same samples as the focus is analyzing the correlation between both omic approaches. After checking it, the next step is the proper analysis.

For the first step of this procedure, the package from the Bioconductor website needs to be downloaded (https://www.bioconductor.org/packages/release/bioc/html/mixOmics.html) and installed on RStudio. After installing and loading the package, the datasets must be also loaded as two different variables (samples on rows, variables on columns). When working with imported .csv files, the data, sometimes, might be in a different format (MixOmics accepts data frames with only numeric values) than the accepted by the package. HINT! If the data set is not being recognized, it is recommended to compare it with the provided data example by the toolkit using the command: `data <-breast.TCGA` and observe the differences between them.

After verifying if datasets are in the accepted format, it is necessary to load a vector Y within the R environment as a factor using “as.factor” before the vector name in order to set it in the correct format. Each of the omics used will be treated as a “block”. DIABLO also requires that the used datasets are structured as a list. Use the list command to build a new one with both datasets and then load this new list and the vector Y into the package using the function:




```
MyResult.diablo <- block.splsda(X,Y)
```

Then, with the `plotIndiv(MyResult.diablo)` function, it is possible to visualize the dispersion of each data, as shown in Figure 9.

With the `plotVar` function, it is possible to observe the correlation circle plots between the datasets, which shows the contribution of each variable to each correlation component [68], as observed in Figure 10.

Figure 9. Dispersion of each block using the `plotIndiv` function in DIABLO package [67].

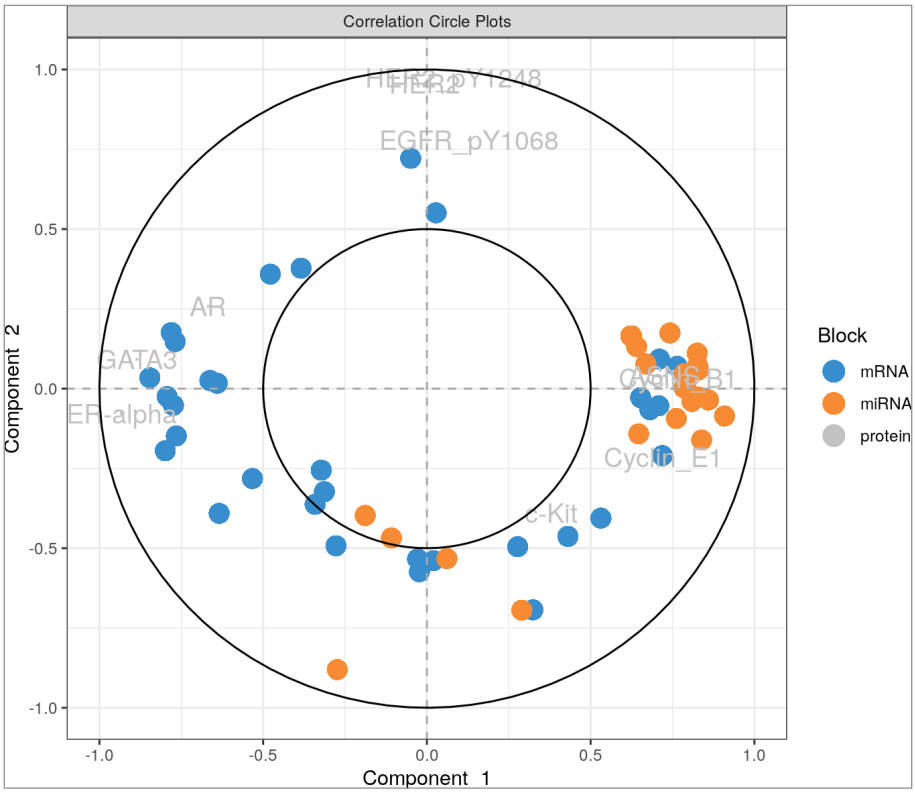


Figure 10. Correlation circle plots of inputted data plotted with `plotVar` function [68].

The `plotDiablo` function allows analyzing the correlation between inputted datasets, as the example in Figure 11.

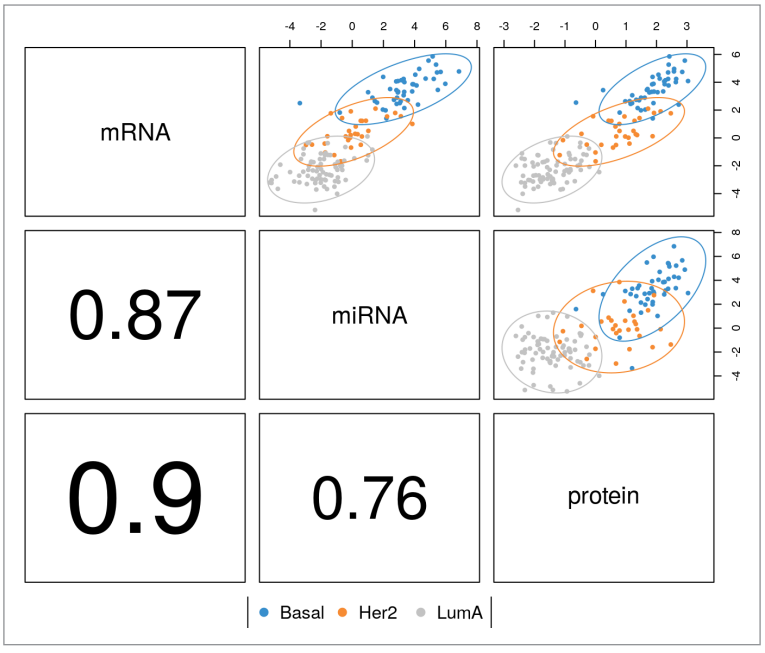


Figure 11. Correlation structure using the `plotDiablo` function [67].

The most useful function of this platform is undoubtedly the `circosPlot`. This plot represents the correlations between the variables represented on the side quadrants, and it is based on a similarity matrix extended to multiple datasets [62]. An example of the `circosPlot` function is presented in Figure 12.

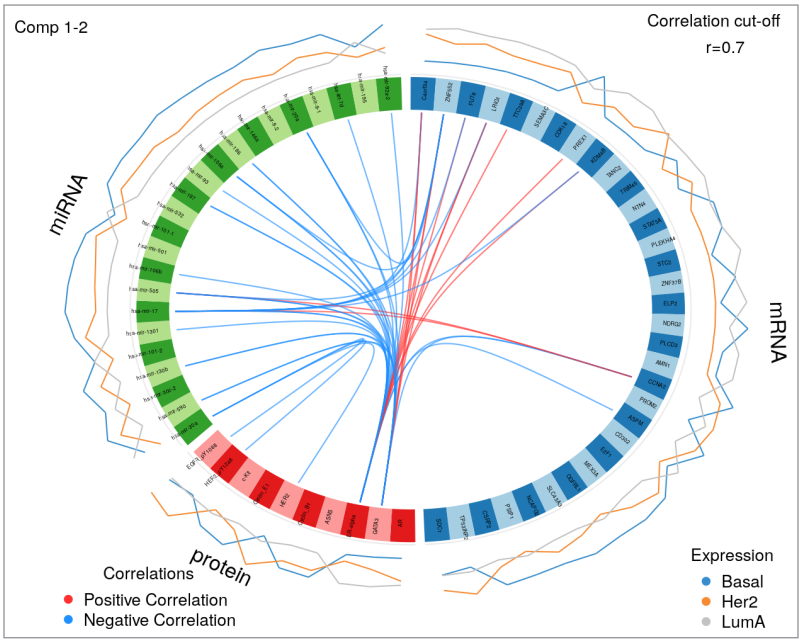


Figure 12. Correlation between variables using the `circosPlot` function [62].

The variables' names are small and can be challenging to observe. However, by using RStudio export function, a file in .pdf format can be exported with enough resolution to allow the observation of which variables correlate to each other, whether it is a positive or negative correlation and the expression of each dataset.

The main drawback of these platforms is the necessity of a previous identification and validation of the metabolites/proteins in order to visualize the possible interactions appropriately among the omic experiments. Another drawback of MixOmics is that all data need to be acquired from the same samples. OmicsNet is excellent at seeing a general linkage between the datasets and identifying the differential pathways, visually. MixOmics can present the correlations between variables through two datasets, indicating the most robust interaction, so the results obtained in OmicsNet can be filtered, showing the most relevant interactions among the metabolites and proteins. They can be used individually; however, together, they can provide a large amount of relevant biological interactions, which can unravel new biochemical mechanisms in several studies worldwide.

BIOLOGICAL INTERPRETATION: ID CROSS-CONNECTIONS

Multi-omics data provide a holistic view of the biological system. In this review, the description on how to create a multi-omics dataset and some platforms to perform multi-omics analysis were presented. Currently available methods approach the integration and interpretation at the downstream process, *i.e.*, using the identifiers (ID) from genomics, proteomics, and metabolomics, to understand the gaps left from single-omic studies [69]. Important biological information has been obtained, especially from integrative approaches using metabolomics and proteomics data. Previous studies have provided evidence concerning the organizational aspects of biomolecules [70], with great relevance to systems knowledge, such as the network topologies and spatial organization of enzyme interactions that correlate with metabolic efficiency analysis [70].

In this section, some publicly available-databases and the most common workflows used for biological interpretation are described. In this context, Yugi and collaborators [71] provided an important overview about the connection steps between all interaction procedures. They classified the available methods based on regulation levels into five categories: (1) metabolic, (2) transcriptional factors, (3) kinase-substrate relationship (KSR), (4) protein-protein interaction (PPI), and (5) enzyme allosteric regulation by small molecules (Figure 13A) [71]. For each molecule identified, the ID is used to integrate the single-omic approaches (Figure 13B).

Amongst metabolomics researchers, the Kyoto Encyclopedia of Genes and Genomes (KEGG) is the main option for ID conversion, regarding the cross-connection along with the other omic strategies (Figure 13B), *i.e.*, the IDs are used to perform multi-omics analysis and interpretation. KEGG database plays a pivotal role in connecting multiple omics data by ID manipulation, allowing conversion among molecular entities, giving access to different platforms to build pathway maps in order to understand the interaction levels. Another option is MetaboAnalyst [53], a web-based tool suite that host a platform to integrate the omics data based on changes in both gene/protein expression and metabolite concentrations concerning the biochemical pathway and phenotype.

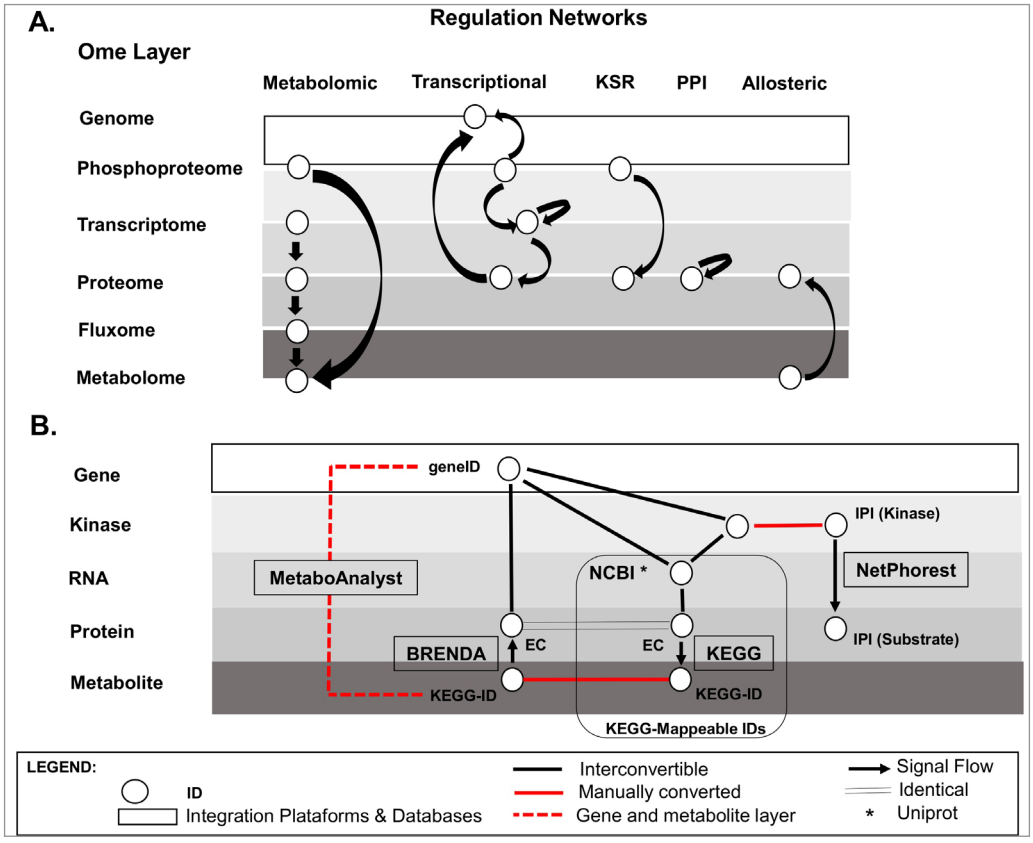


Figure 13. Cross-connection and multi-omics interpretation. **(A)** Single-omic classification and connected layers by regulation IDs. From metabolic to allosteric regulation and omics correlation are indicated in horizontal and vertical positions, respectively. The arrows indicate the directions of regulation processes across the layers. **(B)** Connecting IDs (circles) across multiple omic layers. Lines drawn between circles indicate conversion between IDs. Black lines indicate that an ID association or conversion can be performed by using cross-reference tables provided by, *e.g.*, KEGG. Red lines indicate manual conversions required for IDs. Abbreviations: IPI, International Protein Index; KSR, kinase–substrate relationship; PPI, protein–protein interaction; BRENDA: BRaunschweig ENzyme Database (modified from [71]).

CHALLENGES IN OMIC APPROACHES

The simultaneous integration of multi-omic approaches represents a powerful strategy to disclose the mechanisms connecting identified genetic variations to several conditions. Nevertheless, many sources of variability are combined into statistical models when identifying key drivers and pathways that aim to represent the most significant contributions to a biological process.

As mentioned before, the problems observed in multi-omics analysis begin with the individual omic approaches. In this way, bias is a relevant drawback to emphasize when developing an omics project. Ransohoff has defined bias as "the systematic erroneous association of some characters with a group in a way that distorts the comparison with another group" [72]. It is likely to be introduced because omic measurements are usually performed without a specific hypothesis in mind for lacking some biological reasoning. In reductionist approaches such as the measurement of glucocorticoid levels in response to acute stress [73] or the methylation of DNA as an indicator of repressed gene expression [74], for example, some of the biological relationships these markers present are established and underlie the experiment's initial hypothesis. In this way, the variables are mostly known and tightly controlled. In omic approaches, which refer to the analysis of global sets of biomolecules (*omes*), this is not the case. The complex interactions between genes, proteins, lipids, among other biomolecules, and the interplay with environmental, social, psychological and etiological factors make the associations of cause and effect difficult; thus, many omics-based experiments are poorly defined at the beginning [6,24]. Therefore, the findings from omic approaches need further confirmation by reductionist experiments.

The high potential of biomarkers in the clinical field creates an atmosphere in science urging to discover them as quickly as possible, since these biomarkers can detect diseases and potentially save lives. Although this rush, it is essential that measurements and experiments are designed as carefully as possible with their most addressed drawbacks and limitations. Their implications in the social and health care systems are significant and must be performed with attention.

Moreover, bias is complicated to solve; it cannot be intentionally introduced and is even harder to address its sources. It can be brought if a group is treated differently from other ones. That includes different collection protocols for test and control groups, different protocols for sample extraction and preparation, as well as the number of freezing-thaw cycles and different storage conditions, since the samples can change due to molecular responses to altering conditions [24]. Besides, during analysis, bias might be carried by no proper selection of subjects or participants, such as match of age range or gender among the conditions. For example, if a group of patients has a median age of 70 years and the control group a median of 25, then bias can be introduced due to age. Ethical, genre and socio-economic data should also be addressed and reported, since they are factors that can influence the experiment outcomes [75]. Another source is analytical bias when the measured signal shifts over time. Though, in this case, sample randomization is a good practice to avoid this effect [76].

Different analysis approaches also have different sources of bias (*e.g.*, specimen collection, sample preparation). For example, studies involving disease prognosis by RNA analysis might be affected by the specimen collection, while for the same experiment conducted with DNA, it would be less influenced, since DNA is more stable than RNA. There is no statistical way to solve bias, nor has it any relationship with reproducibility. The best way to minimize its impact is firmly controlling the experimental design and, if possible, addressing its probable sources. If researchers can detect and consider the most likely magnitude and direction of impact of bias, then it can be judged as being present, but not relevant. In any case, bias does not need to be entirely managed; nevertheless, it needs to be considered and the protocols reported in detail, so reviewers and other scientists can interpret the reliability of the study [72].

Another problem with the omic sciences in regard to statistics is called the "Anna Karenina effect". It occurs when the null hypothesis is rejected without any true observation and can also be called "chance". It means that the statistical significance does not equate to biological relevance. Chance is overfitting of the data in a way that discriminates with high precision between test and control groups, which is indicated by the *p*-value. This determines if an observation occurs by chance [77,78]. Many researchers believe that

a low *p*-value is a guarantee of the reproducibility and strength of the study; however, it is not always true. The reproducibility of a statistically significant result, when the probability is true, is substantially lower than one might expect. It does not relate directly to sample size effect, but varies along experimental replicates generated from the same population [79].

Lay Jr. *et al.* [6] suggest that most researchers in the omics field believe that a large number of measurements involved within the approaches could somehow compensate the small sample number. It does not occur, even assuming very low *p*-values such as 0.005. If there is no real association, even increasing the number of samples infinitely cannot render it valid. In order to overcome this issue, overfitting can be assessed through a random dataset for validation and further with external validation – application of the method within an external group to prove its validity [72]. As an alternative, cross-validation should be performed when the latter is not satisfied [80].

The difficulty in reproducing omic results obtained in different laboratories, in order to propose new biomarkers, brings barriers to the development of the field, since biomarkers are not thoroughly validated (Box 1). Therefore, it points out a need for better control and descriptions to follow the protocols applied in the research. Therefore, scientists can have a better understanding of the experimental designs and help to detect the sources of errors and develop alternatives to minimize them.

In order to overcome this issue, the proper design of the study and the control of the pre-analytical variables must be assessed and the influences of the analytical techniques should be reduced through the application of internal controls and standards for calibration within quality control [81]. A random sample selection from a studied population is ideal and should also be matched: the same median age, genres, routines, diet, types of medication, etc. Samples with uncertainty in response diagnoses or other clinical data should be excluded. It is highly indicated sample blocking since it can avoid the introduction of analytical and technical bias. All the steps should be done carefully and described accurately; moreover, standard operating procedures (SOPs) should be established in each stage. Figures of methods: variability, sensitivity and specificity should also be determined [82].

Once again, it is highly essential to demonstrate intra-laboratory and inter-laboratory reproducibility and standardization, reporting a detailed work in all stages. Some researchers claim the codes used in data treatment, which should be submitted together with the scientific articles containing omic studies [83]. Both experts and scientists of diverse fields, such as biostatisticians, program developers, and bioinformaticists, need to work together in order to overcome the difficulties found in omic sciences [6]. It does not seem to be an easy work; however, it brings reliable results. Moreover, it is necessary for the evolution of multi-omic approaches for further clinical application.

Box 1. Stages for biomarker validation [84-87]

Stage I - Pre-clinical exploratory studies
Biomolecule detection strategies (such as proteomics and metabolomics) enable the discovery of candidate biomarkers, achieved by screening through modern imaging techniques and other high-throughput techniques. Those identified markers are prioritized based on diagnostic, prognostic, or predictive characteristics, which may suggest their development to be of clinical use.

Stage II - Development of clinical trial for a disease (validation phase)
Clinical trial, based on the use of non-invasively obtained (non-surgical) samples, has two essential components. The first must record clinical utility, which assays need to be validated for reproducibility and demonstrated to be portable among different laboratories (full validation). The second consideration is that the assessment should analyze for clinical performance in terms of "sensitivity" and "specificity" within the determined preliminaries for the intended clinical use. Very often, biomarkers do not continue beyond this stage due to the lack of reliable and accurate assay tests or validation studies do not confirm whether the markers have a proper sensitivity or specificity to continue their development.

Stage III - Retrospective longitudinal repository studies
Researchers evaluate the sensitivity and specificity of the test for detecting diseases that have not been clinically detected yet. Diseased patients are compared with control patients before their clinical diagnosis, providing evidence of the biomarker ability to detect preclinical disease. If biomarker levels in the studied individuals only present a little divergence from those in control subjects near the time of clinical diagnosis, then the biomarker reveals little promise for screening. Otherwise, when it exhibits significative different levels between disease and control patients within a broader time before diagnosis, the biomarker validation is fulfilled, and it is ready for clinical use.

Stage IV - Screening studies perspective
A positive test triggers procedure for a definitive diagnostic, which is often invasive, and it could lead to an increased burden of economic health care. Thus, the study consists in determining the operational characteristics (disease state at the time of detection) of the biomarker-based screening test in a relevant population by determining the proportion of detection and the proportion of false references.

Stage V - Disease control studies
The final phase is to evaluate how the biomarker test performs in a population. In order to determine whether the screening test reduces the burden (on morbidity and mortality) of the disease in a predefined community, large-scale studies are necessary.

In this review, (1) a description on how multi-omics can enhance the understanding of the biological complexity under a systems perspective was provided; (2) the application of omics data integration to fill gaps generated from individual omic studies and (3) a workflow detailing how multi-omic approaches can be incorporated into filtering protocols that aim at identifying molecular candidates of specific biological processes were discussed; and (4) essential considerations and future directions that are relevant to the success of molecular targets selection supported by multi-omics data were pointed out, in view of the systems biology concept. It is important to emphasize that the omic strategies, as well as mathematical (*e.g.*, statistical) and computational tools, and the integrative methods mentioned in this review, are a subset

of the current methods available and additional ones can also be used to identify biomarker candidates and describe biological systems successfully.

FINAL CONSIDERATIONS

Since the beginning and development of omic sciences, investigations have been directed towards the discovery of biomarkers and their biochemical ability to explain biological conditions, *e.g.*, health-disease, host-parasite, even molecular factors that promote systems' dynamics. These search methods are perpetuated until the present day; however, for some researchers, with many caveats. The main one is the way that the experimental designs have been conducted in order to determine specific molecular patterns based on reductionist methods [88-91], even though these same targets belong to a physiological mechanism of high complexity, being part of a system.

Hence, methods for the integrative analysis of multi-omics data arise as an opportunity to thoroughly describe ideas regarding data mining using an integration algorithm to test the target molecules and data generated from omic approaches under a complex context. At the same time, biomarker discovery approaches require a completer and more accurate picture of the molecular systems' dynamics. The complexity of biological systems, the technological boundaries, the large number of biological variables and the relatively low number of biological samples analyzed are challenges in multi-omics, but it is still essential to recognize it as a potential tool to fill important gaps and respond questions not answered by omic strategies alone.

REFERENCES

1. Nielsen, J. *Annu. Rev. Biochem*., **2017**, *86*, pp 245–275 (http://doi.org/10.1146/annurev-biochem-061516-044757).
2. Kitano, H. *Nature*, **2002**, *420*, pp 206–210 (http://doi.org/10.1038/nature01254).
3. Hasin, Y.; Seldin, M.; Lusis, A. *Gen. Biol*., **2017**, *18*, pp 1–15 (http://doi.org/10.1186/s13059-017-1215-1).
4. Zhang, A.; Sun, H.; Wang, Z.; Sun, W.; Wang, P.; Wang, X. *Planta Med*., **2010**, *76*, pp 2026–2035 (https://doi.org/10.1055/s-0030-1250542).
5. Forshed, J. J. *Proteome Res*., **2017**, *16*, pp 3954-3960 (https://doi.org/10.1021/acs.jproteome.7b00418).
6. Lay Jr., J. O.; Borgmann, S.; Liyanage, R.; Wilkins, C. L. *Trends Anal. Chem*., **2006**, *25*, pp 1046-1056 (https:// 10.1016/j.trac.2006.10.007).
7. Villas-Bôas, S. G.; Mas, S.; Åkesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spec. Rev*., **2005**, *24*, pp 613-646 (https://doi.org/10.1002/mas.20032).
8. Garlow, S. J. *Neuron*, **2002**, *34*, pp 327-328 (https://doi.org/10.1016/S0896-6273(02)00680-3).
9. Hanash, S. *Nature*, **2003**, *422*, pp 226–232 (https://doi.org/10.1038/nature01514).
10. Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol*., **2000**, *18*, pp 1157-1161 (https://doi.org/10.1038/81137).
11. Kopczynski, D.; Coman, C.; Zahedi, R. P.; Lorenz, K.; Sickmann, A.; Ahrends, R. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, **2017**, *1862*, pp 808-811 (https://doi.org/10.1016/j.bbalip.2017.02.003).
12. Boccio, P.; Del Rossi, C.; Cicalini, I.; Sacchetta, P.; Pieragostino, D. *Proteomics Clin. Appl*., **2016**, *10*, pp 470-484 (https://doi.org/10.1002/prca.201500083).
13. Cambiaghi, A.; Díaz, R.; Martinez, J. B.; Odena, A.; Brunelli, L.; Caironi, P.; Masson, S.; Baselli, G.; Ristagno, G.; Gattinoni, L.; De Oliveira, E.; Pastorelli, R.; Ferrario, M. *Sci. Rep*., **2018**, *8*, pp 1-12 (https://doi.org/10.1038/s41598-018-25035-1).
14. Gui, S.-W.; Liu, Y.-Y.; Zhong, X.-G.; Liu, X.; Zheng, P.; Pu, J.-C.; Zhou, J.; Chen, J.-J.; Zhao, L.-B.; Liu, L.-X.; Xu, G.; Xie, P. *Neuropsychiatr. Dis. Treat*., **2018**, *14*, pp 1451-1561 (https://doi.org/10.2147/NDT.S164134).
15. Pinu, F. R.; Beale, D. J.; Paten, A. M.; Kouremenos, K.; Swarup, S.; Schirra, H. J.; Wishart, D. *Metabolites*, **2019**, *9*, pp 1-31 (https://doi.org/10.3390/metabo9040076).

16. Misra, B. B.; Langefeld, C.; Olivier, M.; Cox, L. A. *J. Mol. Endocrinol.*, **2019**, *62*, pp R21-R25 (https://doi.org/10.1530/JME-18-0055).
17. Fondi, M.; Liò, P. *Microb. Res.*, **2015**, *171*, pp 52-64 (https://doi.org/10.1016/j.micres.2015.01.003).
18. Álvares-Sanchez, B.; Priego-Capote, F.; Luque de Castro, M. D. *Trends Anal. Chem.*, **2010**, *29*, pp 111-119 (https://doi.org/10.1016/j.trac.2009.12.003).
19. Tuck, M. K.; Chan, D. W.; Chia, D.; Godwin, A. K.; Grizzle, W. D.; Krueger, K. E.; Rom, W.; Sanda, M.; Sorbara, L.; Stass, S.; Wang, W.; Brenner, D. E. *J. Proteome Res.*, **2009**, *8* (1), pp 113-117 (https://doi.org/10.1021/pr800545q).
20. Lauridsen, M.; Hansen, S. H.; Jaroszewski, J. W.; Cornet, C. *Anal. Chem.*, **2007**, *79* (3), pp 1181-1186 (https://doi.org/10.1021/ac061354x).
21. Cuevas-Córdoba, B.; Santiago-García, *J. OMICS*, **2014**, *18* (2), pp 87-97 (https://doi.org/10.1089/omi.2013.0064).
22. Fan, X.; Peters, B. A; Min, D.; Ahn, J.; Hayes, R. B. *PLoS One*, **2018**, *13* (4), e0194729 (https://doi.org/10.1371/journal.pone.0194729).
23. Teunissen, C. D; Petzold, A.; Bennett, J. L; Berven, F. S.; Brundin, L.; Comabella, M; Franciotta, D.; Frederiksen, J. L.; Fleming, J. O.; Furlan, R.; et al. *Neurology*, **2009**, *73* (22), pp 1914-1922 (https://doi.org/10.1212/WNL.0b013e3181c47cc2).
24. Comes, A. L; Papiol, S.; Mueller, T.; Geyer, P. E.; Mann, M.; Schulze, T. G. *Transl. Psychiatry*, **2018**, *8*, p 160 (https://doi.org/10.1038/s41398-018-0219-2).
25. Salem, M.; Bernach, M.; Bajdzienko, K.; Giavalisco, P. *J. Vis. Exp.*, **2017**, *124*, e55802 (https://doi.org/10.3791/55802).
26. Coman, C.; Solaris, F. A.; Hentschel, A.; Sickmann, A.; Zahedi, R. P.; Ahrends, R. *Mol. Cell. Proteomics*, **2016**, *15* (4), pp 1453-1466 (https://doi.org/10.1074/mcp.M115.053702).
27. Breil, C.; Vian, M. A.; Zemb, T.; Kunz, W.; Chemat, F. *Int. J. Mol. Sci.*, **2017**, *18* (4), p 708 (https://doi.org/10.3390/ijms18040708).
28. Olszowy, P.; Buszweski, B. *J. Sep. Sci.*, **2014**, *37* (20), pp 2920-2928 (https://doi.org/10.1002/jssc.201400331).
29. Gutiérrez, A.; Cerón, J. J.; Razzazi-Fazeli, E.; Schlosser, S.; Tecles, F. *BMC Vet. Res.*, **2017**, *13* (1), p 375 (https://doi.org/10.1186/s12917-017-1296-9).
30. Yu, Y.; Suh, M.-J.; Sikorski, P.; Kwon, K.; Nelson, K. E.; Pieper, R. *Anal. Chem.*, **2014**, *86* (11), pp 5470-5477 (https://doi.org/10.1021/ac5008317).
31. Álvarez-Sánchez, B.; Priego-Capote, F.; Luque de Castro, M. D. *J. Chromatogr. A*, **2012**, *1248*, pp 178-181 (https://doi.org/10.1016/j.chroma.2012.05.029).
32. Katajamaa, M.; Orešič, M. *J. Chromatogr. A*, **2007**, *1158*, pp 318–328 (https://doi.org/10.1016/j.chroma.2007.04.021).
33. Holman, J. D.; Tabb, D. L.; Mallick, P. *Curr. Protoc. Bioinformatics*, **2014**, *46* (1) pp 1–9 (https://doi.org/10.1002/0471250953.bi1324s46).
34. Smit, H. C.; Walg, H. L. *Chromatographia*, **1975**, *8* (7), pp 311–323 (https://doi.org/10.1007/BF02350794).
35. Schiffman, C.; Petrick, L.; Perttula, K.; Yano, Y.; Carlsson, H.; Whitehead, T.; Metayer, C.; Hayes, J.; Rappaport, S.; Dudoit, S. *BMC Bioinformatics*, **2019**, *20* (1), pp 1–10 (https://doi.org/10.1186/s12859-019-2871-9).
36. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.*, **2012**, *84* (11), pp 5035–5039 (https://doi.org/10.1021/ac300698c).
37. Libiseller, G.; Dvorzak, M.; Kleb, U.; Gander, E.; Eisenberg, T.; Madeo, F.; Neumann, S.; Trausinger, G.; Sinner, F.; Pieber, T.; Magnes, C. *BMC Bioinformatics*, **2015**, *16* (118), pp 1–10 (https://doi.org/10.1186/s12859-015-0562-8).
38. Petera, M.; Le Corguillé, G.; Martin, J.-F.; Guitton, Y. *LC-MS analysis* (Galaxy Training Materials). Available from: https://training.galaxyproject.org/training-material/topics/metabolomics/tutorials/lcms/tutorial.html [accessed May 15 2020].

39. Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. *Anal. Chem*., **2012**, *84*, pp 283–289 (https://doi.org/10.1021/ac202450g).
40. Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. *Genome Biol*., **2004**, *5*, pp R80.1-R80.16 (https://doi.org/10.1186/gb-2004-5-10-r80).
41. Simonet, B. M. *Trends Anal. Chem*., **2005**, *24* (6), pp 525–531 (https://doi.org/10.1016/j.trac.2005.03.011).
42. Gromski, P. S.; Xu, Y.; Hollywood, K. A.; Turner, M. L.; Goodacre, R. *Metabolomics*, **2015**, *11* (3), pp 684–695 (https://doi.org/10.1007/s11306-014-0738-7).
43. Salek, R. M.; Maguire, M. L.; Bentley, E.; Rubstov, D. V.; Hough, T.; Cheeseman, M.; Nunez, D.; Sweatman, B. C.; Haselden, J. N.; Cox, R. D.; et al. *Physiol. Genomics*, **2007**, *29* (2), pp 99–108 (https://doi.org/10.1152/physiolgenomics.00194.2006).
44. Wang, P.; Tang, H.; Zhang, H.; Whitewalker, J.; Paulovich, A. G.; *McIntosh, M. Biocomputing*, **2006**, pp 315–326 (https://doi.org/10.1142/9789812701626_0029).
45. Sauve, A.; Speed, T. *Proc. Gensips*, **2004**, pp 1–4.
46. Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. *Sci. Rep*., **2018**, *8* (1), pp 1–10 (https://doi.org/10.1038/s41598-017-19120-0).
47. Bijlsma, S.; Bobeldijk, I.; Verheij, E.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; Ommen, B. V.; Smilde, A. K. *Anal. Chem*., **2006**, *78* (2), pp 567–574 (https://doi.org/10.1021/ac051495j).
48. Pleil, J. D. J. *Breath Res*., **2016**, *10* (4) (https://doi.org/10.1088/1752-7155/10/4/045001).
49. Antweiler, R. C.; Taylor, H. E. *Environ. Sci. Technol*., **2008**, *42* (10), pp 3732–3738 (https://doi.org/10.1021/es071301c).
50. Troyanskaya, O.; Cantor, M.; Shelock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R. B. *Bioinformatics*, **2001**, *17* (6), pp 520–525 (https://doi.org/10.1093/bioinformatics/17.6.520).
51. Stacklies, W.; Redestig, H; Scholz, M.; Walther, D.; Selbig, J. *Bioinformatics*, **2007**, *23* (9), pp 1164–1167 (https://doi.org/10.1093/bioinformatics/btm069).
52. Do, K. T.; Wahl, S.; Raffler, J.; Molnos, S.; Laimighofer, M.; Adamski, J.; Suhre, K.; Strauch, K.; Peters, A.; Gieger, C; et al. *Metabolomics*, **2018**, *14* (10), pp 1–18 (https://doi.org/10.1007/s11306-018-1420-2).
53. Chong, J.; Wishart, D. S.; Xia, J. *Curr. Protoc. Bioinformatics*, **2019**, *68* (1), pp 1–128 (https://doi.org/10.1002/cpbi.86).
54. Bruch, M.; Bodden, E.; Monperrus, M.; Mezini, M. Proceedings of the FSE/SDP Work. Fut. Softw. Eng. Res. FoSER, **2010**, pp 53–57 (https://doi.org/10.1145/1882362.1882374).
55. Valdman, J. *Applications from Engineering with MATLAB Concepts*. InTech, Rijeka, Croatia, **2016**, Chapter 8, p. 171.
56. Pesaran, M. H. *J. Appl. Econom*., **2012**, *27*, pp 167–172 (https://doi.org/10.1002/jae).
57. Domon, B.; Aebersold, R. *Mol. Cell. Proteomics*, **2006**, *5*, pp 1921–1926 (https://doi.org/10.1021/jasms.8b04081).
58. Cox, J.; Michalski, A.; Mann, M. *J. Am. Soc. Mass Spectrom*., **2011**, *22*, pp 1373-1380 (https://doi.org/10.1021/jasms.8b04081).
59. Tyanova, S.; Temu, T.; Cox, J. *Nat. Protoc*., **2016**, *11* (12), pp 2301–2319 (https://doi.org/10.1038/nprot.2016.136).
60. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. *Nat. Methods*, **2016**, *13* (9), pp 731–740 (https://doi.org/10.1038/nmeth.3901).
61. Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. *PLoS Comput. Biol*., **2017**, *13*, pp 1-19 (https://doi.org/10.1371/journal.pcbi.1005752).
62. Zhou, G.; Xia, J. *Nuc. Acids Res*., **2018**, *46*, pp W514-W522 (https://doi.org/10.1093/nar/gky510).
63. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. *Nuc. Acids Res*., **1999**, *27*, pp 29-34 (https://doi.org/10.1093/nar/27.1.29).
64. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; et al. *Nuc. Acids Res*., **2013**, *41*, pp 801-D807 (https://doi.org/10.1093/nar/gks1065).

65. Zhou, G.; Xia, J. *Curr. Protoc. Bioinformatics*, **2019**, *65* (1), e69 (https://doi.org/10.1002/cpbi.69).
66. Shannon, P; Markiel, A; Ozier, O; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwilowski, B.; Ideler, T. *Genome Res*., **2003**, *13*, pp 2498-2504 (https://doi.org/ 10.1101/gr.1239303).
67. Singh, A.; Shannon, C. P.; Gautier, B.; Rohart, F.; Vacher, M.; Tebbutt, S. J.; Cao, Kim-Anh. L. *Bioinformatics*, **2019**, *35*, pp 3055–3062 (https://doi.org/10.1093/bioinformatics/bty1054).
68. González, I.; Cao, K. L.; Davis, M. J.; Déjean, S. *BioData Mining*, **2012**, *5*, pp 1-23 (https://doi.org/10.1186/1756-0381-5-19).
69. Conesa, A.; Beck, S. *Sci. Data*, **2019**, *6* (1), pp 1-4 (https://doi.org/10.1038/s41597-019-0258-4).
70. Durek, P.; Walther, D. *BMC Syst. Biol*., **2008**, *2* (1), p 100 (https://doi.org/10.1186/1752-0509-2-100).
71. Yugi, K.; Kubota, H.; Hatano, A.; Kuroda, S. *Trends Biotechnol*., **2016**, *34* (4), pp 276-290 (https://doi.org/10.1016/j.tibtech.2015.12.013).
72. Ransohoff, D. F. *Nat. Rev. Cancer*, **2005**, *5*, pp 142-149 (https://doi.org/10.1038/nrc1550).
73. Cattaneo, A.; Riva, M. A. *J. Steroid Biochem*., **2016**, *160*, pp 169-74 (https://doi.org/10.1016/j.jsbmb.2015.07.021).
74. Curradi, M.; Izzo, A.; Bandaracco, G.; Landsberger, N. *Mol. Cell Biol*., **2002**, *22* (9), pp 3157-3173 (https://doi.org/10.1128/MCB.22.9.3157-3173.2002).
75. Renzi, C.; Provencal, N.; Bassil, K. C.; Evers, K.; Kihlbom, U.; Radford, E. J.; Koupil. I.; Mueller-Myshok, B.; Hansson, M. G.; Rutten, B. P. F. *Progr. Mol. Biol. Transl*., **2018**, *158*, pp 299-323 (https://doi.org/10.1016/bs.pmbts.2018.04.011).
76. Suresh, K. *J. Hum. Reprod. Sci*., **2011**, *4*, pp 8-11 (https://doi.org/10.4103/0974-1208.82352).
77. Goodman, S. N. *Ann. Intern. Med*., **1999**, *130*, pp 995-1004 (https://doi.org/10.7326/0003-4819-130-12-199906150-00008).
78. Goodman, S. N. *Semin. Hematol*., **2008**, *45*, pp 135-140 (https://doi.org/10.1053/j.seminhematol.2008.04.003).
79. Goh, W. W. B.; Wong, L. *Trends Biotechnol*., **2018**, *36*, pp 488-498 (https://doi.org/10.1016/j.tibtech.2018.01.013).
80. Browne, M. W. *J. Math. Psy.*, **2000**, *44*, pp 108-132 (https://doi.org/10.1006/jmps.1999.1279).
81. Lumbreras, B.; Porta, M.; Marquez, S.; Pollán, M.; Parker, L. A.; Hernández-Aguado, I. *Proteomics Clin. Appl.*, **2009**, *3*, pp 173-184 (https://doi.org/10.1002/prca.200800092).
82. Parikh, R.; Mathai, A.; Parikh, S.; Sekhar, G. C.; Thomas. R. *Indian J. Ophtalmol*., **2008**, *56*, pp 45-50 (https://doi.org/ 10.4103/0301-4738.37595).
83. Baggerly, K.; Coombes, K. *Clin. Chem*., **2011**, *57*, pp 688-690 (https://doi.org/10.1373/clinchem.2010.158618).
84. Kumar, M.; Sarin, S. K. *Curr. Trends Sci*., **2010**, *15*, pp 403-417.
85. Pepe, M. S.; Etzioni, R.; Feng, Z.; Potter, J. D.; Thompson, M. L.; Thornquist, M. *J. Natl. Cancer Inst*., **2001**, *93*, pp 1054-1061 (https://doi.org/10.1093/jnci/93.14.1054).
86. Sataloff, R. T. *Ear Nose Throat J*., **2004**, *83*, pp 665-665 (https://doi.org/10.1177/014556130408301001).
87. EDRN. Early Detection Research Network Manual of Operations [Internet]. **2009**. Available from: https://edrn.nci.nih.gov/docs [accessed December 10, 2019].
88. Sarkar, S. *Genetics and Reductionism*. Cambridge University Press, Cambridge, **1998**.
89. Li, R.; Ma, T.; Gu, J.; Liang, X.; Li, S. *Sci. Rep*., **2013**, *3*, pp 1543-1549 (https://doi.org/10.1038/srep01543).
90. Chibbaro, S.; Rondoni, L.; Vulpiani, A. *Reductionism, Emergence and Levels of Reality*. Springer International Publishing, **2014**.
91. Calvani, R.; Picca, A.; Cesari, M.; Tosato, M.; Marini, F.; Manes-Gravina, E.; Berbanei, R.; Landi, F.; Marzetti, E. *Curr. Protein Pept. Sci*., **2018**, *19*, pp 639-642 (https://doi.org/10.2174/1389203718666170516115422).

Supplemental Table I. (Bio)chemical techniques to enrich and characterize the main PTM by mass spectrometry analyses

Techniques	Enrichment & Purification methods		PTMs		Large-scale analysis application	Published protocols (examples)*
Biochemical	Enzymatic Labeling		O-GLcNAc S-glutathionylation		Long preparation cycle & multiple steps	1-3
	Immunoaffinity		Acetylation - Lys		Yes	4,5
			Methylation	Lys	Yes, but poor specificity	6-8
				Arg	Yes, but poor specificity	
			Ubiquitination - Lys		Yes	9-11
			Phosphorylation	Tyr	Yes	12,13
				His	No, but poor specificity	14-16
Chemical	Chromatography	IMAC	Phosphorylation	Ser, Thr and Tyr	Yes	17 - 19
		HILIC	N-linked glycopeptides		Yes, relatively poor specificity	20 - 23
		Boric Acid	N-linked glycopeptides		No, poor specificity	23, 24
	Chemical Derivatization	Hydrazide chemistry	N-linked glycopeptides		Yes, high specificity	25,26
		Biotin switch technique	Redox Modification		Comprised by incomplete reaction and side reaction	27-29
		Direct reductive methylation	Cysteine Oxidation		Long preparation cycle & multiple steps	30,31

*In these articles, the authors provide other references concerning protocols and the main parameters for MS analyses.

Supplemental Table I – References

1. Pineda-Molina, E.; Klatt, P.; Vázquez, J.; Marina, A.; de Lacoba, M. G.; Pérez-Sala, D.; Lamas, S. *Biochemistry*, **2001**, *40*, pp 14134-14142 (https://doi.org/10.1021/bi011459o).
2. Cao, W.; Cao, J.; Huang, J.; Yao, J.; Yan, G.; Xu, H.; Yang, P. *PLoS ONE*, **2013**, *8*, e76399 (https://doi.org/10.1371/journal.pone.0076399).
3. Wu, H. Y.; Lu, C. T.; Kao, H. J.; Chen, Y. J.; Chen, Y. J.; Lee, T. Y. *BMC Bioinf*., **2014**, *15*, S1 (https://doi.org/10.1186/1471-2105-15-S16-S1).
4. Junot, C.; Pruvost, A.; Créminon, C.; Grognet, J. M.; Benech, H.; Ezan, E. *J. Chromatogr. B: Biomed. Sci. Appl.*, **2001**, *752*, pp 69-75 (https://doi.org/10.1016/S0378-4347(00)00520-X).
5. Thao, S.; Escalante-Semerena, J. C. *Curr. Opin. Microbiol*., **2011**, *14*, pp 200-204 (https://doi.org/10.1016/j.mib.2010.12.013).
6. Wang, Q.; Liu, Z.; Wang, K.; Wang, Y.; Ye, M. *Anal. Chim. Acta*, **2019**, *1068*, pp 111-119 (https://doi.org/10.1016/j.aca.2019.03.042).
7. Chen, Y. *Methods Mol. Biol*., **2016**, *1410*, pp 23-37 (https://doi.org/10.1007/978-1-4939-3524-6_2).
8. Wesche, J.; Kühn, S.; Kessler, B. M.; Salton, M.; Wolf, A. *Cell. Mol. Life Sci*., **2017**, *74*, pp 3305-3315 (https://doi.org/10.1007/s00018-017-2515-z).
9. Bustos, D.; Bakalarski, C. E.; Yang, Y.; Peng, J.; Kirkpatrick, D. S. *Mol. Cell. Proteomics*, **2012**, *11*, pp 1529-1540 (https://doi.org/10.1074/mcp.R112.019117).
10. Xu, G.; Paige, J. S.; Jaffrey, S. R. *Nat. Biotechnol*., **2010**, *28*, pp 868-873 (https://doi.org/10.1038/nbt.1654).
11. Valdés, A.; Bergström, S. L. *Proteomics*, **2020**, *20*, 1800425 (https://doi.org/10.1002/pmic.201800425).
12. Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A.; Yang, V.; Aguiar, M.; Kornhauser, J.; Jia, X.; et al. *Mol. Cell. Proteomics*, **2014**, *13*, pp 372-387 (https://doi.org/10.1074/mcp.O113.027870).
13. Liang, Y.; Zhu, X.; Zhao, M.; Liu, H. *Methods*, **2012**, *56*, pp 174-179 (https://doi.org/10.1016/j.ymeth.2011.08.006).
14. Fuhs, S. R.; Meisenhelder, J.; Aslanian, A.; Ma, L.; Zagorska, A.; Stankova, M.; Binnie, A.; Al-Obeidi, F.; Mauger, J.; Lemke, G.; et al. *Cell*, **2015**, *162*, pp 198-210 (https://doi.org/10.1016/j.cell.2015.05.046).
15. Kee, J. M.; Oslund, R. C.; Perlman, D. H.; Muir, T. W. *Nat. Chem. Biol*., **2013**, *9*, pp 416-421 (https://doi.org/10.1038/nchembio.1259).
16. Adam, K.; Lesperance, J.; Hunter, T.; Zage, P. E. *Int J. Mol. Sci*., **2020**, *21*, p 3319 (https://doi.org/10.3390/ijms21093319).
17. Ruprecht, B.; Koch, H.; Medard, G.; Mundt, M.; Kuster, B.; Lemeer, S. *Mol. Cell. Proteomics*, **2015**, *14*, pp 205-215 (https://doi.org 10.1074/mcp.M114.043109).
18. Gan, C. S.; Guo, T.; Zhang, H.; Lim, S. K.; Sze, S. K. *J. Proteome Res*., **2008**, *7*, pp 4869-4877 (https://doi.org/10.1021/pr800473j).
19. Kielkopf, C. L.; Bauer, W.; Urbatsch, I. L. *Cold Spring Harbor Protocols*, **2020** (https://doi.org/10.1101/pdb.prot102194).
20. Shu, Q.; Li, M.; Shu, L.; An, Z.; Wang, J.; Lv, H.; Yang, M.; Cai, T.; Hu, T.; Fu, Y.; Yang, F. *Mol. Cell. Proteomics*, **2020**, *19*, pp 672-689 (https://doi.org/ 10.1074/mcp.RA119.001791).
21. Zhang, H.; Lv, Y.; Du, J.; Shao, W.; Jiao, F.; Xia, C.; Gao, F.; Yu, Q.; Liu, Y.; Zhang, Y.; et al. *Anal. Chim. Acta*, **2020**, *1098,* pp 181-189 (https://doi.org/10.1016/j.aca.2019.11.012).
22. Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X. *Trends Anal. Chem.*, **2020**, *124*, 115570 (https://doi.org/10.1016/j.trac.2019.06.020).
23. Saleem, S.; Sajid, M. S.; Hussain, D.; Jabeen, F.; Najam-ul-Haq, M.; Saeed, A. *Anal. Bioanal. Chem*., **2020**, *412*, pp 1509-1520 (https://doi.org/10.1007/s00216-020-02427-9).
24. Jiang, B.; Huang, J.; Yu, Z.; Wu, M.; Liu, M.; Yao, J.; Zhao, H.; Yan, G.; Ying, W.; Cao, W.; Yang, P. *Talanta*, **2019**, *199*, pp 254-261 (https://doi.org/10.1016/j.talanta.2019.02.010).

25. Sajid, M. S.; Jabeen, F.; Hussain, D.; Ashiq, M. N.; Najam-ul-Haq, M. *Anal. Bioanal. Chem*., **2017**, *409*, pp 3135-3143 (https://doi.org/10.1007/s00216-017-0254-5).
26. Napoli, A. Use of Galactose Oxidase in Hydrazide-Capturing Technique to Isolate N-linked Glycoallergens from Olive Tree Pollen. In: *IOP Conference Series: Materials Science and Engineering*, **2020**, *739* (1), 012052. IOP Publishing.
27. McDonagh, B.; Ogueta, S.; Lasarte, G.; Padilla, C. A.; Bárcena, J. A. *J. Proteomics*, **2009***, 72*, pp 677-689 (https://doi.org/10.1016/j.jprot.2009.01.023).
28. Bykova, N. V.; Rampitsch, C. *Proteomics*, **2013**, *13*, pp 579-596 (https://doi.org/10.1002/pmic.201200270).
29. Zhang, T.; Gaffrey, M. J.; Qian, W. J.; Thrall, B. D. Oxidative Stress and Redox Modifications in Nanomaterial–Cellular Interactions. In: Bonner, J. C.; Brown, J. M. (Eds.). *Interaction of Nanomaterials with the Immune System*. Springer, Cham, **2020**, pp 127-148.
30. McShane, A. J.; Shen, Y.; Castillo, M. J.; Yao, X. *J. Am. Soc. Mass Spectrom*., **2014**, *25*, pp 1694-1704 (https://doi.org/10.1007/s13361-014-0951-7).
31. Alcock, L. J.; Perkins, M. V.; Chalker, J. M. *Chem. Soc. Rev*., **2018**, *47*, pp 231-268 (https://doi.org/10.1039/C7CS00607A).

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 45-63
doi: 10.30744/brjac.2179-3425.AR-41-2019

BrJAC

ARTICLE

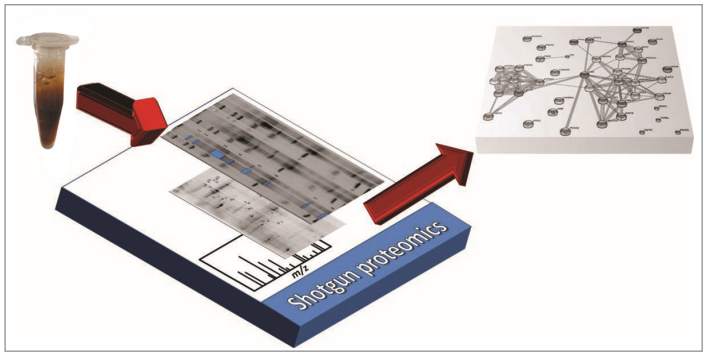
A Contribution to the Analytical Proteomics of Human Intestinal Mucus: Sampling, Sample Preparation, and Protein Identification Strategies

Paola Andrea Gómez Buitrago[1]* ✉ Carlos Augusto González Correa[1] Mario Santacoloma Osorio[2], Gonzalo Taborda Ocampo[1] Hector Henrique Ferreira Koolen[3] Marco Aurélio Zezzi Arruda[3]

[1]Research Group on Electrical Bio-Impedance, Doctoral Program on Biomedical Sciences, Universidad de Caldas, Calle 65 # 26-10, Edificio de Laboratorios, Office 502, Manizales, Colombia

[2]Clinical Department, Universidad de Caldas, Carrera 25 # 48-57, Manizales, Colombia

[3]Institute of Chemistry, University of Campinas (UNICAMP) and National Institute of Science and Technology in Bioanalytics, PO Box 6154, Campinas, Brazil



The colon is inhabited by circa ten bacteria for each cell of our organism (microbiota) and it is estimated that, altogether, they posses between 100 to 150 times more genes (microbiome) that the human genome. Between these two sets of genomes, acting as a biological interface, there is a mucus bilayer biochemically structured by the Mucin 2 glycoprotein. Several analytical proteomics strategies are presented in this study, which contribute to the proteomic map of human intestinal mucus, enlarging knowledge of this important sample. Utilizing mucus obtained from sampling via a transanal irrigation procedure, 15 different protocols for running the samples through 2-D PAGE in the pH ranges 3-10 or 4-7 were evaluated. The protocol that showed no streaking in 2-D PAGE and yielded more protein spots (116 and 220 spots were found in the 3 to 10 and 4 to 7 pH ranges, respectively) was considered the best and was the one that was finally used (protocol C). For improving protein identification, SDS-PAGE, 2-D PAGE and shotgun proteomics are carried out through nLC-MS/MS, and a Venn diagram built-up from identified proteins as well as cellular, bacterial and other proteins of clinical interest. Mucin 2, which is the main component of the mucus, is accurately identified, despite its insolubility and high molecular weight, indicating the complementarity and the success of the optimization conditions of all strategies employed. Finally, the interaction between 23 identified proteins is carried out, indicating two major classes of proteins: energy metabolism and energy consumption.

Keywords: Intestinal mucus, proteins, sampling method, shotgun proteomics.

Cite: Gómez-Buitrago, P-A.; González-Correa, C-A.; Santacoloma-Osorio, M.; Taborda-Ocampo, G.; Koolen, H-H.F.; Arruda, M. A. Z. **A Contribution to the Analytical Proteomics of Human Intestinal Mucus: Sampling, Sample Preparation, and Protein Identification Strategies**. *Braz. J. Anal. Chem.,* 2020, *7* (29), pp 45-63. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-41-2019

Submitted 25 November 2019, Resubmitted 15 May 2020, Accepted 6 July 2020, Available online 6 August 2020.

INTRODUCTION

Human intestinal mucus (HIM) is a highly complex sample of incompletely defined chemical structure. However, it is mainly composed of Mucin 2 (MUC2), a glycoprotein containing *ca*. 5200 amino acids and having a molecular weight of 2.6 MDa when completely *O*-glycosylated in the proline, threonine and serine (PTS) domains [1]. This macromolecule forms a net-like structure, where, among several other substances, small proteins, bacteria, viruses and Achaea are found [2]

In the colon, the mucus gel comprises a bilayer [3]: the first is a luminal layer with a lower concentration of MUC2, allowing the penetration of bacteria into the mucus and their attachment by means of glycans [1,4]. The second is the mucosal layer, which is in contact with the colonic epithelium. Being impenetrable and acting as a barrier, it prevents passage of bacteria due to the highest concentration of MUC2. The structure of colonic mucus is responsible for two of the main physiological functions necessary for our good health: well-being and survival. Additionally, this mucus possesses a complex variety of microorganisms that inhabit the colon, known as the colonic microbiota [5,6], and acts as an important line of defense against any kind of pathogenic microorganism [7], as well as incoming acids, enzymes and all metabolic waste. In this manner, several authors consider glycans attached to MUC2 to be the body's first line of defense [8,9].

It is important to highlight that the colon has approximately 10 bacteria per single cell in the entire human organism, and the size of the microbiome exceeds the human genome by 100 to 150 times [10]. In the small intestine, the bacterial density is lower when compared to the colon, and the mucus is more permeable to bacteria. However, the mucus from the small intestine has a higher concentration of antibacterial peptides, and proteins are secreted by Paneth cells and enterocytes. These cells seem to contribute to the production of Immunoglobulin A and mucins, and are secreted in the crypt mouth, covering the villi to diminish the risk of its contact with bacteria [11].

To the best of our knowledge, those few studies focusing on intestinal mucus proteomics have been conducted using mouse models. Using agarose-1-D and LC/MS-MS, previous results showed that the main component of the colonic mucus bilayer is MUC2 [1]. Later, these same authors performed a proteomic analysis of both layers of mouse colonic mucus, as well as of samples obtained from human biopsies and reported that cell proteins, serum and bacterial proteins and MUC2 were linked to the Fcgbp protein [12].

Besides being a complex sample, sampling of HIM is not an easy task [13,14]. In contrast to rats, which can be sacrificed, colonoscopy is the most frequently used strategy for sampling HIM [15], accounting for the small amount of mucus obtained with each sampling, which makes its handling and acquisition of possible replicates difficult. This hampers the work with such samples and, more probably, explains the scarce literature regarding HIM proteomics.

Sampling from the recto sigmoid using cytology brushes had been applied by evaluating protein concentrations with Surface Enhanced Laser Desorption Ionization – TOF (SELDI-TOF) through mono- and oligo-saccharides analysis [15]. In order to alleviate sampling problems, some authors use *in vitro* cultured human goblet cells, with the proteome of mucus obtained by ultracentrifugation of such cells. Using LC-MS and MS/MS, a set of membrane proteins were identified, but proteins such as FAM62B and ATP6AP2 were reported as structural components of the colonic mucus [16].

Due to the complexity of HIM and the difficulty to sample and handle it, this study focuses not only on some strategies for analyzing HIM proteins, but also on a sampling strategy for obtaining HIM samples in sufficient quantity and quality, while demonstrating how to isolate and treat them. All these strategies may contribute to enhancing analytical proteomic studies of HIM.

MATERIALS AND METHODS

Sampling and treatment of HIM samples

Samples of HIM were collected from four overweight individuals (32.4 ± 6.2 years old) participating in a weight loss program held at the University of Caldas (UC), Colombia. Exclusion criteria for this program included: uncontrolled high blood pressure, diabetes, pregnancy or breast-feeding, weight

loss ≥ 5 kg in the previous six months, major psychiatric conditions (such as nervous bulimia, abuse of psychotropic substances, or under psychiatric treatment), and suspicion of moderate to severe diverticulosis, endoscopically diagnosed. All participants signed an informed consent form approved by the Ethical Committee of the UC. The protocol included 5 days of supplemented fasting, daily consumption of psyllyum and montmorillonite and 1 daily transanal colon irrigation using a Colema Board® (California, USA), as suggested by the manufacturer. Throughout the treatment, volunteers were under the medical supervision of Carlos-Augusto González-Correa and Mario Santacoloma-Osorio, a general doctor and a gastroenterologist, respectively. The material eliminated from the 3rd to the 5th days during the hydrotherapy sessions (from 0.5 mL up to 5.0 mL) was collected using a plastic sieve as, during the first two days, the patients mostly eliminated fecal matter. Visual identifiable mucus associated with this material was separated using tweezers, subsequently suctioned with a Pasteur pipette and collected in 1.5 mL micro centrifuge (Eppendorf) tubes. Mucus collected after the hydrotherapy was then mixed with a cleaning buffer (0.1 M Tris-HCl, in 154 mM NaCl, 0.01% azide, 0.02% Tween-20, pH 7.4) [15] in 1:1 v/v proportion. This material was then centrifuged at 350 *g* for 5 min, and the supernatant (50 µL) discarded to eliminate bacteria and other solid particles. Gramm staining confirmed the absence of bacteria in the remaining supernatant. The obtained pellet was then stored at -80 °C for later analysis.

Extraction of proteins from HIM samples using protocol C

For protein extraction, a 300 µL sample was treated as explained in the preceding section, adding 300 µL of denaturing buffer (8 M urea, Bio–Rad Laboratories Hércules, California, USA), 2% 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate – CHAPS (Bio–Rad Laboratories Hércules, California, USA), 1% Dithiothreitol – DTT (Bio–Rad Laboratories Hércules, California, USA), 50 mM Tris (Bio–Rad Laboratories Hércules, California, USA), pH 7.0 and the addition of 15 µL of phenylmethylsulphonyl fluoride – PMSF (Sigma- Aldrich Milwaukee, Wisconsin, USA) to inhibit proteases. Lysis was carried out for 10 min and the proteins were cleaned through precipitation with *ReadyPrep 2-D Clean-up* kit from Bio-Rad™, following the manufacturer's instructions. The obtained precipitate was washed with 100 µL cold acetone (Carlo Erba, Paris, France) and water (90:10 mixture) by vortex stirring for 1 min and stored for 10 min at -20 ºC. Washing was carried out three times, eliminating the supernatant and adding fresh acetone solution each time. After the third wash and storage at -20 °C, the solution was centrifuged at 350 *g* for 5 min in a 5430 Eppendorf centrifuge, the supernatant discharged and 100 µL of the rehydration buffer (7 M urea, 2 M thiourea, 2% 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate – CHAPS, 50 mM Dithiothreitol – DTT, 1% ampholytes. All reagents were acquired from Bio–Rad Laboratories Hércules, California, USA) [17]. The sample required 1 h to dissolve and it was, intermittently, manually stirred during that time. Protein quantity was determined after protein extraction. Total protein content was measured using the *2-D Quant Kit* from GE Health Care Life Sciences (Uppsala, Sweden) following the manufacturer's instructions. A sample from a single volunteer was included for protein extraction, and the quantification performed three times in the same day. Protein extraction was carried out for all the samples obtained from volunteers that expelled mucus during the third day of treatment.

Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Samples from three volunteers were run in duplicate, representing two analytical and three biological replicas. The sample protein fraction with protocol C was also evaluated by SDS-PAGE under denaturing conditions. Gels of 17 cm were prepared, and each well was filled with a 1:1 mix of protein fraction and buffer (0.5 M Tris-HCl, 10% (w/v) glycerol (Sigma-Aldrich Milwaukee, Wisconsin, USA), 10% (v/v) SDS, 0.001% (m/v) bromophenol blue and 5% (v/v) *β*-mercaptoethanol with a pH of 6.8, reagents were acquired from Bio–Rad Laboratories Hércules, California, USA. Gels were run using an Ettan™ Daltsix electrophoresis system (GE Health Care Life Sciences, Uppsala, Sweden), and the processing carried out in two steps: 30 min at 90 V, 15 mA/gel and 5 h at 600 V and 25 mA/gel.

Two-dimensional gel electrophoresis

For this purpose, 15 different protocols of protein extraction were tested (see Table I) in order to standardize a method for extraction and separation of protein from HIM, carrying out one biological and three analytical replicas for each protocol. Table I lists all the protocols that were used in the search to obtain the best results in 2-D PAGE. After studying the results, protocol C was chosen for protein extraction. The sample from one volunteer was used for 2-D PAGE in order to generate three gels (17 cm-strips) with pH values in the range 3-10 and another three gels with pH values of between 4-7 (13 cm-strips). Two biological replicas were processed, but only the gels from one individual are shown in Figure 3. The sample protein fraction was applied to IPG gels, using 250 µL for hydration of the 13 cm-strips and 300 µL for hydration of the 17 cm-strips, and then incubated for 12 h.

IEF on the 17 cm gel strips was performed using PROTEAN® IEF Cell equipment (Bio-Rad, Hércules, California) at CENICAFE laboratories (Chinchiná, Caldas, Colombia) using a program with four steps: 1) constant 500 V for 500 Vh; 2) a gradient at 1000 V for 800 Vh; 3) a gradient at 1000 V for 16500 Vh, and 4) constant 1000 V for 3700 Vh. IEF on the 13 cm gels was conducted at the GEPAM laboratory (Unicamp, Campinas, Brazil) using Amersham Biosciences equipment (GE Healthcare, model 18113002 EPS601). The program used was the same as that for 17 cm gels, except for step 3, which was performed at 11300 V h. The IEF process was optimized in the laboratory. Once the sample was equilibrated with reducing and alkylating agents, the second dimension was carried out in a 12% (m/v) polyacrylamide; the final concentration of the gels was: 12% Acrilamida/Bis, 0.37 M Tris, 0.1% SDS (Sodium dodecyl sulfate), 0.05% APS (Ammonium Persulfate) and 0.05%TEMED. All reagents were acquired from Bio–Rad Laboratories Hércules, California, USA. The power source was operated in two stages under the following running conditions: 1) 30 min, 15 mA/gel at 90 V, and 2) 4-5 h 25 mA/gel at 600 V. Finally, the gels were stained with either Coomassie colloidal or Sypro Ruby (Bio–Rad Laboratories Hércules, California, USA) and scanned in a Pharos FX Molecular Imager (Bio-Rad, Hércules, California, USA) at CENICAFE laboratories, as well as by an ImageScanner TM II (GE healthcare, Sweden) at the GEPAM laboratory.

Table I: Protein extraction protocols evaluated for 2-D PAGE of HIM

Protocols	Brief description
A	Protein precipitation using acetone, 10% v/v trichloroacetic acid and 0.007% DTT, and incubation during 24h at -8 ºC.
B	[18]
C	[19] utilizing a denaturing buffer (8 M urea, 2% (m/v) CHAPS, 1% (m/v) DTT, 50 mM Tris pH 7.0), PMSF, cleaning with Protein Clean Up Kit de Bio-Rad (California, USA) and acetone.
D	[18] and [19] utilizing a pool of 3 samples.
E	[19] utilizing the chemical M-PER for extraction and 300 µg of sample.
F	[19] utilizing the chemical M-PER for extraction and 100 µg of sample.
G	[20] for glucosidases denaturation and separation of small proteins.
H	Protocol C and double cleaning with Protein Clean Up Kit de Bio-Rad (California, USA).
I	[21]

Table I: Protein extraction protocols evaluated for 2-D PAGE of HIM (Continuation)

Protocols	Brief description
J	Protein precipitation with ammonium sulfate. Precipitation of 300 µL of cleaned mucus sample with ammonium sulfate 0 – 20%. Precipitated incubated overnight. 300 µL of sample clean mucus is precipitated with ammonium sulphate ($(NH_4)_2SO_4$) in the range of 0 to 20% saturation (0- 20%), then allowed to stand overnight and centrifuged at 14548 *g* for 5 minutes. The supernatant of 20% saturation was precipitated again to 60% saturation (20-60%), repeating the process; two precipitates (ranges 20-60% 0-20% and saturation) comprising 200 µL diluted in hydration buffer containing 7 M urea, 2 M thiourea, 2% CHAPS, 50 mM DTT and 1% ampholytes were then finally obtained.
K	[21] and microdialysis.
L	Protein precipitation with metasulfonic acid (MSA). 300 µL of 1 M MSA was added to 300 µL of clean mucus sample, and was allowed to precipitate on ice for 40 min. The sample was centrifuged at 14,548 *g* for 15 min. Then it was cleaned using the Clean Up Kit (Bio-Rad - California, USA), and finally resuspended in 200 µL of hydration buffer which contains 7 M urea, 2 M thiourea, 2% CHAPS, 50 mM DTT and 1% ampholytes.
M	Acid hydrolysis and microdialysis. 62 µL of 1M HCl is added to 300 µL of clean mucus sample, to pH 4.5. Subsequently microdialysis is performed with a membrane pore size 3 kDa for 12 h against deionized water. Finally, the sample was centrifuged at 14,548 *g* for 5 min, the supernatant is discarded and 100 to 200 µL of hydration buffer added which contains 7 M urea, 2 M thiourea, 2% CHAPS, 50 mM DTT and 1% ampholytes.
N	[19] utilizing a denaturing buffer, PMSF and triple cleaning with acetone.
O	[19] utilizing a denaturing buffer, PMSF and triple cleaning with acetone and TCA for precipitating the proteins.

Shotgun proteomics

A sample from a single volunteer was used for this experiment, giving one biological replica which was processed twice: 150 µL of acetonitrile – ACN (JT Baker, Pennsylvania, USA) at 53% (v/v) [22] was added to 100 µL of extracted protein, in addition to a denaturing buffer and the absence of any cleaning treatment. This preparation was stirred for one minute and placed in an ultrasonic bath for 10 min. The process was repeated once and the solution was then centrifuged at 12000 *g* for 10 min. A digestion with trypsin (Promega Corporation Madison, Wisconsin, USA) was carried out in the supernatant and the precipitate was eliminated [23]. Following [22], the precipitate obtained with acetonitrile was eliminated as a depletion method of abundant proteins, especially albumin.

Because proteins did not undergo any cleaning treatment, peptides were desalted using a C_{18}-solid phase extraction cartridge (Allcrom 8B-S001-EAK). The cartridges were activated with 3 mL ACN 100%, equilibrated with 1 mL 50/50 ACN/H_2O with 0.1% (v/v) formic acid (JT Baker Center Valley, Pennsylvania, USA) and then loaded in 3 mL of 0.1% (v/v) Trifluoroacetic acid – TFA (Carlo Erba, Paris, France). Desalting was carried out after sample loading. The sample as washed with 3 mL 0.1% (v/v) TFA and equilibrated with 1 mL 0.1% (v/v) formic acid. The sample was eluted in 2 mL 50/50 v/v ACN/H_2O with 0.1% (v/v) formic acid and 1 mL 80/20 v/v ACN/H_2O.

Mass spectrometry for protein identification

Samples were concentrated to *ca*. 50 µL using a Speed Vac Concentrator (model SPD131DDA, Thermo Scientific). The peptides in the digested samples were analyzed using a UPLC nanoAcquity Waters chromatograph coupled to a Waters SYNAPT HDMS spectrometer equipped with a nanoESI source. A

total of 2 to 5 µL of the aqueous sample was injected using the UPLC autosampler and placed in a guard pre-column (Waters Symmetry C_{18}, 20 mm x 180 µm) for desalination with a 5 µL/min flow of 97/3 v/v H_2O/acetonitrile (0.1% v/v formic acid) for 3 minutes. The sample was then transferred to an analytical column (Waters, BEH130 C_{18}, 100 mm x 100 µm i.d., 1.7 µm particles) and eluted with a 1 µl/min flow of formic acid. The peptides were detected on line using a mass spectrometer configured to operate in data dependent acquisition (DDA) mode using an MS function (*fullscan* of *m/z* 200 to 2000), three MS/MS functions, and an external calibration curve function (*lockmass*). The spectrometer operated under the following parameters: 3 kV capillary voltage, 30 V cone voltage, 100 °C source temperature, 0.5 L/h nanoESI gas flow, 6 and 4 eV collision energy for the Trap and Transfer cells, respectively, and 1700 V detector. The MS (*fullscan*) and MS/MS (spectrum of ions fragmented by collision-induced dissociation) spectra were acquired at a rate of 1 spectrum/s. The instrument was calibrated before the analysis using phosphoric acid oligomers (0.5% v/v H_3PO_4 solution in 50/50 v/v ACN/H_2O) with an *m/z* from 90 to 1960. Argon at 9.7 10^{-3} mbar was used as the collision gas. The LC-MS runs were processed using ProteinLynx Global Server v.2.2 software (Waters) and all mass spectra were analyzed using a peak list format in the MASCOT program v.2.2 (Matrix Science, London, UK). The selected search parameter was: digestion ignoring trypsin until reaching a cleavage site, oxidation (M) or Methionine oxidation and, then, labelling them because there are oxidant conditions during sample preparation, a variable modification corresponding to those modifications that may or may not be present. For the specific case of M oxidation, MASCOT explains that, when a peptide contains more than 3 methionines, the coincidence with the experimental data has to be evaluated for that peptide that contains 0, 1, 2 and 3 residuals of oxidized Methionine. Precursor and fragment mass tolerance of ±0.1 Da was also used as modification variables. For the searches, the SwissProt database (available since 1986) was used. Identification was based on the high specificity of the combination of the 20 essential aminoacids in sequences of peptides with 6 or more aminoacids. SwissProt is characterized because it contains entries that show contrasted, experimental evidence that has been manually reviewed by experts.

RESULTS AND DISCUSSION

Sample cleaning

Buffering and centrifugation have been previously implemented [15] for cleaning of intestinal mucus but there is no indication of the effectiveness of the process. In our study, Gram staining was employed as a simple methodology to verify the cleanliness of the sample. For proteomic studies of HIM, besides the quantity of sample, an effective cleaning process is of paramount importance. With this approach, only the mucus forming proteins are extracted, otherwise the sample will be contaminated with proteins from food, faeces, microbiota and external cells. A micrograph of Gram stained bacteria from HIM is shown in Figure 1A, before the sample had undergone the bacterial cleaning process; Figure 1B corresponds to conditions after performing the bacterial cleaning process, thus indicating a sample completely devoid of bacteria.

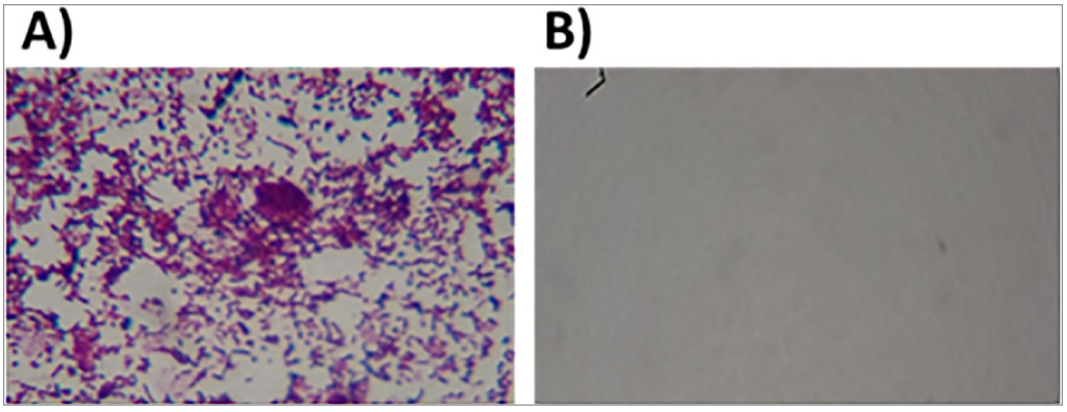


Figure 1. Elimination of bacteria from mucus sample. Gram stain of: A) the upper part of the supernatant, rich in bacteria, and B) the lower part of the supernatant, devoid of bacteria.

Sodium dodecyl sulphate polyacrylamide gel electrophoresis and two-dimensional gel electrophoresis

According to the results obtained with the patients involved in this study, only 5 to 150 µL of mucus sample can be obtained from a single person during colonoscopy. In fact, this amount of material is not sufficient for evaluation by 2-DE, even using staining reagents like colloidal coomassie, which improves the detection of proteins [24]. The protein quantification showed concentrations ranging from 0.65 to 1.12 µg/µL [25]. On the other hand, a single session of transanal irrigation with one patient can provide from 0.5 to 5 mL of HIM, with protein concentrations ranging from 20.0 to 22.4 µg/µL [25], after quantification. This concentration is enough for analysis by SDS-PAGE (Figure 2) or by 2-D PAGE (Figure 3). In this way, and using the material obtained through this sampling process, Figure 2 shows the SDS-PAGE for protein separation from mucus samples. Although this separation was carried out from three different patients, good similarities in terms of protein distribution are observed (similar protein weight), as well as the good efficiency in the sample cleaning process. After protein separation, all processes for protein identification through MS/MS were conducted (as indicated in section 'Mass spectrometry for protein identification') and supplementary material Table 1 summarizes the list of proteins identified using SDS-PAGE.

As the sampling strategy was efficient in terms of quantity of sample, a proteomic map was also obtained using 2-D PAGE based on samples from the same two patients. Then, gels presenting pH ranges from 3 to 10 and from 4 to 7 were used for protein identification according to procedures presented in section 'Two-dimensional gel electrophoresis'. However, 15 different protocols were tested with the objective of standardizing protein separations from HIM. All the protocols are visualized in Table I. In this way, using protocol A, only poor protein precipitation and partial dilution of the mucus were obtained and the gel presents a bad resolution with only 15 spots observed. On the other hand, protocol B presents a good resolution for protein identification. Even better was protocol C, because it presents the best resolution (*ca*. 220 spots, see Figure 3), requiring *ca*. 8 h for the run. Protocol G presents *ca*. 55 spots but a long time for running the gel. Similar behavior was observed by employing protocol I (only 12 spots), and too much streaking was observed in the gel. No or a minimal amount of proteins were observed employing the other protocols (D – F, H, J – O).

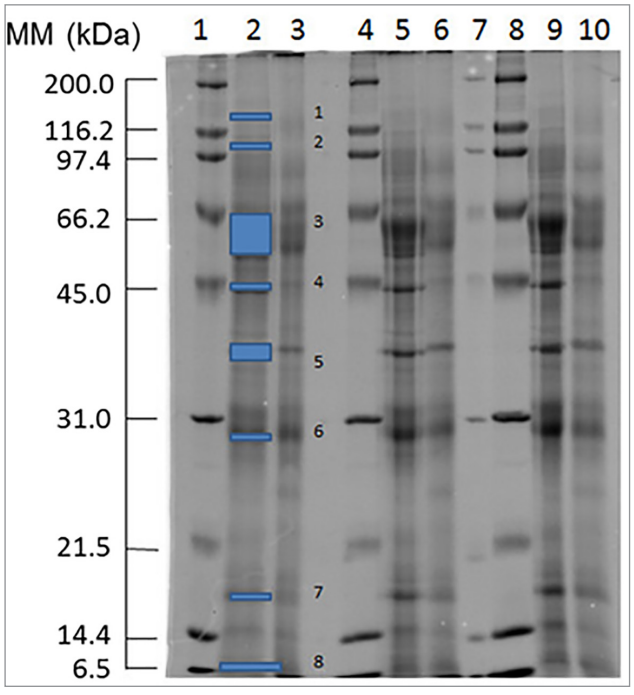


Figure 2. Sodium dodecyl sulphate polyacrylamide gel electrophoresis made with the mucus samples obtained from 3 patients. The gel is 17 cm and stained with SYPRO ruby. Lanes 1, 4, 7 and 8 are molecular weight standards (Bio-Rad, Hércules, California, USA). Lanes 2 and 3 are from volunteer 1, lanes 5 and 6 from volunteer 2, and lanes 9 and 10 correspond to volunteer 3. Blue bands indicate the bands analyzed by mass spectrometry.

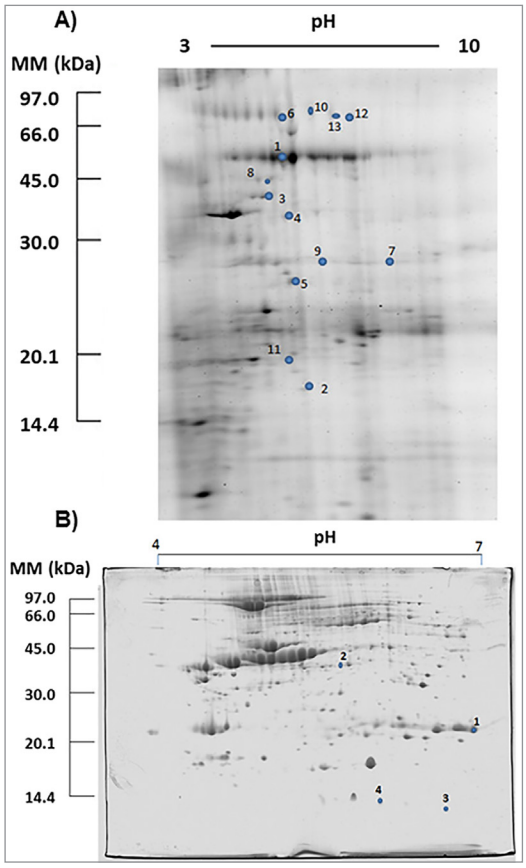


Figure 3. Two-dimensional gel electrophoresis of human intestinal mucus. A) Gel with a pH range from 3 to 10, stained with SYPRO ruby. A total of 116 spots were visually identified with samples from one volunteer. B) Gel presenting a pH range from 4 to 7, stained with Coomassie blue. A total of 220 spots were visually identified with samples from one volunteer. Blue spots indicate the spots analyzed by mass spectrometry.

Protocol C was identified as the best for 2D-PAGE and the image of the best gel is shown in Figure 3. In fact, 116 spots were counted between 3 and 10, while 220 were present between 4 and 7 pH ranges. Additionally, the identified proteins can be visualized in Tables 2 and 3 in supplementary materials for gels from 3-10 and 4-7 pH ranges, respectively.

Finally, it is noteworthy that all steps involved in protein analysis must be conducted without stopping.

Shotgun proteomics

Due to the difficulty in solubilizing some proteins, even using chemicals for improving their solubilization [1], as well as the inherent limitation of electrophoretic techniques in separating large molecular weight proteins, shotgun proteomics was selected for examining these problems.

Identified proteins

Some preliminary considerations

Figure 4 presents a Venn diagram showing the overlap between the proteins identified by all the strategies employed in this work (*i.e.* SDS-PAGE; 2-D PAGE and shotgun), indicating the need for such strategies for identification due to their complementarities. It seems that a single technique is not suitable in order to study the largest number of the HIM proteins. It is important to comment that most of the identified proteins present in HIM are cellular proteins. This may be a consequence of the renovation of the cellular epithelium, also observed in the mouse colon mucus, or that such proteins have a specific function in the mucus. However, at present, there is insufficient information to provide a complete explanation of this fact. Besides cellular proteins, plasma proteins were also found. Some authors [17] suggest that the permeability of the intestine allows the passage of the extracellular liquid, or that such proteins are natural components of mucus and possibly have fundamental functions associated with the microbiota. As an example, in the cervical mucus these proteins are abundantly observed [17].

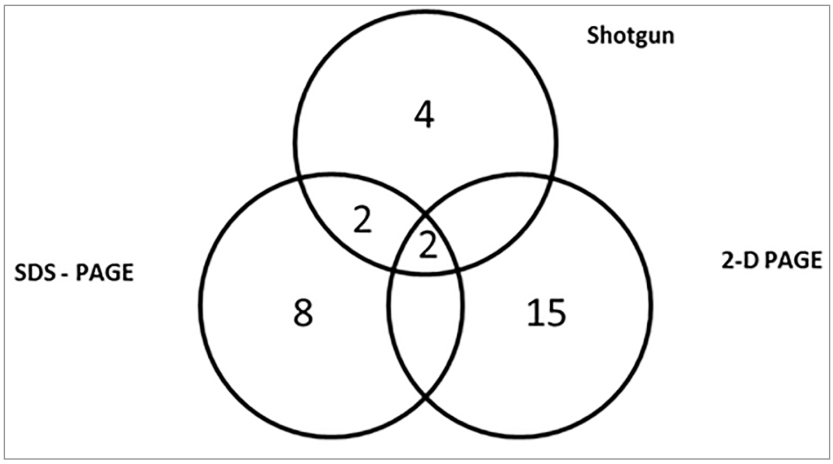


Figure 4. This Venn diagram shows the overlap of proteins identified for three separation/identification methods.

From the 31 proteins identified, 27 of them come from *Homo sapiens*, 3 from *Escherichia coli* and 1 from *Saccharomydes cerevisiae*. *E. coli* is a gram-negative intestinal bacterium which participates in the digestion process, although it can produce some intestinal problems. The *S cerevisiae* is an anaerobic unicellular fungus, which could be trapped in the mucus during ingestion of food. It is evident from the literature [12] and the HIM obtained through transanal irrigation, that it contains proteins from luminal (loose) and mucosa (tight) layers, as identified in Table II.

Table II. Proteins identified in the loose (L) and tight (T) layers of the colon mucus.

Protein	L	T	Remarks
Calcium activated	x	x	
Protein disulfide-isomerase	x	x	
ATP synthase subunit beta, mitochondrial precursor	x	x	
Actin, cytoplasmic 1		x	Identify from *Dreissena polymorpha*
IgA heavy 2	x	x	
Malate dehydrogenase mitochondrial	x	x	
Galectin-4			Protein found after the reduction treatment with GuHCl
Annexin 2	x	x	
Ig kappa light chain VLJ region	x		
Zymogen granule membrane protein 16 precursor	x	x	
Mitogen-activated protein kinase scaffold protein 1		x	
Hemoglobin subunit beta A	x	x	
Glutamine amidotransferase [Escherichia coli]	x	x	
Ubiquitin	x	x	
Polymeric immunoglobulin receptor	x		
MUC2	x	x	Protein found after the reduction treatment with GuHCl

About the identified proteins

To make this topic more relevant, only those proteins which may have some clinical interest are discussed. Additionally, all the proteins identified were classified according to the biological processes in which they are involved.

Transport and motility

Proteins identified within this group were: hemoglobine subunit alfa, calcium-activated chloride channel protein 1, Actin cytoplasmic 1, and Zymogen granule membrane protein 16 or Secretory lectin ZG16. The calcium-activated chloride channel protein 1 is one of the proteins most present in the intestinal epithelium [26], and has an important function in the regulation of production and/or secretion of mucus when considering some health problems, such as metaplasia and cystic fibrosis, among others [27], [28]. Additionally, it also acts as a signaling agent between the intestinal environment and some immune perturbations, such as in inflammatory bowel disease. Finally, it is a tumor suppressor and induces the production of the Muc5C [29]. The actin cytoplasmic 1 and actin aortic smooth muscle are proteins involved in a diversity of cellular motility functions; for example, the actin cytoplasmic 1 participates in the secretion of gastric acid. The zymogen granule membrane protein 16 or secretory lectin ZG16 recognizes some pathogens' fungus, such as *Candida* and *Malassezia* species in the digestive system. The iron regulatory protein, Fur, is present in the microbiota, participating as a cofactor in iron (as Fe^{2+}) transport.

Immune response

The Ig A, Ig alpha-2 heavy chain, polymeric immunoglobulin receptor, Ig kappa light chain VLJ, Ig lambda chain C, and Chain A Crystal structure of human calcineurin complexed with cyclosporin A were identified within this group. The chain A, crystal structure of human calcineurin complex with cyclosporine A and human cyclophilin, are calcium receptor enzymes which participate in different cellular regulatory ways. The cyclosporine is an immunosupressor drug that complexes with cyclophilin. Both can also be associated with the calcineurin for inhibiting its catalytic activity [30]. It is curious that one of the patients remembers that when she was 15 years-old (she was 45 when the mucus sampling was carried out) the doctor prescribed cyclosporine for her. This result is something peculiar, because it highlights the protective importance of the intestinal mucus in trapping some waste substances.

Metabolism

The proteins identified within this group included: Glutamine-fructose-6-phosphate aminotransferase, and Bifunctional UDP-glucuronic acid decarboxylase/UDP-4-amino-4-deoxy-L-arabinose formyltransferase. The latter, for instance, confers the capacity of accepting antibiotics by mutations to bacteria that hampers the bond between the antibiotic and the protein, or by gene amplification, which codifies the protein. It also participates in lipid metabolism.

DNA damage response

Polyubiquitin, ubiquitin, and TIP41-like protein were those identified associated with this group. Regarding the TIP41-like protein, it can be an allosteric regulator of serine / threonine protein phosphatase 2A. Isoforms may play a role in regulating the signaling pathway ATM / ATR and control in DNA replication and repair.

Cellular response

Within this group, mitogen-activated protein kinase scaffold protein 1 or Regulator complex protein LAMTOR3 and UPF0539 protein C7orf59, also known as Regulator complex protein LAMTOR4, were the proteins identified. LAMTOR3 and LAMTOR4 participate in the regulator complex which functions as a factor in guanine nucleotide exchange and activation of small GTPases Rag. In a recent study [31], LAMTOR3 was identified as participating in cellular responses induced by gastrin which is involved, in turn, in proliferation and homeostasis of gastric mucosa.

Angiogenesis

Within this group, the following proteins were identified: ATP synthase subunit beta, mitochondrial precursor, and annexin A2. This last protein is involved in the regulation of epithelial cell migration. When an intestinal lesion happens, these cells migrate to seal the wound. Its action is carried out by controlling the levels of B1 integrin protein, which is responsible for allowing the interaction between the epithelial cells and the matrix [32]. Annexin A2 was evaluated as a biomarker in serum for early diagnosis of hepatocellular carcinoma, which is related to the virus for hepatitis B [33].

Enzymatic activity

Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform 1, and Chain A, crystal structure of the 20s proteasome were the proteins constituting this group. The last one is a proteinase complex whose function is to carry out the proteolysis of those unnecessary or damaged proteins localized in the cytoplasm or in the nucleus.

Cellular differentiation and growth

Within this group the nucleolar protein 8 was found. This protein plays an important role in the apoptosis of gastric cancer cells [34]. It can also be involved in the genic expression regulation at post-transcriptional level or in the biogenesis of ribosomes in cancer cells.

Cell adhesion

Galectin-4, and Mucin 2 (MUC2) were identified in this group, Mucin, the target of this work, was described and commented on in the introduction. Galectin-4 has a similar function to gelactin-3, which is related to human colon cancer [35]. A higher abundance of galectin-3 was related to the malignant behavior of colon cancer cells [34]. However, it is still not clear how this protein affects the intracellular signal pathway.

Other functions

Within this group the ferric uptake regulator was identified. This protein [*Escherichia coli* O157: H7EDL933] is grouped with the biological process of transcription and uses Fe^{2+} as a cofactor that binds to operator repressed genes and regulates the expression of several proteins of the outer membrane.

Unknown functions

One protein was identified inside this group: the Scll+ suppressor protein [*Saccharomyces cerevisiae*] is a dominant suppressor of cycloheximide resistance.

Interactions between some proteins identified

When considering protein-protein interactions (Figure 5) using the STRING program, from the total amount of proteins identified, 23 are correlated. From these proteins, a good correlation between their functions is then noted, and at least two major clusters observed. Firstly, in the center of Figure 5 is Ubiquitin, a protein that mediates transcriptional activation of target genes and plays an important role in controlling the progress of cell cycle and differentiation, in the error-free DNA repair pathway, and in the survival of cells after DNA damage. Secondly, on the left side of this figure, both ATP synthase (ATP5B) and malate dehydrogenase (MDH1) were correlated, belonging to a cluster in which a diversity of ATP synthases is present. In fact, the ATP synthases are involved in the production of ATP from ADP in the presence of a proton gradient across the membrane, which is generated by electron transport complexes in the respiratory chain. As this enzyme is involved in energy production, its correlation with the synthesis of glucose through MDH is then correct.

Some other proteins identified (see Tables 1-4 supplementary material), such as annexin A2 (ANXA2), Gelactin 4 (LGALS4), calcium binding protein 1 (CABIN1) and others, are not connected to each other, but

play important roles regarding biological processes, as described in section about the identified proteins. Looking at the right side of Figure 5, another cluster is noted in which mitogen-activate protein kinase 1 (MAPK1), crystalline alpha B (CRYAB), ferritin (FTH1) and actin (ACTA1) are present. The presence of such proteins denotes the link between the clusters: while the cluster on the left side is involved in energy storage, the cluster on the right hand side is involved in energy consumption due to the work developed by the biochemical machinery in relation to cell growth, adhesion, survival (MAPK1), motility (ACTA1) and iron uptake (FTH1).

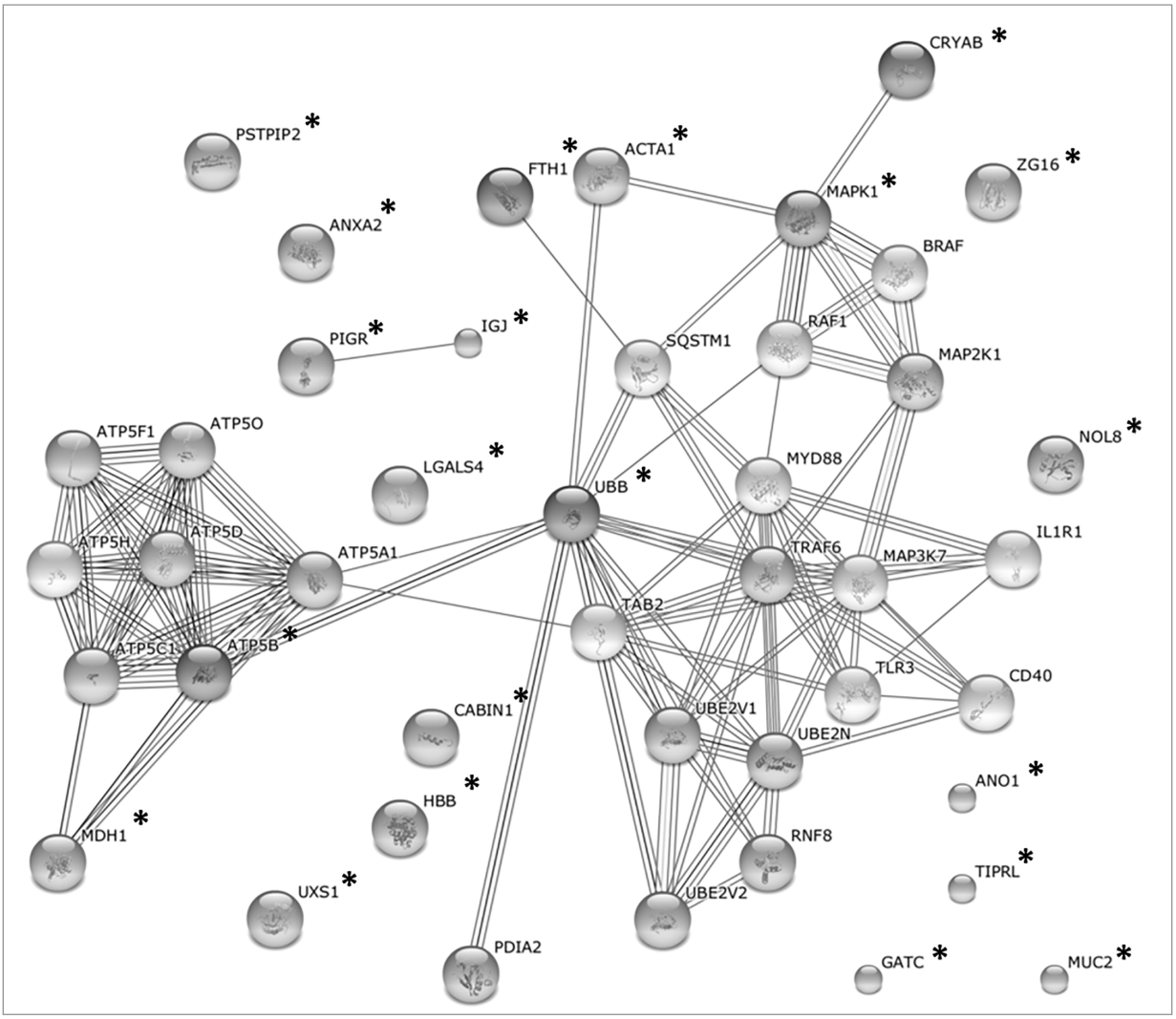


Figure 5. Protein interaction network generated and visualized with STRING 9.1 program for proteins identified in our study. Only those proteins identified in this study are namedand marked with an asterisk in this Figure. The others are inserted into the program in order to compose the network. Center: UBB - Ubiquitin B; Left side: ATP5B – ATP synthase; MDH1- Malate dehydrogenase; UXS1 – UDP glucuronic acid descarboxylase; HBB – hemoglobin; CABIN1 – calcinerium binding protein 1**;** LGALS4 – Lectin; PIGR – Polymeric immunoglobulin receptor; IGJ – Immunoglobulin J peptide; ANXA2 – Annexin A2; PSTPIP2 – Proline-serine-threonine phosphatase; Right side: GATC – glutamyl-tRNA amiditransferase; MUC2 – Mucin 2; TIPRL – TIP41-like protein; ANO1 – calcium activated chloride channel; NOL8 – Nucleolar protein 8; ZG16 – zymogen granule protein 16; CYAB – Crystallin alpha B; MAPK1 – mitogen-activated protein kinase; ACTA 1 – actin; FTH1 – ferritin.

Most probably this “hard work” developed by the organism may give rise to some stress conditions, thus explaining the presence of CRYAB in this cluster, as one of its functions is just to prevent aggregation of various proteins under a wide range of stress conditions. Some other identified proteins, such as Mucin 2 (MUC2), zymogen granule protein 16 (ZG16), calcium-activated chloride channel (ANO1) and others, are also present, but apparently with no correlation to that cluster. However, some of them are proteins which also participate in the process of energy consumption, such as gastrointestinal smooth muscle contraction (ANO1), protein trafficking (ZG16), and exclusion of bacteria from the inner mucus layer (MUC2), among others.

Finally, according to the analysis of the identified proteins and their correlations, some of the most important biological processes are then identified (as indicated in the section about the identified proteins) and correlated with each other.

CONCLUSIONS

The main objectives of this study were successfully achieved. By using a sampling method based on transanal irrigation and some strategies involved in analytical proteomics, the first, and most detailed, proteomic human intestinal mucus map was obtained. It is necessary to note the extreme difficulty in working with such samples, especially when attaining accurate identification of those proteins belonging exclusively to HIM. In this way, all analytical steps for treating the samples were efficiently performed, producing a selective proteomic map, free of those proteins from food, faeces and microbiota. With all the strategies applied here, a good number of proteins was then identified using SDS-PAGE, 2-D PAGE and shotgun analysis, and 9 of them related with other studies, playing an important role in some diseases, particularly considering those related to the digestive system. The majority are secreted proteins, making their purification possible for future studies. This study contributes to a better understanding of the HIM proteome through a methodology for obtaining, treating, separating and identifying its proteins. Additionally, it provides a methodology for analysis of the HIM proteins, without protein separation. This includes their extraction, desalination, depletion and digestion for MUC2 identification, the main protein of HIM, generally separated by SDS-agarose or ultracentrifugation [36]. In summary, the content of this work should facilitate further studies of complex bodily fluids, such as HIM, as well as the identification of the differential proteins in a variety of pathological conditions that could be used as therapeutic agents or biomarkers.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

The authors thank: a) In Colombia: The University of Caldas, Cenicafé and COLCIENCIAS, for their partial financial support and permission to use their equipment and facilities; b) In Brazil: “Fundação de Amparo a Pesquisa do Estado de São Paulo” (FAPESP, São Paulo, Brazil), the “Conselho Nacional de Desenvolvimento Científico e Tecnológico” (CNPq, Brasília, Brazil), and the “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES, Brasília, Brazil) for financial support and fellowships.

REFERENCES

1. Johansson, M. E.; Phillipson, M.; Petersson, J.; Velcich, A.; Holm, L.; Hansson, G. C. *Pro. Natl. Acad. Sci.,* **2008**, *105* (39), pp 15064–15069 (https://doi.org/10.1073/pnas.0803124105).
2. Gum, J. R.; Hicks, J. W.; Toribara, N. W.; Siddiki, B.; Kim, Y. S. *J Biol Chem.,* **1994,** *269* (4), pp 2440–2446.
3. Matsuo, K.; Ota, H.; Akamatsu, T.; Sugiyama, A.; Katsuyama, T. *Gut.,* **1997**, *40* (6), pp 782–789.
4. Sommer, F.; Adam, N.; Johansson, M. E. V.; Xia, L.; Hansson, G. C.; Bäckhed, F. *PLoS One,* **2014**, *9* (1), e85254 (https://doi.org/10.1371/journal.pone.0085254).
5. Blaut, M.; Collins, M. D.; Welling, G. W.; Doré, J.; Van Loo, J.; de Vos, W. M. *Br. J. Nutr.,* **2002**, *87* (S2), pp S203-S211 (https://doi.org/10.1079/BJN/2002539).

6. Tennyson, C. A; Friedman, G. *Curr Opin Endocrinol Diabetes Obes.,* **2008**, *15* (5), pp 422–427.
7. Tamboli, C. P; Neut, C.; Desreumaux, P.; Colombel, J. F. *Gut.,* **2004**, *53* (1), pp 1–4.
8. Johansson, M. E. V.; Sjövall, H.; Hansson, G. C. *Nat Rev Gastroenterol Hepatol.,* **2013**, *10* (6), pp 352–361 (https://doi.org/10.1038/nrgastro.2013.35).
9. Ouwerkerk, J. P.; De Vos, W. M.; Belzer, C. *Best Pract Res Clin Gastroenterol.,* **2013**, *27* (1), pp 25–38.
10. Bäckhed, F.; Ley, R. E.; Sonnenburg, J. L.; Peterson, D.; Gordon, J. I. *Science*, **2005**, *307* (5717), pp 1915–1920 (https://doi.org/10.1126/science.1104816).
11. Hansson, G. C. *Curr Opin Microbiol.,* **2012**, *15* (1), pp 57–62 (https://doi.org/10.1016/j.mib.2011.11.002).
12. Johansson, M. E. V.; Thomsson, K. A.; Hansson, G. C. *J Proteome Res.,* **2009**, *8* (7), pp 3549–3557.
13. Du, L. Y.; Qian, D. W.; Shang, E. X.; Jiang, S.; Liu, P.; Guo, J. M.; Su, S. L.; Duan, J.; Xu, J.; Zhao, M. *Anal Methods.,* **2015**, *7* (13), pp 5574–5582 (https://doi.org/10.1039/C5AY00931F).
14. Licea-Perez, H.; Junnotula, V.; Zohrabian, S.; Karlinsey, M. *Anal Methods,* **2016**, *8* (15), pp 3023–3033.
15. Hamer, H. M.; Jonkers, D.; Loof, A.; Vanhoutvin, S. A. L. W.; Troost, F. J.; Venema, K.; Kodde, A.; Koek, G.; Schipper, R.; van Heerde, W.; et al. *Dig Liver Dis.,* **2009**, *41* (8), pp 559–564.
16. Rodríguez-Piñeiro, A. M.; Post S, V. D.; Johansson, M. E. V.; Thomsson, K. A.; Nesvizhskii, A. I.; Hansson, G. C. *J Proteome Res.*, **2012**, *11* (3), pp 1879–1890 (https://doi.org/10.1021/pr2010988).
17. Panicker, G.; Ye, Y.; Wang, D.; Unger, E. R. *Clin Proteomics.,* **2010**, *6* (1-2), pp 18–28.
18. Dellios, N. L.; Lappas, M.; Young, I. R.; Palliser, H. K.; Hirst, J. J.; Oliva, K.; Ayhan, M.; Rice, G. *Eur J Obs. Gynecol Reprod Biol.*, **2010**, *153* (1) pp 16–22 (https://doi.org/10.1016/j.ejogrb.2010.06.014).
19. Panicker, D. R.; Lee, E. R.; Unger, E. *J Proteomics*., **2009**, *71* (6) pp 637–646.
20. Bovee-Oudenhoven, I.; Termont, D.; Dekker, R.; Van der Meer, R. *Gut*., **1996**, *38* (1) pp 59–65.
21. Nice, E. C., Fabri, L.; Whitehead, R. H.; James, R.; Simpson, R. J.; Burgess, A. W. *J Biol Chem.,***1991**, *266* (22) pp 14425–14430.
22. Kay, R.; Barton, C. *Rapid Commum Mass Spectrom.,* **2008**, pp 3255–3260.
23. Ribeiro, D.; Cota, J.; Alvarez, T. M.; Brüchli, F.; Bragato, J.; Pereira, B.; Pauletti, B.; Jackson, G.; Pimenta, M.; Murakami, M.; et al. *PLoS One.,* **2012**, *7* (12), e50571 (https://doi.org/10.1371/journal.pone.0050571).
24. Dyballa, N.; Metzger, S. *J. Vis. Exp.,* **2009**, (30), e1431 (https://doi.org/10.3791/1431).
25. Gómez-Buitrago, P-A.; González-Correa, C-A.; Santacoloma-Osorio, M.; Taborda-Ocampo, G.; Zezzi-Arruda, M. A. *Rev Colomb Quim.,* **2014**, *43* (3), pp 5–10 (http://dx.doi.org/10.15446/rev.colomb.quim.v43n3.53586).
26. Nyström, E.; Arike, L.; Ehrencrona, E.; Hansson, G. C.; Malin, E. V. *J. Biol. Chem.,* **2019**, *1*, pp 1–25.
27. Hoshino, M.; Morita, S.; Iwashita, H.; Sagiya, Y.; Nagi, T.; Nakanishi, A.; Ashida, Y.; Nishimura, O.; Fujisawa, Y.; Fujino, M. *Am J Respir Crit Care Med.,* **2002**, *165* (8), pp 1132–1136.
28. Qiu, J.; Heller, J.; Guo, X.; Chen, Z.; Fish, K.; Fu, Y.; Zhou, L. *Immunity.,* **2012**, *36* (1), pp 92–104.
29. Bustin, S.; Li, S.; Dorudi, S. *DNA Cell Biol.,* **2001**, *20* (6), pp 331–338.
30. Jin, L.; Harrison, S. *Proc Natl Acad Sci.,* **2002**, *99* (21) , pp 13522–13526.
31. Steigedal, T.; Prestvik, W.; Selvik, L.; Fjeldbo, C.; Bruland, T.; Laegreid, A.; Thommesen, L. *In Vitro Cell. Dev. Biol. - Animal,* **2013**, *49* (3), pp 162–169 (https://doi.org/10.1007/s11626-013-9588-2).
32. Rankin, C.; Hilgarth, R. S.; Leoni, G.; Kwon, M.; Den Beste, K.; Parkos, C.; Nusrat, A. *J Biol Chem.,* **2013**, *288* (21), pp 15229–15239 (https://doi.org/10.1074/jbc.M112.440909).
33. Sun, Y.; Gao, G.; Cai, J.; Wang, Y.; Qu, X.; He, L.; Liu, F.; Zhang, Y.; Lin, K.; Ma, S.; et al. *Carcinogenesis,* **2013**, *34* (3), pp 595–604.
34. Jinawath, N.; Furukawa, Y.; Nakamura, Y. *Cancer Sci.,* **2004**, *95* (5), pp 430–435.
35. Kim, D. W.; Kim, K. H.; Yoo, B. C.; Hong, S. H.; Lim, Y. C.; Shin, Y. K.; Park, J. *Cancer Sci.,* **2008**, *99* (10), pp 1884–1891.
36. Holmén, J. M.; Karlsson, H.; Sjövall, H.; Hansson, G. C. *Glycobiology,* **2009,** *19* (7), pp 756-766 (https://doi.org/10.1093/glycob/cwp048).

Supplementary Material

Table 1. Proteins identified of HIM through SDS-PAGE

Line	Protein name	Access	Theoretical MM(Da)	Experimental MM(Da)	Score	% Coverage	Peptides	# Peptides
4	Actin, cytoplasmic 1; AltName: Full=Beta actin.	gi|113271	42169	42200	868	55	K.AGFAGDDAPR.A, R.AVFPSIVGRPR.H, R.HQGVMVGMGQK.D, K.DSYVGDEAQSK.R, K.DSYVGDEAQSKR.G,K.IWHHTFYNELR.V, R.VAPEEHPVLLTEAPLNPK.A, R.TTGIVMDSGDGVTHTVPIYEGYALPHAILR.L, R.LDLAGRDLTDYLMK.I, R.DLTDYLMK.I,K.ILTERGYSFTTTAER.E, R.GYSFTTTAER.E,R.GYSFTTTAEREIVR.D, K.SYELPDGQVITIGNER.F, K.DLYANTVLSGGTTMYPGIADR.M, R.MQKEITALAPSTMK.I, K.EITALAPSTMK.I,K.IIAPPER.K, K.IIAPPERK.Y,K.QEYDESGPSIVHR.K, K.QEYDESGPSIVHRK.C	21
5	Annexin A2.	gi|18645167	38784	38500	288	24	K.AYTNFDAER.D,R.DALNIETAIK.T, K.GVDEVTIVNILTNR.S, K.SALSGHLETVILGLLK.T, K.TPAQYDASELK.A,R.TNQELQEINR.V, K.SYSPYDMLESIR.K	7
3	ATP synthase subunit beta, mitochondrial precursor.	gi|32189394	56525	56000	1254	59	K.VLDSGAPIK.I, K.TVLIMELINNVAK.A, K.IGLFGGAGVGK.T,K.IPVGPETLGR.I, K.VVDLLAPYAK.G,R.TIAMDGTEGLVR.G, R.IMNVIGEPIDER.G,K.AHGGYSVFAGVGER.T, R.FTQAGSEVSALLGR.I,R.VALTGLTVAEYFR.D, K.GFQQILAGEYDHLPEQAFYMVGPIEEAVAK.A	13
2	Calcium-activated chloride channel protein 1.	gi|4585469	100894	103000	212	10	K.LAEYGPQGR.A, R.IHLTPDFIAGK.K, K.LAEYGPQGR.A, R.WGVFDEYNNDEK.F, K.TTPMTTQPPNPTFSLLQIGQR.I, K.MAYLQIPGIAK.V, K.TVTLELLDNGAGADATK.D	9
5	Galectin-4.	gi|5453712	36034	38500	514	30	R.FFVNFVVGQDPGSDVAFHFNPR.F, R.FDGWDK.V,K.VVFNTLQGGK.W, K.VVVNGNPFYEYGHR.L,R.LQGGLTAR.R, K.SFAINFK.V,K.VGSSGDIALHINPR.M, K.KITHNPFGPGQFFDLSIR.C, K.ITHNPFGPGQFFDLSIR.C	10

Table 1. Proteins identified of HIM through SDS-PAGE (Continuation)

Line	Protein name	Access	Theoretical MM(Da)	Experimental MM(Da)	Score	% Coverage	Peptides	# Peptides
5	Ig alpha-2 heavy chain constant region.	gi|184761	37226	38500	587	20	R.DASGATFTWTPSSGK.S, R.GFSPKDVLVR.W,R.WLQGSQELPR.E, R.WLQGSQELPR.E, R.WLQGSQELPREK.Y, R.WLQGSQELPREK.Y,R.EKYLTWASR.Q, K.YLTWASR.Q, R.VAAEDWK.K, K.YLTWASRQEPSQGTTTFAVTSILR.V, R.QEPSQGTTTFAVTSILR.V	11
6	Ig kappa light chain VLJ region.	gi|21669449	28825	29000	65	18	R.TVAAPSVFIFPPSDEQLK.S, R.TVAAPSVFIFPPSDEQLK.S, K.DSTYSLSSTLTLSK.A, K.VYACQVTHQGLSSPVTK.S	4

Table 2. Proteins identified of HIM from 2-D PAGE, pH 3-10

Spot	Protein name	Access	Theoretical pI/MM(Da)	Experimental pI	Experimental MM(Da)	Score	% Coverage	Peptides	# Peptides
10	Bifunctional UDP-glucuronic acid decarboxylase [Escherichia coli]	gi|15802804	6.34/74859	6.40	74000	110	5	R.APLNWVLVNGETETGVTLHR.M, R.VLILGVNGFIGNHLTER.L	3
3	Chain A, Crystal structure of human calcineurin complexed with cyclosporin A and Human cyclophilin	gi|24987750	5.26/43392	5.30	40000	101	8	R.IITEGASILR.Q, R.YLFLGDYVDR.G	2
5	Chain A, Crystal structure of the 20s proteasome From Yeast At 2.4 Angstroms Resolution	gi|24987750	5.90/27440	5.90	28000	74	10	R.TIGMVVNGPIPDAR.N, R.MANLSQIYTQR.A	2
11	Ferric uptake regulator [Escherichia coli O157:H7 EDL933]	gi|15800386	5.68/17016	5.60	18000	93	16	R.VLNQFDDAGIVTR.H, K.VIEFSDDSIEAR.Q	2
6	Glutamine amidotransferase [Escherichia coli]	gi|290577	5.52/67065	5.52	68000	69	4	R.DVAEILLEGLR.R, R.FIFLEEGDIAEITRR.S	2

Table 2. Proteins identified of HIM from 2-D PAGE, pH 3-10 (Continuation)

Spot	Protein name	Access	Theoretical pI/MM(Da)	Experimental pI	Experimental MM(Da)	Score	% Coverage	Peptides	# Peptides
2	Hemoglobin subunit beta A	gi|122539	6.36/16012	6.35	16200	331	55	K.VHVDEVGGEALGR.L, R.FFESFGDLSTADAVMNNPK.V, K.VLDSFSDGMK.H, K.HLDDLK.G, K.LHVDPENFK.L, K.LLGNVLVVVLAR.N, K.EFTPVLQADFQK.V	7
4	Ig A-2 heavy chain	gi|184761	5.71/37226	5.60	37000	132	17	R.DASGATFTWTPSSGK.S, K.SAVQGPPER.D, R.WLQGSQELPR.E, K.YLTWASR.Q, R.QEPSQGTTTFAVTSILR.V	6
7	Ig kappa light chain VLJ region	gi|21669449	8.15/28885	8.11	29000	76	13	K.VDNALQSGNSQESVTEQDSK.D, K.VYACQVTHQGLSSPVTK.S	2
13	Polyubiquitin	gi|2627129	7.13/68448	7.10	68000	193	6	MQIFVK.T, R.TLSDYNIQK.E , K.ESTLHLVLR.L,K. TITIEVEPSDTIENVK.A	4
9	Scll+ suppressor protein [Saccharomyces cerevisiae]	gi|172546	6.7/30552	6.70	29000	74	9	R.TIGMVVNGPIPDAR.N, R.MANLSQIYTQR.A	2
1	Serine/threonine-protein phoshatase 2B catalytic subunit alpha isoform 1	gi|6715568	5.58/59346	5.48	59000	313	19	K.EVFDNDGKPR.V , R.LEESVALR.I, R.IITEGASILR.Q, K.LFEVGGSPANTR.Y, R.YLFLGDYVDR.G , K.TQEHFTHNTVR.G, K.TQEHFTHNTVR.G, R.AHEAQDAGYR.M, R.DAMPSDANLNSINK. AR.EESESVLTLK.G, K.YENNVMNIR.Q	11
8	Unnamed protein product [Homo sapiens]	gi|34039	5.04/44079	5.00	49000	269	20	R.ALEAANGELEVK.I,R. DYSHYYTTIQDLR.D, K.ILGATIENSR.I,R.IVLQIDNAR.L, K.FETEQALR.M ,R.MSVEADINGLR.R, R.VLDELTLAR.T, R.TDLEMQIEGLK.E	8

Table 2. Proteins identified of HIM from 2-D PAGE, pH 3-10 (Continuation)

Spot	Protein name	Access	Theoretical pI/MM(Da)	Experimental pI	Experimental MM(Da)	Score	% Coverage	Peptides	# Peptides
12	Unnamed protein product [Homo sapiens]	gi|189054178	7.62/66154	7.39	67000	341	12	R.SLVNLGGSK.S,R. SLVNLGGSKSISISVAR.G, K.SLNNQFASFIDK.V,R. FLEQQNQVLQTK.W, R.TNAENEFVTIK.K,R. SLDLDSIIAEVK.A, K.AQYEDIAQK.S,K.YEELQITAGR.H	9

Table 3. Proteins identified of HIM from 2-D PAGE, pH 4-7

Spot	Protein name	Access	Theoretical pI/ MM(Da)	Experimental pI/MM (Da)	Score	% Coverage	Peptides	# Peptides
1	Alpha-crystallin B chain	CRYAB_HUMAN	6.76/20146	6.80/20150	104	35	R.RPFFPFHSPSR.L, R.FSVNLDVK.H, R.QDEHGFISR.E, K.YRIPADVDPLTITSSLSSDGVLTVNGPR.K, R.IPADVDPLTITSSLSSDGVLTVNGPR.K, R.TIPITR.E, R.IPADVDPLTITSSLSSDGVLTVNGPR.K, R.IPADVDPLTITSSLSSDGVLTVNGPR.K	11
2	TIP41-like protein	TIPRL_HUMAN	5.6/31652	5.75/34000	470	43	K.LADELHMPSLPEMMFGDNVLR.I, R.IQHGSGFGIEFNATDALR.C, K.GTLLGESLK.L, R.VMPSSFFLLLR.F, R.IDGVLIR.M, R.LYHEADK.T, R.LYHEADK.T, R.LYHEADKTYMLR.E, K.TYMLR.E, K.ISSLMHVPPSLFTEPNEISQYLPIK.E, K.LIFPER.I,R.IDPNPADSQK.S, R.IDPNPADSQK.S	12
3	Ubiquitin	UBIQ_HUMAN	6.56/8560	6.55/10100	19	32	K.TITLEVEPSDTIENVK.A, K.EGIPPDQQR.L, K.EGIPPDQQR.L, K.EGIPPDQQR.L	4
4	UPF0539 protein C7orf59	CG059_HUMAN	6.07/10792	5.91/13400	97	26	R.GMNVPFK.R, R.GMNVPFKR.L, R.LSVVFGEHTLLVTVSGQR.V	4

Table 4. Proteins identified of HIM from shotgun analyses

Protein name	Access	Theoretical pI/ MM(Da)	Score	% Coverage	Peptides	# Peptides
Calcium-activated chloride channel regulator	CLCA1_HUMAN	5.97/100847	63	8	K.TKADYVRPK.L, K.NADVLVAESTPPGNDEPYTEQMGNCGEKGER.I, R.IHLTPDFIAGKK.L, R.AFVHEWAHLR.W, R.WGVFDEYNNDEKFYLSNGR.I	6
Galectin-4	LEG4_HUMAN	9.21/36032	136	5	R.FFVNFVVGQDPGSDVAFHFNPR.F, R.FDGWDKVVFNTLQGGK.W, K.VGSSGDIALHINPR.M, K.VGSSGDIALHINPR.M	5
IgA-1 chain C region	IGHA1_HUMAN	6.08/38486	148	23	R.DASGVTFTWTPSSGK.S, K.TFTCTAAYPESK.T, R.WLQGSQELPR.E, R.EKYLTWASR.Q, R.QEPSQGTTTFAVTSILR.V, K.KGDTFSCMVGHEALPLAFTQK.T	6
IgA-2 chain C region	IGHA2_HUMAN	5.71/37301	148	38	K.HYTNPSQDVTVPCPVPPPPPCCHPR.L, R.DASGATFTWTPSSGK.S, R.DLCGCYSVSSVLPGCAQPWNHGETFTCTAAHPELK.T, R.WLQGSQELPR.E, R.EKYLTWASR.Q, R.QEPSQGTTTFAVTSILR.V, K.KGDTFSCMVGHEALPLAFTQK.T	9
Ig kappa chain C region	IGKC_HUMAN	5.58/11773	229	53	K.VQWKVDNALQSGNSQESVTEQDSK.D, K.VDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSK.A, K.DSTYSLSSTLTLSK.A, K.VYACEVTHQGLSSPVTK.SK.HKVYACEVTHQGLSSPVTK.S, K.HKVYACEVTHQGLSSPVTK.S	8
Ig lambda chain C regions	LAC_HUMAN	6.92/11401	36	49	K.ADSSPVKAGVETTTPSK.Q, K.AGVETTTPSK.Q, K.QSNNKYAASSYLSLTPEQWK.S, K.YAASSYLSLTPEQWK.S, R.SYSCQVTHEGSTVEK.T	6
Muc-2	MUC2_HUMAN	5.49/552218	28	<1	D RPIYEEDLKK, K.TVHMMPMQVQVQVNR.Q, K.GVCVHGNAEYQPGSPVYSSK.C	4
Polymeric immunoglobulin receptor	PIGR_HUMAN	5.58/84429	95	15	R.GGCITLISSEGYVSSK.Y,R.TVTINCPFKTENAQK.R, K.NADLQVLKPEPELVYEDLRGSVTFHCALGPEVANVAK.F, K.RAPAFEGR.I, R.ILLNPQDKDGSFSVVITGLR.K, K.QGHFYGETAAVYVAVEERK. AK.QGHFYGETAAVYVAVEERK.A,	8

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 64-83
doi: 10.30744/brjac.2179-3425.AR-41-2020

BrJAC

ARTICLE

In Situ Proteomic Analysis of Glioblastoma Multiforme: A Translational Approach to Improve Prognostic/Diagnostic Routines

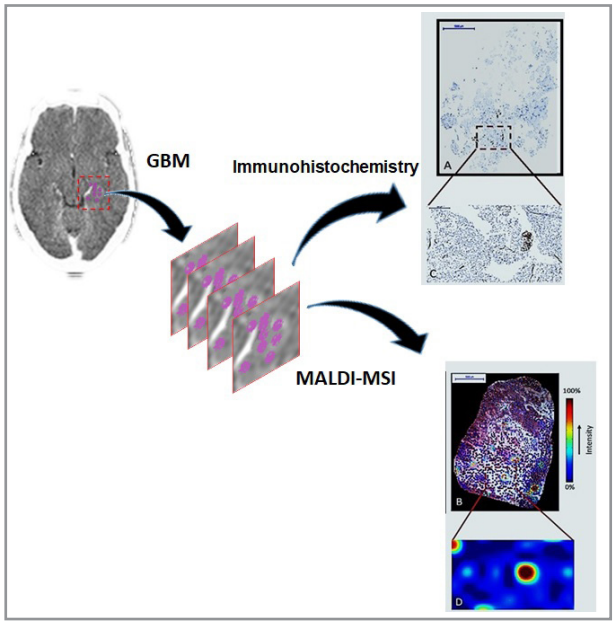
Anally Ribeiro da Silva Menegasso[1], Marcel Pratavieira[1], Lucilene Delazari dos Santos[2], Flávio de Oliveira Lima[2], Marcelo Padovani T. Moraes[3], Marco Antonio Zanini[3], Mario Sergio Palma[1]*

[1]Institute of Biosciences, Department of Basic and Applied Biology, Center of the Study of Social Insects, São Paulo State University (UNESP), Rio Claro, SP, Brazil

[2]Center for the Study of Venoms and Venomous Animals (CEVAP), Sao Paulo State University (UNESP), Botucatu, SP, Brazil

[3]Botucatu Medical School (FMB) – São Paulo State University (UNESP), Botucatu, SP, Brazil

**Correspondence: Prof. Dr. Mario Sergio Palma. CEIS-IBRC- UNESP, Av. 24A nº 1515, Bela Vista - 13506-900, Rio Claro, SP, Brazil*



In order to evaluate the use of matrix-assisted laser desorption ionization (MALDI) Mass Spectral Imaging (MSI) to Glioblastoma (GBM) studies, some sections of formalin fixed paraffin embedded samples of GBM tumors were submitted to classical immunoassays, to define the profile of distribution of some classical and well recognized molecular markers of GBM grade IV (Ki-67, S100, Glial GFAP, CD31 and CD34), while other sections of the same samples were submitted to in-tissue proteomic analysis by MALDI MSI, and both results compared to each other. The overlapping of the MALDI spectra obtained for the tryptic peptides with the immunohistochemical reactions of each marker protein were used to build a distribution map of the marker proteins all over the GBM tissue section. The results revealed a high correlation between both methods, indicating that MALDI MSI has enough sensitivity to be compared to the immunohistochemical methods, as well is sufficiently reliable to be used in biomarkers identification.

Keywords: MALDI-TOF MS, MALDI Imaging, Tumor biomarkers, Glioblastoma multiforme, In situ proteomics.

Cite: Menegasso, A. R. S.; Pratavieira, M.; dos Santos, L. D.; Lima, F. O.; Moraes, M. P. T.; Zanini, M. A.; Palma, M. S. In Situ Proteomic Analysis of Glioblastoma Multiforme: A Translational Approach to Improve Prognostic/Diagnostic Routines. *Braz. J. Anal. Chem.,* 2020, *7* (29), pp 64–83. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-41-2020

Submitted 3 August 2020, Resubmitted 14 October 2020, Accepted 15 October 2020, Available online 20 October 2020.

INTRODUCTION

GBM is a very common malignant primary brain tumor in adults, being one of the most lethal human cancers [1]. In US there are about 50,000 GBM patients and approximately 10,000 new cases are diagnosed annually [2–4]. In addition to this, GBM tumors are extremely resistant to radiation and chemotherapy [1,5]; even after aggressive surgical resection GBM patients survive only a few months [6]. GBMs are malignant gliomas arising from glial cells of the brain, and according to World Health Organization classification system, these types of gliomas are classified based on the resemblance of their histology and immunohistochemical profile to the presumed cells of origin [7]. The tumors are then graded according to the histological features associated with biological aggressiveness (i.e., mitotic figures, necrosis, vascular endothelial hyperplasia) [8]. According to this classification, grade I represents the least severe pathology, while grade IV corresponds to the most severe one. The diagnosis and grading of these gliomas take into account the understanding about the heterogenous histology and its relationship with a complex molecular scenario of these tumors [8].

Mutations suffered by normal cells are important for the classification of these tumors, and may be used to track the progress of the gliomagenesis [9]. Clinically, some patients may present a grade IV lesion which developed de novo from normal glial cells suffering mutations (primary GBM), while other patients may present GBM tumor which evolved from the progression of lower grade gliomas (secondary GBMs) [8]. Despite both GBM tumors are graded IV, they present distinct correlation with different molecular markers [10]. Secondary GBMs are rare, and accounts for less than 10% of GBMs cases; their unique immunohistochemical profile, and the clinical course of these tumors have clarified the understanding about GBM as a heterogeneous disease [10].

The cellular and histologic heterogeneity, as well the genetic and clinical differences amongst the GBM patients, and the difficulties to access fresh-frozen samples of tumor tissues, limited initially the contributions of proteomics to the study of GBM. This situation evolved due to the improvement of proteomic techniques, and availability of novel cellular and technical tools to overcome cell heterogeneity of these tumors [11]. Thus, as example, the novel techniques of quantitative proteomics permitted to investigate the serum proteomes of GBM patients, resulting in the identification of potential biomarkers to monitor patient responses to therapy [12]. Proteomics also contributed for understanding of gliomas pathobiology, as well identifying novel therapeutic targets for this disease; thus, as examples were reported the presence of ceruloplasmin, vitamin D binding protein, and serum amyloid P, which were not reported directly in the context of GBM formation, but presented a good correlation with tumor growth and metastasis formation [12].

The expression (or repression) of some important proteins are critical for understanding GBM tumorigenesis; the identification (or not) of these proteins have been used as criteria for diagnosis and prognosis of the disease [9]. The proteins reported as molecular markers of GBM tumorigenesis includes proteins such as epidermal growth factor receptor *(*EGFR*)*, platelet-derived growth factor receptor (PDGFR) [13], O-6-methylguanine-DNA-methyltransferase (MGMT), *p53*, Isocitrate dehydrogenase gene 1 and 2 (IDH1/2*)*, phosphatase and tensin homolog (PTEN), phosphoinositide 3-kinase (PI3K*)* [9] amongst many others. Since the GBMs are histologically heterogeneous, a series of different molecular makers are used to obtain the immunohistochemical profile of each tumor, for the purpose of diagnosis and prognosis. The most usual markers used are the proteins Ki-67 [14], S100, Glial fibrillary acidic protein – GFAP [15], CD31 and CD34 [16,17].

Mass spectrometry emerged as a very attractive method for the examination of proteins in different types of extracts and tissue sections. Currently, MALDI MSI associates the classical histology and in situ proteomics, permitting to map specific proteins into the organizational structure of a tissue section, at level of cutting edge for detection of proteins and peptides [18,19]. The development of MALDI MSI provided a powerful tool for proteomic research in pathology, enabling the identification and relative quantification of proteins directly in their histological framework. The matching of histomorphology with this level of proteomic analysis opens wide possibilities for a true revolution in this area of medical knowledge [20–22]. In MALDI MSI the proteins considered as molecular markers can be detected and identified without the requirement

of specific and highly selective antibodies (as required by the immunoassays). Because of this feature, the technique may be used both as an independent method for detecting and mapping the molecular maker in GBM tissue section, as well to validate the results of immunoassays. In order to evaluate the application of MALDI MSI to the study of GBM, some sections of formalin fixed paraffin embedded samples of GBM were submitted to classical immunohistochemistry, for defining the profile of molecular markers of GBM grade IV (Ki-67, S100, Glial GFAP, CD31 and CD34). Meanwhile, other sections of the same samples were submitted to in-tissue proteomic analysis and MALDI MSI, and both results compared to each other. The results revealed a great correlation between both methods, indicating that MALDI MSI has enough sensitivity to corroborate the immunehistochemical methods, as well is sufficiently reliable to be used in biomarkers identification.

MATERIAL AND METHODS

Histopathology methods

Processing the tissue samples

The tissues of human glioblastoma were obtained from the surgical pathology routine of the Department of Pathology of the Hospital of Clinics of the Faculty of Medicine of Botucatú (HCFMB), in conjunction with the Neurosurgery service of the same hospital. The project was approved by the Research Ethics Committee (CEP) of Botucatu Medical School UNESP (CEP 378/2012, protocol CEP 4305/2012). Frozen sections of non-formalin-fixed material were used for shotgun proteomic analysis, and histological sections of formalin fixed paraffin embedded GBM tissues were used for immunohistochemistry and MALDI MSI analysis. Histological sections with 3 µm thickness of each block were stained by the hematoxylin-eosin technique, and reviewed by a pathologist for the evaluation of histopathological findings.

Immunohistochemistry

The histological sections, with the respective positive and negative controls, were submitted to examination using an automated immunohistochemical system, with antigenic recovery mode. TPLink (Dako®). The incubation, development and counterstaining were performed with an AutoStainer Link48 (Dako®), using high sensitivity polymer and FLEX antibodies ready for use. The primary antibodies used were acquired from DAKO/AGILENT for using in Autostainer Link 48 system as follows: Glial Fibrillary Acidic Protein (GFAP) polyclonal rabbit anti-human), Ki-67 (clone MIB-1 mouse anti-human), CD31 (clone JC70A mouse anti-human), CD34 (clone QBEnd 10 mouse anti-human), and S100 (clone IR504 mouse anti-human). The slides were mounted on Entellan resin (Sigma).

Interpretation of the labeling reactions

The cases presenting any expression in the target were considered positive cells for the immunohistochemical markers GFAP, CD34, CD31, S100 and Ki-67. For interpretation of CD31 and CD34 any membrane or cytoplasmic staining was considered. For the interpretation of GFAP, any cytoplasmic and fibrillary immunostaining was considered positive, meanwhile for Ki-67 any nuclear labeling was considered. The interpretation for S100 protein, considered any nuclear or cytoplasmic labeling.

Mass spectrometric analysis and data treatment

GBM tissue section preparation for analysis

The fresh GBM tumors were frozen and maintained in the ultra-freezer at -80ºC The frozen tumors were sliced at 12 µm thick sections using a cryostat at -20 ºC (LEICA CM1850). Subsequent histological sections were analyzed immediately through histological staining technique by hematoxylin and eosin (H&E). The slices analyzed by MALDI-MSI were previously washed successively with 70% (v/v) ethanol, subsequently with 90% (v/v) ethanol during one minute in each solution, for the elimination of salts and lipids present in the samples. Subsequently, the histological sections were subjected to desiccation for 30 minutes under ambient temperature in a vacuum sealed desiccator containing silica gel as desiccant.

Some slices were submitted to staining with H&E for localization of GBM structures. For this technique, the slice preparations were submerged in 95% (v/v) ethanol during 20 minutes, washed under running water and then stained with hematoxylin for 20 seconds.

After this procedure, the slides were washed with distilled water. They were then stained with eosin for 20 seconds more and washed with distilled water. At the end of the process, the slides were rinsed with solutions of 95% (v/v) ethanol for removal of excess dye and in Xylol solutions. The digital images of the tissue sections were generated through a microscope Olympus BX51TF, connected to a camera Olympus U-LH100HG.

In the analyses of the tryptic digests, trypsin deposition was carried-out through a chemical printer CHIP-1000 (Shimadzu). The printer was programmed to generate a micro-arrangement of about 15,000 spots/tissue section analysis, where each point of application of the matrix (and trypsin solution) solution was spaced by 200 µm from each other (from center to center in lateral resolution), totaling 100 spots/ mm^2 of histological section. Initially in each spot was applied 20 nL of a solution containing 20 mg/mL of trypsin in 20 mM NH_4HCO_3 (pH 8.0) and incubated during 2 h, at 37 ºC. The section was then incubated for 90 min in a humid atmosphere at 37 ºC. Next, a solution containing 10 mg/mL of cinnamic acid (CHCA) in 50% (v/v) acetonitrile (containing 1% (v/v) trifluoroacetic acid (TFA)) was used as a matrix for analysis of the tryptic digests, being applied 500 pL (5 drops of 100 pL) of matrix/point solution. After the application of the matrix, the histological sections were submitted to the process of desiccation during 15 minutes, under room temperature. Planar coordinates (x and y) created by the ChIP-1000 were saved and exported to the mass spectrometer.

Proteome approach

LC-MS and LCMS^n analysis

One slice of GBM tumor (12 µm thick sections) was homogenized in (~5 µg tissue) in 300 µL of 50 mM ammonium bicarbonate, pH 7.9, and centrifuged at 10.000 x g during 10 min at 4 ºC. The supernatant was dried using a speed-vac system (Edwards Co,), and maintained at -80 ºC until be analyzed. The samples were solubilized in 50 mM ammonium bicarbonate, pH 7.9, containing 7.5 M urea and incubated for 60 min at 37 ºC to denature the proteins, which were then reduced with 10 mM DTT at 37 ºC for 60 min. After this treatment, the proteins were alkylated with 40 mM iodoacetamide at 25 ºC for 60 min in the dark. The samples were diluted two-fold with 100 mM ammonium bicarbonate, pH 7.8, and 1 M calcium chloride was added to the samples to a final concentration of 1 mM. Non-autolytic trypsin (Promega) was added to the denatured protein solution (1:50 trypsin : protein, w/w) and incubated for 18 h at 37 ºC. The samples were frozen in liquid nitrogen to interrupt the enzymatic digestion. The digested samples were desalted using an SPE C18 column (Discovery DSC-18, SUPELCO, Bellefonte, PA, USA) conditioned with MeOH, rinsed with 1 mL 0.1% TFA and washed with 4 mL of 0.1% (v/v) TFA / 5% (v/v) ACN. Peptides were eluted from the SPE column with 1 mL of 0.1% TFA/80% ACN and concentrated to dryness using a Speed-Vac. The digested samples were stored at -80 ºC until needed for analysis; the tryptic peptides were solubilized in 50% ACN and subjected to LC-MS and MS^n analysis. The samples were solubilized in 100 µL of 50% (v/v) ACN and fractionated in an LC-MS system using an X-Bridge BEH 130 C-18 column (100 mm × 2.1 mm; 3.5 µm) (Waters, Massachusetts, USA) at a flow rate of 200 µL/min. Elution was performed under gradient conditions from 5 to 95% (v/v) ACN (containing 0.1% (v/v) TFA) between 0 and 95 min at 30 °C. The eluent was monitored at 215 nm with a UV-DAD detector, mod. SPD-M10A (SHIMADZU, Kyoto, Japan) coupled to an IT-TOF/MS and MS^n mass spectrometer system equipped with an electrospray ionization source (Shimadzu, Kyoto, Japan). Spectra were acquired in positive mode, with activation of data dependent acquisition (DDA), which permits an automatic switching from MS to MS2 mode. The electrospray voltage was set to 4.5 kV, the CDL temperature was set to 200 °C, the block heater temperature was adjusted to 200 °C, the nebulizer gas (N_2) flow was 1.5 L/min, the trap cooling gas (Ar) flow was 95 mL/min, the ion trap pressure was 1.7×10^{-2} Pa, the TOF region pressure was 1.5×10^{-4} Pa, and the ion accumulation time was 50 ms. The top five ions from each MS spectrum were selected as precursors (Top N) for fragmentation

in MS^2, as typically used in DDA experiments. The collision energy was set at 35% both for MS^2 and MS^3, and the collision gas set to 20%. Auto-tuning was performed in the presence of Na-TFA solution (10 mM NaOH + 0.1% (v/v) at pH 3.5). The mass spectral resolution was approximately 10,000 FWHM, and error was approximately 3.08 ppm.

Mass spectrometry conditions for MALDI spectral acquisition

The tissue section preparation was dried under vacuum during 10 min and submitted to the acquisition of MALDI spectra in the positive mode, using a MALDI-TOF-TOF instrument mod. AXIMA Performance (Shimadzu Corp., Kyoto, Japan), equipped with a laser SmartBeam system, with the reflectron device activated. The instrument was controlled using the Launchpad v2.8 software (Shimadzu). The setting conditions were: CDL temperature adjusted to 200 ºC, block heater temperature at 200 ºC, TOF region pressure 1.5×10^{-4} Pa, ion accumulation time 50 ms; helium was used as collision gas. It was applied an accelerating voltage of 20 kV and delayed extraction, peak density of maximum 50 peaks per 200 Da, minimal S/N ratio of 10 and maximum peak at 60. The instrument was calibrated using a standard calibration mixture of ACTH (fragment 18-39), Angiotensin II and P14R. The MS spectra were acquired in the *m/z* range 700 - 3600, with the laser power set to 70%, adjusted to perform delayed extraction; the density of peaks was set at 50 for each 200 peaks presenting S/N ratio ≥ 10. The spectra were acquired with 50 shots per movement from the center of each spot, up to a distance of 50 µm straight, performing a total of 250 laser shots per spot. Under the conditions reported above the error obtained was 3.08 ppm and resolution 9,000 FWHM.

Proteins identification

The proteins identification was performed with the algorithm MASCOT (v 2.3) to analyze the MS^2 spectra, using the protein databank restricted to the genus *Homo sapiens* (NCBI). The search parameters were set as follows: trypsin was selected as enzyme, one maximum missing cleavage sites selected; peptide mass tolerance was adjusted to 0.3 Da for MS and 0.2 Da for MS^2 spectra; carbamidomethyl (C) specified as a fixed modification, and methionine oxidation was selected as a variable modification. After protein identification, an error-tolerant search was performed to detect nonspecific cleavage. The proteins identified after the database search were subjected to additional filtering using Scaffold 4.3.2 (Proteome Software Inc., Portland, OR) to validate the peptide identification and to obtain a false discovery rate (FDR) of less than 1%; FDR was calculated from forward and decoy matches by requiring significant matches for at least two distinct peptide sequences. According to a Local FDR algorithm implemented in Scaffold, the peptide probability was set to a minimum of 90%, whereas the protein probability was set at 95%. The databanks mentioned above were screened for common external contaminants (laboratory proteins and standards proteins provided through the Global Proteome Machine Organization (http://www.thegpm.org/crap/index.html). Functional and Gene Ontology (GO) annotations were performed using the algorithm developed by the GO Consortium (http://www.geneontology.org/).

Spectral imaging data processing

Using of the Launchpad algorithm v 2.8 (Shimadzu) the raw data (***.raw***) were converted to the format ***.mzXML***. The images were rebuilt in MSiReader v 0.05 algorithm from the data in the format ***.mzXML***, using a mass tolerance of ± 0.3 Da and spacing of 250 µm. In this algorithm, the following parameters were configured: the method used to calculate the intensity displayed in molecular images was the ***Max of Window*** (**Max**); subtraction of noise and *background*; baseline correction and cubic interpolation of order 5 of the images' pixels.

Molecular images of GBM sections were constructed using the corresponding *m/z* values of the molecular ions from the tryptic fragment peptides of the molecular markers selected, in their monoprotonated form $[M+H]^+$ for each specific molecular marker as described elsewhere [23]. The scale of color used in the border of these images correspond to a semi-quantitative method of representation of molecules distribution in a

snap-shot of sample collection. The images used to represent the distribution of each molecular marker were built based on the overlapping of the Extract of Individual Ion Chromatogram (XIC) of all the tryptic fragments identified for the marker.

RESULTS AND DISCUSSION

The GBM section preparation using a chemical printer applied reagents in 25 spots/mm^2 of tissue section; considering that the total surface of each section presented 300 mm^2, the whole surface of each GBM tissue section was covered by an array of 7,500 spots. Thus, if each spot was shot by 250 laser shots, this means that the complete analysis of each section required 1.875 x 10^6 laser shots, and consequently the same number of spectra were acquired for the analysis of each GBM tissue section. The tryptic digestion of GBM tissue section, followed by the MS and MS/MS analysis resulted in the detection and reliable sequencing of 166 peptides, which in turn permitted the identification of 25 proteins (shown in Table S.I in Supplementary Material). The identifications were performed using a minimum of two tryptic peptides per protein; the protein scores changed from 33 to 83, with sequence coverage from 3% to 58 % (Table S.I in Supplementary Material). These proteins probably are the most abundant ones in the tissue slices of GBM tumor cells and/or correspond to the proteins most ionizable from the samples, under the experimental conditions reported in material and methods. The proteins with functions identified by GO analysis were classified into different groups; those apparently related to processes involved with the development of cancer tumors are described as follows:

i) **Organization/regulation of the cytoskeleton:** Thymosin beta-10, Calponin-2, and Microcephalin, and F-actine capping protein subunit beta;
ii) **Adhesion cell-matrix / cellular differentiation:** Collagen alpha-1 chain;
iii) **Signal transduction:** Guanine nucleotide-binding protein G, Cocaine/ amphetamine-regulated transcript protein, RAB3A interacting protein, and Raftlin;
iv) **Regulation of T-cell activation:** Dual specificity tyrosine-phosphorylation-regulated kinase 1B;
v) **Regulation of Immune Response:** Tumor necrosis factor receptor superfamily member 17, Ubiquitin-associated and SH3 domain-containing protein A, Kelch-like protein 6, and serpin-like protein HMSD;
vi) **Energetic metabolism:** alpha amylase 2B, ATP-synthase subunit-S like protein;
vii) **Protein degradation:** LON-peptidase N-terminal domain and ring finger protein;
viii) **Cel-to-cell communication:** cysteine rich PD2-binding protein;
ix) **Transcriptional process:** host cell factor 2.

Table S.I in Supplementary Material also shows a small group of proteins (Ki-67, glial fibrillary acidic protein - GFAP, protein S100-A, hematopoietic progenitor cell antigen CD34, and platelet endothelial cell adhesion molecule - CD31) that have been frequently used as markers of proliferation of GBM tumor cells by the pathologists [14-16]. Thus, considering the use of these five proteins in the routine diagnostic of GBM in humans, the images generated for mapping the distribution of these markers all over the GBM section using immunohistochemistry were compared to the images resulting from the application of MALDI MSI protocols to the equivalent tissue sections, as shown in Figures 1 to 5. The images of these proteins, mapping their pattern of distribution in the GBM tumor section, were generated by overlapping the *m/z* values of all tryptic peptides detected for each marker; the individual spatial distribution was made in a relative scale, using the peak of highest intensity amongst those corresponding to the tryptic peptides of each marker protein.

The Figures 1A and 1B show the pattern of labelling obtained by immunohistochemistry and the general profile of MALDI MSI, respectively, for the GBM marker CD31 over the GBM tissue section. The Figure 1C is showing the magnified image of the area assigned by the dashed lines in the Figure 1A; meanwhile, the figure 1D is showing the magnified image corresponding to the MALDI MSI of the region assigned by a dashed rectangle in Figure 1B.

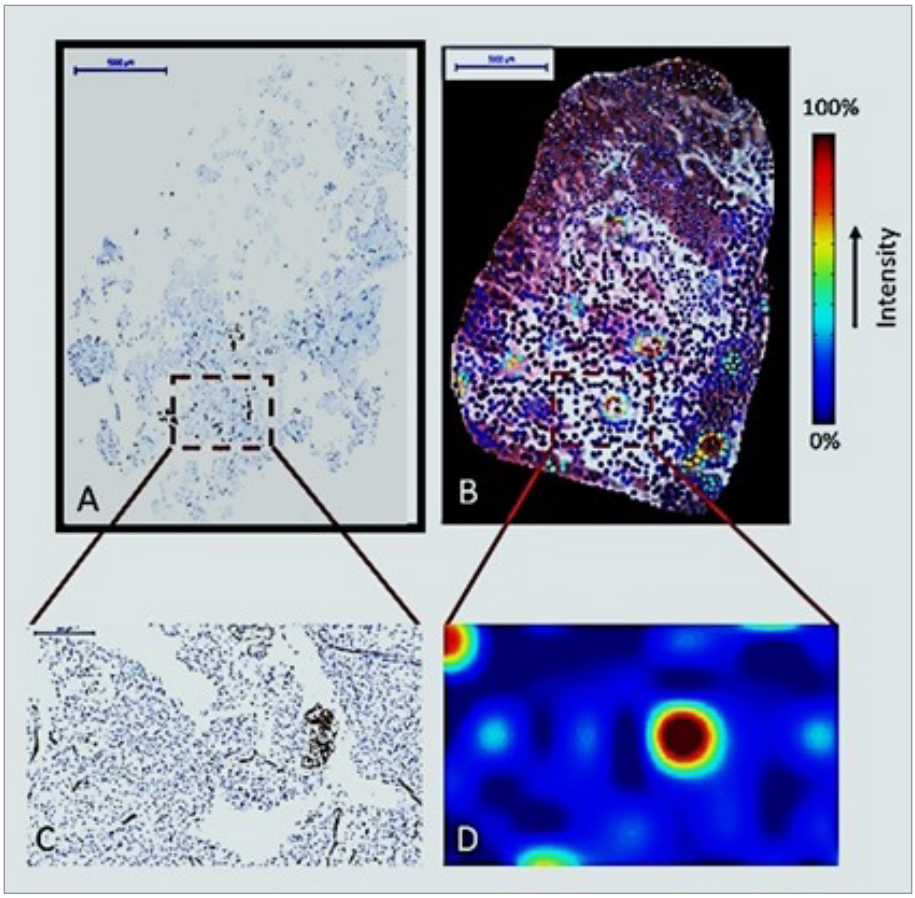


Figure 1. Immunohistochemistry and MALDI-MSI for the molecular marker CD31. Comparative overlapping of images between the immunohistochemistry labeling for CD31 as molecular marker (A), and the molecular mapping of the same biomarker, generated by MALDI Spectral Imaging (B). The selected region in (A) was magnified (C) to show the evident glomeruloid vascular formation, with the reactional endothelial proliferation associated to capillary vessels, characterized by the positive immunohistochemical reaction (brown goldish labeling). In the corresponding figure generated by MALDI Spectral Image (D), the glomeruloid structures are identified by their "hot colors", in the same spatial localization of the positive foci of the immunohistochemical reactions, spaced by negative or less positive areas. The images of CD31 distribution were produced by overlapping of the *m/z* values of the tryptic peptides detected for this marker: *m/z* 1244.62 (SLPDWTVQNGK), *m/z* 1744.81 (PELESSFTHLDQGER), *m/z* 1609.65 (EQEGEYYCTAFNR), and *m/z* 1198.55 (DTETVYSEVR). The spatial distribution of CD31 was represented quantitatively in a relative scale, shown in the right side of the figure.

The Figures 2A and 2B show the pattern of labelling obtained by immunohistochemistry and the general profile of MALDI MSI, respectively, for the GBM marker CD34 over the GBM tissue section. The Figure 2C is showing the magnified image of the area assigned by the dashed lines in the Figure 2A; meanwhile, the figure 2D is showing the magnified image corresponding to the MALDI MSI of dashed rectangle in Figure 2B.

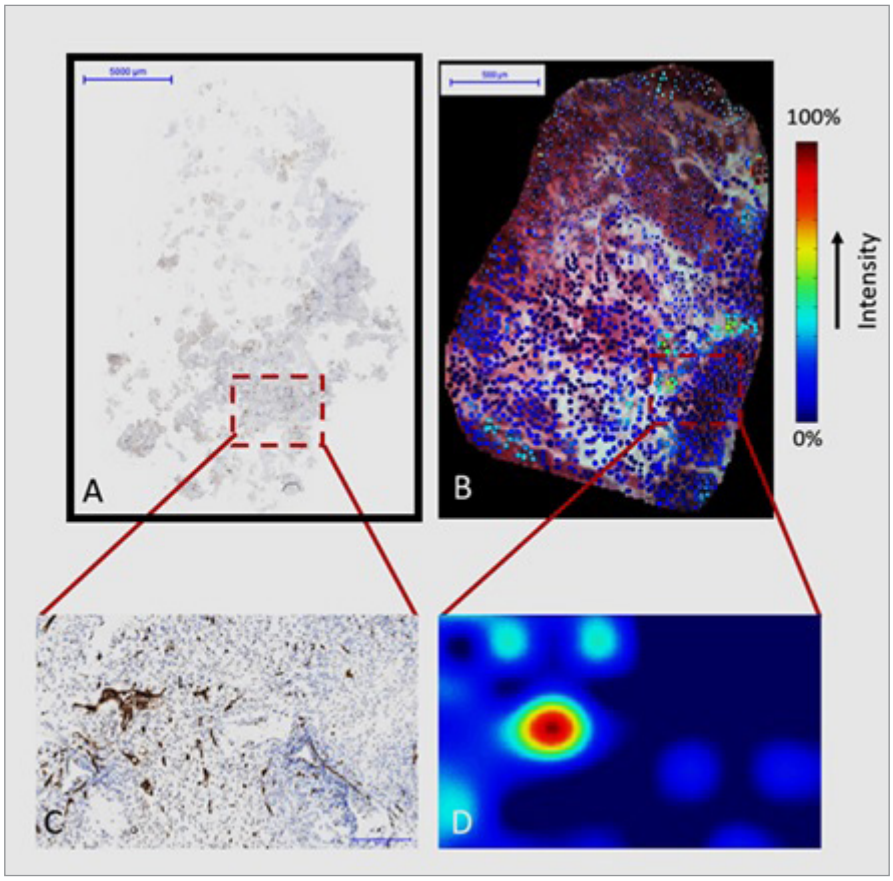


Figure 2. Immunohistochemistry and MALDI-MSI for the molecular marker CD34. Comparative overlapping of images between the immunohistochemistry labeling for CD34 (A), and the molecular mapping of the same biomarker, generated by MALDI Spectral Imaging (B). The selected region in (A) was magnified (C) to show the evident glomeruloid vascular formation, where the reactional endothelial proliferation is associated to capillary vessels, characterized by the positive immunohistochemical reaction (brown goldish labeling). In the corresponding figure generated by MALDI Spectral Image (D), the glomeruloid structures are identified by their "hot colors", in the same spatial localization of the positive foci of the immunohistochemical reactions, spaced by negative or less positive areas. The images of CD34 distribution were produced by overlapping of the *m/z* values of the tryptic peptides detected for this marker: *m/z* 2348.18 (PSLSPGNVSDLSTTSTSLATSPTK), *m/z* 1407.73 (PYTSSSPILSDIK), *m/z* 664.34 (TEISSK), *m/z* 2266.09 (LGILDFTEQDVASHQSYSQK), and *m/z* 919.42 (SWSPTGER). The spatial distribution of CD34 was represented quantitatively in a relative scale, shown in the right side of the figure.

The Figures 3A and 3B show the pattern of labelling obtained by immunohistochemistry and the general profile of MALDI MSI, respectively, for the GBM marker GFAP over the GBM tissue section. The Figure 3C is showing the magnified image of the area assigned by the dashed lines in the Figure 3A; meanwhile, the figure 3D is showing the magnified image corresponding to the MALDI MSI of dashed rectangle assigned in Figure 3B.

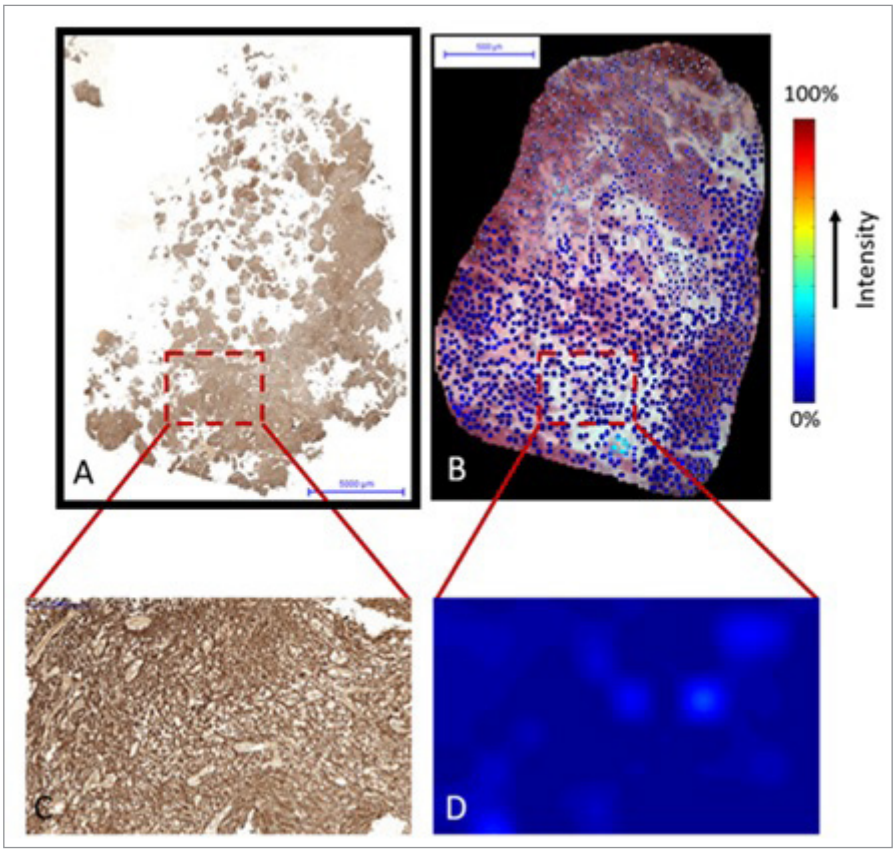


Figure 3. Immunohistochemistry and MALDI-MSI for the molecular marker GFAP. Comparative overlapping of images between the immunohistochemistry labeling for GFAP (A), and the molecular mapping of the same biomarker, generated by MSI (B). The pattern of reaction throughout the tissue is uniform and homogeneous, as characterized by the distribution of GFAP all over the nervous neoplastic tissue section. The selected region in (A) was magnified (C) to show the diffuse distribution pattern all over the tumor tissue in GBM, characterized by a brown labeling. The magnified image generated by MSI (Figure 3D), was characterized by low intensity blue spots, corresponding to the overall distribution of GFAP. The images GFAP distribution were produced by overlapping of the *m/z* values of the tryptic peptides detected for this marker: *m/z* 1697.79 (SYVSSGEMMVGGLAPGR), *m/z* 1108.47 (AEMMELNDR), *m/z* 986.52(ELQEQLAR), and *m/z* 1208.58 (EAASYQEALAR). The spatial distribution of GFAP was represented quantitatively in a relative scale, shown in the right side of the figure.

The Figures 4A and 4B show the pattern of labelling obtained by immunohistochemistry and the general profile of MALDI MSI, respectively, for the GBM marker S-100 over the GBM tissue section. The Figure 4C is showing the magnified image of the area assigned by the dashed lines in the Figure 4A; meanwhile, the figure 4D is showing the magnified image corresponding to the MALDI MSI of dashed rectangle assigned in Figure 4B.

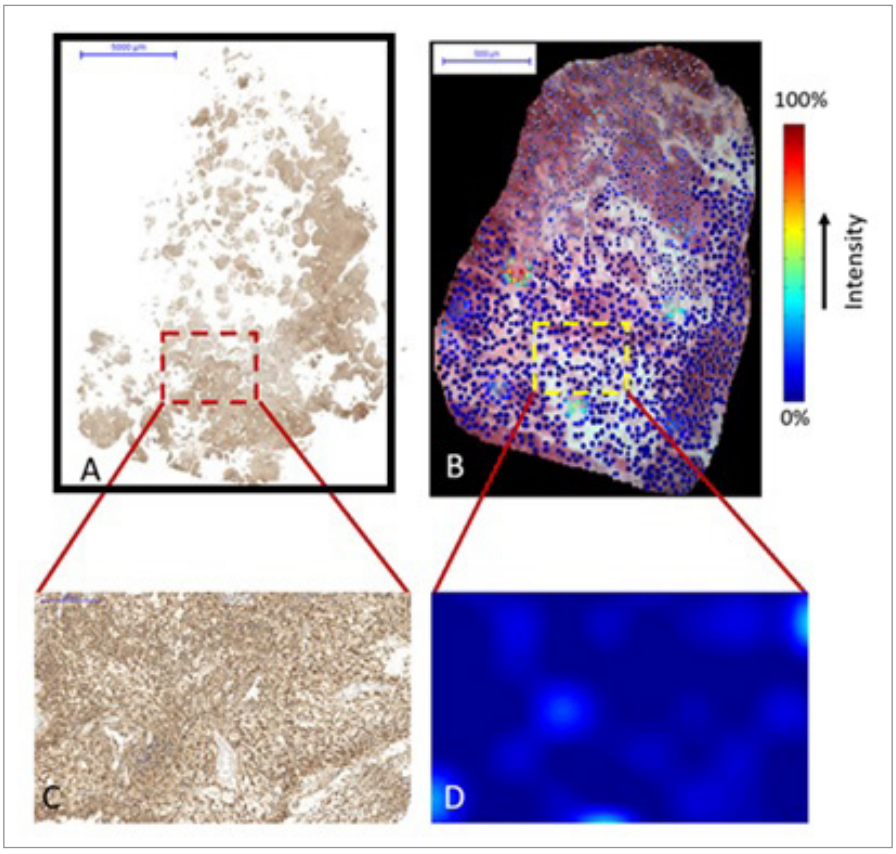


Figure 4. Immunohistochemistry and MALDI-MSI for the molecular marker S100. Comparative overlapping of images between the immunohistochemistry labeling for the S100 (A), and the molecular mapping of the same biomarker, generated by MSI (B). A pattern of uniform homogeneous reaction characterized the distribution of S100 protein all over the tissue section. The selected region in (A) was magnified (C) to show the diffuse distribution pattern all over the tumor tissue in GBM, characterized by a brown labeling. The magnified image generated by MSI (Figure 4D), was characterized by low intensity blue spots, corresponding to the distribution of S100 protein. The images of S100 distribution were produced by overlapping of the *m/z* values of the tryptic peptides detected for this marker: *m/z* 1691.98 (ELLQTELSGFLDAQK), *m/z* 761.36 (DVDAVDK), *and m/z* 2531.18 (MGSELETAMETLINVFHAHSGK). The spatial distribution of S100 was represented quantitatively in a relative scale, shown in the right side of the figure.

The Figures 5A and 5B show the pattern of labelling obtained by immunohistochemistry and the general profile of MALDI MSI, respectively, for the GBM marker Ki-67 over the GBM tissue section. The Figure 5C is showing the magnified image of the area assigned by the dashed lines in the Figure 5A; meanwhile, the figure 5D is showing the magnified image corresponding to the MALDI MSI assigned by a dashed rectangle in Figure 5B.

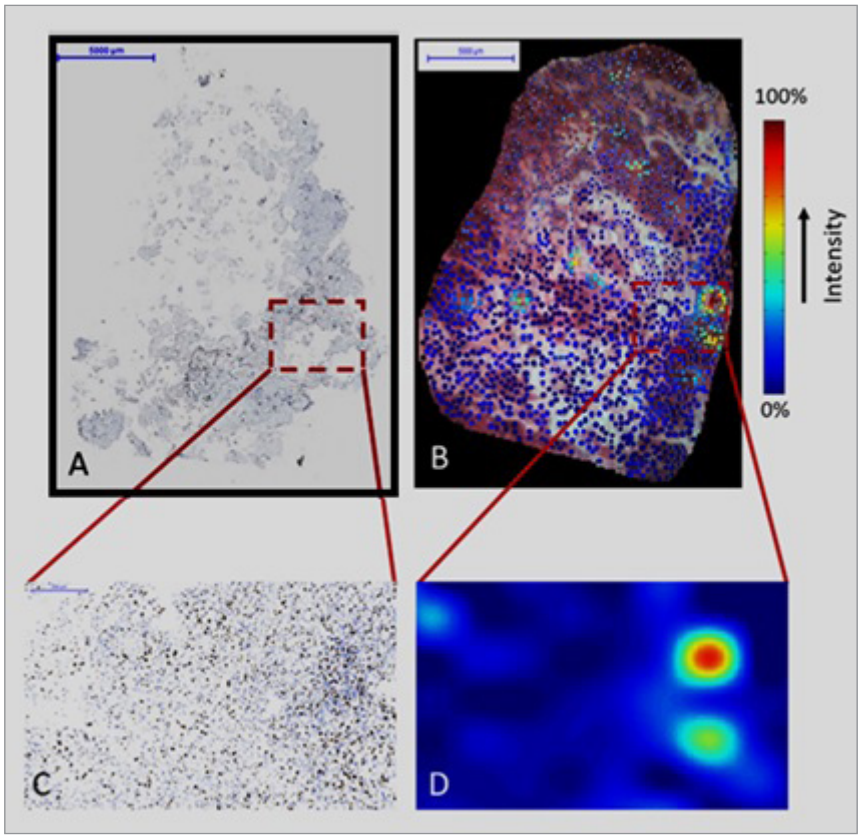


Figure 5. Immunohistochemistry and MALDI-MSI for the molecular marker Ki-67. Comparative overlapping of images between the immunohistochemistry labeling for the Ki-67 protein as molecular marker of cellular proliferation (A), and the molecular mapping of the same biomarker, generated by MSI (B). A pattern of reaction characterized the distribution of Ki67 protein all over the tissue section, with areas of relatively low reaction pervaded with areas of high Ki67 reactivity (the "hot spots"). The selected region in (A) was magnified (C) to show the irregular cluster distribution pattern in the tumor tissue in GBM reflecting different cellular proliferation rates among different spots (brown/grey nuclear labeling). The image generated by MSI (Figure 5D) reveals areas of higher cellular proliferation ("hot spots") identified by their "hot colors", in the same pattern of spatial localization of the positive foci of the immunohistochemical reactions, spaced by negative or less positive areas, corresponding to the irregular distribution of Ki67 protein. The images of Ki67 distribution were produced by overlapping of the m/z values of the tryptic peptides detected for this marker: *m/z* 858.49 (EELLAVGK), *m/z* 1830.95 (ESADGLQGETQLLVSRK), *m/z* 673.78 (LTPSAGK), *m/z* 2023.32 (MPCQSLQPEPINTPTHTK), and *m/z* 1862.94 (TEAEQQITEVFVLAER). The spatial distribution of Ki-67 was represented quantitatively in a relative scale, shown in the right side of the figure.

Discussion

The outstanding achievements in current oncology includes the use of state-of-art neuroimaging technology, associated to aggressive surgical resections, new strategies in radiotherapy and chemotherapy. Despite to this, the malignant gliomas such as GBM continues challenging the advances in the modern medicine; the current prognosis for GBM patients is not good. The future of GBM therapy seems to depend on the studies at molecular and cellular levels, to improve our understanding about the genetics, biochemistry and metabolism of this disease. The identification of the genetic events and the regulatory pathways involved in the immunological responses of the tumors, as well as the understanding of the

role of stem cells in tumor aggressiveness/resistance may enlighten the development newer strategies of therapy of this fatal disease. In this sense, the validation of MALDI Imaging technology to identify novel molecular markers, not depending of the development of highly selective antibodies, may contribute for the application of novel imaging strategies, improving the current knowledge in imaging diagnostics of GBM.

In the present study were used two complimentary strategies, i.e., immunohistochemistry and MALDI MSI, both performed in histological sections of GBM tumors from the same patient. Despite the samples came from the same tumor, the treatment of GBM sections required for each approach is different. For MALDI-Imaging assays we obtained cuts of frozen fresh material with a mean thickness of 12 µm, while the tissue sections used in immunohistochemical reactions were fixed in formaldehyde, paraffin-included and cut into histological sections with 3 µm thickness.

The results of in situ proteomic analysis in the GBM tumor sections performed in the present study identified proteins from different functional categories such as: organization/regulation of the cytoskeleton, adhesion cell-matrix / cellular differentiation, signal transduction, regulation of t-cell activation, regulation of immune response, energetic metabolism, protein degradation, cel-to-cell communication, transcriptional process.

It is important to emphasize that the use of immunohistochemistry (for the purpose of mapping proteins in tissue sections), requires the use of previously developed and produced primary antibodies, specific for each protein under investigation [24]. This fact limits severely the study to a few known proteins, which are markers for confirmation of histogenetic diagnosis or accessing the proliferation rate of the tumors [18]. Meanwhile, MALDI MSI is a technique that may be potentially applied to this situation, hypothetically without the use any standard protein, and does not requiring the use of any type of antibody, permitting more freedom of choice and a better chance to identify candidate proteins to become molecular prognostic/diagnostic markers, or therapeutic targets [25]. The use of MALDI MSI generated from the overlapping of the spectra of all tryptic peptides detected and sequenced, resulted in a reliable mapping of the distribution of each molecular maker all over the tissue section.

However, for this purpose it is necessary to validate the use of MALDI MSI to the study of proteins related to GBM tumors; thus, a comparative study between both techniques was performed. Amongst the 25 proteins identified by in situ proteomic analysis, were identified five proteins which are well accepted standardized markers, with commercially available antibodies routinely used in surgical pathology to assist the diagnosis of GBM tumors by immunohistochemistry [26]: protein Ki-67, glial fibrillary acidic protein - GFAP, protein S100-A, hematopoietic progenitor cell antigen CD34, and platelet endothelial cell adhesion molecule - CD31. Their detection and the exact location in GBM tissue sections were performed by both immunohistochemistry and MALDI MSI, and the comparative results are shown in Figures 1 to 5.

The markers CD31 and CD34 are characteristic of the vascular endothelial cells, localized in the inner lining of the blood vessels [27]. Blood vessels in glioblastomas form complex vascular glomeruloid structures that are part of histopathological criteria of this neoplasm. A careful observation of Figures 1 and 2 reveals the similar vascular pattern obtained both for CD31 and CD34, highlighting the classical glomeruloid formations of Glioblastomas (Figures 1A and 1C for CD31, and 2A and 2C for CD34) labeling both markers, as expected for proteins associated to endothelial tissues. The images of molecular mapping of these markers obtained by MALDI MSI (Figures 1B and 1D for CD31, and Figures 2B and 2D for CD 34) match well the immunohistochemical glomeruloid pattern. The results of figures 1C and 1D, as well 2C and 2D, suggests that the markers CD31 and CD34 seem to be distributed all over the tumor section, and form spots of high concentrations of these markers, matching the immunohistochemical pattern of the glomeruloid formations.

The neural markers GFAP and S-100 tend to have distribution throughout the brain and also the neoplastic tissue, with less heterogeneity in their location in these target tissues. Both proteins are markers of histogenesis of this neoplasm in cases of uncharacteristic morphology [7,8]. These proteins have a non-homogeneous tissue distribution, with areas of more and of less intense labeling, reflecting little differences of distribution [28]. The immunohistochemical patterns of detection of GFAP and S-100 (Figures 3A and 3C, and 4A and 4C, respectively), are expected for constitutive proteins in the nervous tissues and their tumors.

The images of molecular mapping of these markers obtained by MALDI MSI (Figures 3B and 3D, and 4B and 4D, respectively) match well the patterns obtained by immunohistochemistry, and even enhancing the little differences of concentration that are quite not detectable in the immunoassays. The results reported above are suggesting that GFAP and S-100 seems to be homogeneously distributed all over the tumor section, not forming hot spots due to the accumulation of high concentrations of the markers.

The immunohistochemical detection of Ki-67 (a cell proliferation marker) has a lower presence in tumors of low proliferation, and a higher presence in aggressive neoplasias such as GBM. Ki-67 is a nuclear antigen expressed in the G1, S, G2, and M phases of the cell cycle [29], thus labeling proliferating cells. A careful comparison of Figures 5A and 5B reveals a very similar pattern of distribution of the protein Ki-67 over the GBM section by both techniques, showing formation of clusters of high cell proliferation at different areas of the GBM section, also called “hot spots”. The results shown in Figure 5A are corroborated by the results obtained by MALDI Imaging, shown in of Figure 5B; the clusters of immunohistochemistry labeling shown in Figure 5C are easily confirmed by the observation of similar hot spots in the image of the corresponding region of the tissue section, produced MALDI MSI (Figure 5D).

CONCLUSIONS

The results above are clearly indicating that the pattern of distribution of classical immunohistochemical molecular markers (protein Ki-67, GFAP, protein S100-A, CD34, and CD31) through the sections of GBM tumors are very similar to the pattern obtained by MALDI Imaging. This observation may validate the use of MALDI Imaging as an experimental strategy for mapping molecular markers of GBM tumors. The results above are indicating that the use of MALDI Imaging in medicine may provide access to an innovative and revolutionary technology that will allow studies in Proteomic Pathology. This will permit to design future projects of comparative studies of the constitutional proteins of normal and neoplastic cells, as well to identify and quantify cell-signaling proteins, with enormous potential of application in diagnosis and prognosis of cancer tumors.

Acknowledgements

This work was supported by a grant from the São Paulo State Research Foundation (FAPESP, BIOprospecTA program 2016/16212-5 and INCT Program/iii-CNPq/MCT). M.S.P. is researching for CNPq Proc. 305566/2017-2.

Conflict of Interest

The authors declare that they have no conflict of interest.

Authorship agreement

All authors have read the journal’s authorship agreement and the manuscript has been reviewed by and approved by all named authors.

Ethics approval and consent to participate

Ethics approval was obtained from the local ethics committee. Written informed consent was obtained either from the patient or their appointed legal guardian.

REFERENCES

1. Dunn, G. P; Rinne, M. L.; Wykosky, J.; Genovese, G.; Quayle, S. N.; Dunn, I. F.; Pankaj K.; Agarwalla, P. J.; Chheda, M. G.; Campos, B. *Genes Dev*., **2012**, *26*, pp 756 – 784 (http://www.genesdev.org/cgi/doi/10.1101/gad.187922.112).
2. Davis, F. G.; Kupelian, V.; Freels, S.; McCarthy, B.; Surawicz, T. *Neuro-Oncology*, **2001**, *13*, pp 152 – 168 (https://doi.org/10.1093/neuonc/3.3.152).
3. Dolecek, T. A.; Propp, J. M.; Stroup, N. E.; Kruchko, C. *Neuro-Oncology*, **2012**, *4*, pp 1 – 49 (https://doi.org/10.1093/neuonc/nos218).

4. Porter, K. R.; McCarthy, B. J.; Freels, S.; Kim, Y.; Davis, F. G. *Neuro-Oncology*, **2010,** *12*, pp 520 – 527 (https://doi.org/10.1093/neuonc/nop066).
5. Masui, K.; Cloughesy, T. F.; Mischel, P. S. Neuropathol. *Appl. Neurobiol*., **2012**, *38*, pp 271 – 291(https://dx.doi.org/10.1111/j.1365-2990.2011.01238.x).
6. Stupp, R.; Mason, W. P.; van den Bent, M. J.; Weller, M.; Fisher, B.; Taphoorn, M. J. B.; Belanger, K.; Brandes, A. A.; Marosi, C.; Bogdahn, U.; et al. *N. Engl. J. Med*., **2005**, *352*, pp 987– 996 (https://www.nejm.org/doi/full/10.1056/NEJMoa043330).
7. Louis, D. L.; Ohgaki, H.; Wiestler, O. D.; Cavenee, W. K.; Burger, P. C.; Bernd, W.; Scheithauer, B. W.; Kleihues, P. *Acta Neuropathol*., **2007,** *114*, pp 97 – 109 (https://dx.doi.org/10.1007/s00401-007-0243-4).
8. Cloughesy, T. F.; Cavenee, W. K.; Mischel, P. S. *Annu. Rev. Pathol. Mech. Dis*., **2014**, *9*, pp 1-25 (https://dx.doi.org/10.1146/annurev-pathol-011110-130324).
9. Karsy, M.; Huang, T.; Kleinman, G.; Karpel-Massler, G*. Front. Biosci*., **2014**,*19*, pp 1065 – 1087.
10. Ohgaki, H.; Kleihues, P. *Clin. Cancer Res*., **2013**, *19*, pp 764 – 772 (https://doi.org/10.1158/1078-0432.CCR-12-3002).
11. Thirant, C.; Gavard, J.; Junier, M. P.; Chneiweiss, H*. Proteomics. Clin. Appl*., **2013**, *7*, pp 79 – 90 (https://doi.org/10.1002/prca.201200102).
12. Gollapalli, K.; Ray, S.; Srivastava, R.; Renu, D.; Singh, P.; Dhali, S.; Dikshit, J. B.; Srikant, R.; Moivadi, A.; Srivastaya, S. *Proteomics*, **2012**, *12*, pp 2378 – 2390 (https://doi.org/10.1002/pmic.201200002).
13. Huse, J. T.; Holland, E. C. *Nat. Rev. Cancer*, **2010**, *10*, pp 319 – 331 (https://doi.org/10.1038/nrc2818).
14. Skjulsvik, A. J.; Mørk, J. N.; Torp, M. O.; Torp, S. H. *Int. J. Clin. Exp. Pathol*., **2014**, *7*, pp 8905 – 8910 (PMCID: PMC4313958).
15. Shiras, A.; Bhosale, A.; Shepal, V.; Shukla, R.; Baburao, V. S.; Prabhakara, K.; Shastry, P. *Neoplasia*, **2003**, *5*, pp 520 – 532 (https://dx,doi.org/10.1016/s1476-5586(03)80036-2).
16. Charalambous, C.; Chen, T. C.; Hofman, F. M. *Neurosurg*., **2006**, *20*, E22 (https://doi.org/10.3171/foc.2006.20.4.e22).
17. McNamara, N. G.; Sahebjam, S.; Mason, W. P. *Cancers*, **2013**, *5*, pp 1103 – 1119 (https://doi.org/10.3390/cancers5031103).
18. Longuespe, R; Casadonte, R.; Kriegsmann, M.; Pottier, C.; Picard-de-Muller, G.; Delvenne, P. *Proteomics Clin. Appl*., **2016**, *10*, pp 701– 719 (https://doi.org/10.1002/prca.201500140).
19. Cole, L. M.; Clench, M. R. *Proteomics Clin. Appl*., **2015**, *9,* pp 335 – 341 (https://doi.org/10.1002/prca.201400103).
20. Lemaire, R.; Desmons, A.; Tabet, J. C.; Day, R.; Salzet, M.; Fournier, I. *J. Prot. Res*., **2007**, *6*, pp 1295 – 1305 (https://doi.org/10.1021/pr060549i).
21. Ronci, M.; Bonanno, E.; Colantoni, A.; Pieroni, L. *Proteomics*, **2008**, *8*, pp 3702 – 3714 (https://doi.org/10.1002/pmic.200701143).
22. Seeley, E. H.; Oppenheimer, S. R.; Mi, D.; Chaurand, P.; Caprioli, R. M. *J. Am. Soc. Mass Spectrom*., **2008**, *19*, pp 1069 – 1077 (https://doi.org/10.1016/j.jasms.2008.03.016).
23. Pratavieira, M.; Menegasso, A. R. S.; Esteves, F. G.; Sato, K. U.; Malaspina, O.; Palma, M. S. *J. Proteome Res*., **2018**, *17* (7), pp 2358 – 2369 (https://doi.org/10.1021/acs.jproteome.8b00098).
24. Nuñez, F. J.; Mendez, F. M.; Garcia-Fabiani, M. B.; Pardo, J.; Edwards, M.; Pedro R. Lowenstein, P. R.; María G. Castro, M. G. *J. Vis. Exp*., **2019**,*143* (https://doi.org/10.3791/58931).
25. Briggs, M. T.; Condina, M. R.; Ho, Y. Y.; Everest-Dass, A. V.; Mittal, P.; Kaur, G.; Oehler, M. K.; Packer, N. H.; Hoffmann, P. *Proteomics*, **2019**, *19* (21-22), e1800482 (https://doi.org/10.1002/pmic.201800482).
26. Cohen, A. L.; Colman, H. *Cancer Treat Res*., **2015**, *163*, pp 15 - 30 (https://doi.org/10.1007/978-3-319-12048-5_2).
27. Pisacane, A. M.; Picciotto, F.; Risio, M. *Anal. Cell. Pathol*., **2007**, *29*, pp 59 - 66.
28. Jung, C. S.; Foerch, C.; Schänzer, A.; Heck, A.; Plate, K. H.; Seifert, V.; Steinmetz, H.; Raabe, A.; Sitzer, M. *Brain*, **2007**, *130*, pp 3336 - 3341 (https://doi.org/10.1093/brain/awm263).
29. Mastronardi, L.; Guiducci, A.; Puzzilli, F.; Ruggeri, A. *J. Neurosurg. Sci*., **1999**, *43*, 263 – 270 (https://doi.org/10.1023/A:1016261900413).

Table S1. Proteomic data of the proteins identified in GMB tissue section

	Accession code	Protein identification	Function	Protein Score	% Sequence coverage	Tryptic Peptides Sequenced	*m/z*	Charge state
1	A8MTL9	Serpin-like protein HMSD	Inhibitor of serine endopeptidase	36	38	TDTEYVLR	996.64	+1
						TANGLFGEK	936.49	+1
						VNSWVADK	918.70	+1
2	P63313	Thymosin beta-10	Organization of the cytoskeleton	31	38	MADKPDMGEIASFDK	1670.31	+1
						ADKPDMGEIASFDKAK	1722.65	+1
3	P47756	F-actin-capping protein subunit beta	Organization of the cytoskeleton	38	25	MSDQQLDCALDLMR	1669.30	+1
						VVGKDYLLCDYNR	1557.29	+1
						LTSTVMLWLQTNKSGSGTMNLGGSLTR	2868.91	+1
4	Q5TAT5	Collagen alpha-1 chain	Adhesion cell-matrix; cellular differentiation	33	15	GFPGFPGPIGLDGKPGHPGPK	2028.41	+1
						GDMGLTGPPGQPGPQGQKGEK	2051.07	+1
						GQCGEYPHRECLSSMPAALR	2203.75	+1
						GEIGLPGPPGHDGEKGPR	1769.02	+1
5	Q02223	Tumor necrosis factor receptor superfamily member 17	Immunity; receptor and proto-oncogene	32	58	GTNAILWTCLGLSLIISLAVFVLMFLLR	3114.25	+1
						ISSEPLK	3396.41	+1
						SKPKVDSDHCFPLPAMEEGATILVTTK	932.51	+1
						NTGSGLLGMANIDLEKSR	2051.07	+1
						MLQMAGQCSQNEYFDSLLHACIPCQLR	2993.76	+1

Table S1. Proteomic data of the proteins identified in GMB tissue section (Continuation)

	Accession code	Protein identification	Function	Protein Score	% Sequence coverage	Tryptic Peptides Sequenced	*m/z*	Charge state
6	Q96QF1	RAB3A interacting protein	Protein transportation	38	17	LRSPSVLEVR	1154.40	+1
						TLVLSSSPTSPTQEPLPGGK	1995.04	+1
						STSSAMSGSHQDLSVIQPIVKDCK	2517.45	+1
						EADLSLYNEFR	1355.77	+1
						KCALTGQSK	934.57	+1
						LGYFKEEL	997.41	+1
7	Q16568	Cocaine- and amphetamine-regulated transcript protein	MAPKK activation; positive regulation of nerve impulses	38	26	VRLLPLLGAALLLMLPLLGTR	2258.31	+1
						LLPLLGAALLLMLPLLGTR	2003.57	+1
						LLPLLGAALLLMLPLLGTRAQEDAELQPR	3141.17	+1
8	Q99439	Calponin-2	Cytoskeleton regulation; cellular response to mechanical stimulus	32	30	NFDDATMKAGQCVIGLQMGTNK	2341.45	+1
						HLYDPKNHILPPMDHSTISLQMGTNK	3018.22	+1
						CASQVGMTAPGTRR	1449.44	+1
9	P57075	Ubiquitin-associated and SH3 domain-containing protein A	negative regulation of T-cell receptor	32	14	MAAGETQLYAKVSNK	1689.74	+1
						SRSSPSLLEPLLAMGFPVHTALK	2466.59	+1
						LSNLTRASFVSHYILQK	2056.09	+1
						MYTFSLATDLNSR	1533.41	+1
						DFENDPPLSSCGIFQSRIAGDALLDSGIR	3092.67	+1
10	Q9P2W3	Guanine nucleotide-binding protein G	Hormone-mediated signalization pathway	34	52	EVESLKYQLAFQR	1689.74	+1
						YQLAFQREMASK	1470.71	+1

Table S1. Proteomic data of the proteins identified in GMB tissue section (Continuation)

	Accession code	Protein identification	Function	Protein Score	% Sequence coverage	Tryptic Peptides Sequenced	*m/z*	Charge state
11	Q4G0S7	Coiled-coil domain-containing protein 152	unknown	40	29	MDQSSEGCMKK	1274.76	+1
						NNILDIQLEKSNCLLK	1857.04	+1
						GENEQLKISADLIK	1557.29	+1
						EEGYKK	752.88	+1
						EMEISELNAK	1178.17	+1
						LRSQEK	759.08	+1
						SQEKEK	747.01	+1
						KLQHFQEEK	1186.02	+1
12	Q8WZ60	Kelch-like protein 6	Signalization pathway of B-cell receptor	33	14	MLMAGQRGAWTMGDVVEK	2010.62	+1
						TGDLVEILNGEKVK	1514.10	+1
						HKPSER	752.88	+1
						FVAEVTCLDPLR	1362.09	+1
						LPLTEHELESENK	1537.69	+1
						TQCYDPSTNKWSLK	1669.30	+1
						AAMPVEAKCINAVSFR	1721.46	+1
13	Q9Y463	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	myoblast fusion to protein	32	17	MAVPPGHGPFSGFPGPQEHTQVLPDVR	2868.91	+1
						LSVDLIKTYK	1178.17	+1
						LLELMNQHDTEMKYYIVHLK	2549.25	+1
						HFMFR	752.88	+1
						NTHFRGVSLNLTR	1514.10	+1
						MTGGRPPLPPPDDPATLGPHLGLR	2642.29	+1

Table S1. Proteomic data of the proteins identified in GMB tissue section (Continuation)

	Accession code	Protein identification	Function	Protein Score	% Sequence coverage	Tryptic Peptides Sequenced	*m/z*	Charge state
14	P19961	Alpha-amylase 2B	Carbohydrate catabolism	31	11	WVDIALECER	1232.24	+1
						NMVTRCNNVGVR	1361.69	+1
						NWGEGWGFMPSDR	1537.45	+1
						MAVGFMLAHPYGFTR	1696.85	+1
15	Q9NW81	ATP-synthase subunit s-like protein	ATP synthesis	35	26	GIHRLGAAVAPEGNQK	1616.72	+1
						LGAAVAPEGNQKK	1281.57	+1
						TILQFLTNYFYDVEALRDYLLQR	2893.17	+1
						CCHVDDWCLSRLYPLADSLQELSLAGCPR	3262.76	+1
16	Q8NEM0	Microcephalin	Organizating center of microtubulins	33	11	QVTHVIFKDGYQSTWDK	2051.07	+1
						ENLSPTSSQMIQQSHDNPSNSLCEAPLNISR	3396.41	+1
						LSPTLSSTK	932.51	+1
						DLIKPHEELKK	1348.66	+1
						TLVMTSMPSEKQNVVIQVVDK	2344.97	+1
						VCAPENYLLSQ	1235.79	+1
17	Q14699	Raftlin	Membrane protein	37	17	FLEFTTLSAAELPGSSAVR	1995.04	+1
						LIPEFIKK	986.45	+1
						FVGVIPQYHSSVNSAGSSAPVSTANSTEDARDAK	3448.93	+1
						NQSPEPSSGPR	1154.40	+1
						AGDMGNCVSGQQQEGGVSEEMK	2242.40	+1

Table S1. Proteomic data of the proteins identified in GMB tissue section (Continuation)

	Accession code	Protein identification	Function	Protein Score	% Sequence coverage	Tryptic Peptides Sequenced	*m/z*	Charge state
18	Q9P021	Cysteine-rich PDZ-binding protein	Cell junction	38	42	MVCEKCEK	968.46	+1
						VCEKCEK	837.73	+1
						LNENKALTSK	1116.73	+1
						ARFDPYGK	1032.05	+1
						SSVHQPGSHYCQGCAYK	2010.72	+1
19	Q1L5Z9	LON peptidase N-terminal domain and RING finger protein	ATP-dependent proteolysis	38	13	EFLYCLALNPECNSVK	1841.89	+1
						VMCEVLFSATANVHENLTSSIQSR	2650.88	+1
						NFNITVLAEELIFRYLPDELSDR	2766.79	+1
						RIYDEEMSELSNLTR	1854.87	+1
20	Q9Y5Z7	Host cell factor 2	co-activator of transcription process	38	12	GDIPPGCAAHGFVCDGTR	1771.94	+1
						GVVPSPRESHTAVIYCK	1841.89	+1
						MDPHRQGSNNIVPNSINDTINSTK	2650.88	+1
						VETHATATPFSKETPSNPVATVK	2411.29	+1
						VAAINGCGIGPFSKISEFK	1937.37	+1
21	P46013	Proliferation marker protein Ki-67	to maintain individual mitotic chromosomes dispersed in the cytoplasm following nuclear envelope disassembly	61	3	EELLAVGK	858.49	+1
						MPCQSLQPEPINTPTHTK	2023.32	+1
						LTPSAGK	673.78	+1
						ESADGLQGETQLLVSRK	1830.95	+1
						TEAEQQITEVFVLAER	1862.94	+1

Table S1. Proteomic data of the proteins identified in GMB tissue section (Continuation)

	Accession code	Protein identification	Function	Protein Score	% Sequence coverage	Tryptic Peptides Sequenced	*m/z*	Charge state
22	P14136	Glial fibrillary acidic protein (GFAP)	is a cell-specific marker that, during the development of the central nervous system, distinguishes astrocytes from other glial cells.	75	12	SYVSSGEMMVGGLAPGR	1697.79	+1
						AEMMELNDR	1108.47	+1
						ELQEQLAR	986.52	+1
						EAASYQEALAR	1208.58	+1
23	P23297	Protein S100-A	binds calcium but binds zinc very tightly-distinct binding sites with different affinities exist for both ions on each monomer	83	53	ELLQTELSGFLDAQK	1691.98	+1
						DVDAVDK	761.36	+1
						MGSELETAMETLINVFHAHSGK	2531.18	+1
24	P28906	Hematopoietic progenitor cell antigen CD34	Possible adhesion molecule with a role in early hematopoiesis by mediating the attachment of stem cells to the bone marrow extracellular matrix or directly to stromal cells	78	19	PSLSPGNVSDLSTTSTSLATSPTK	2348.18	+1
						PYTSSSPILSDIK	1407.73	+1
						LGILDFTEQDVASHQSYSQK	2266.09	+1
						SWSPTGER	919.42	+1
25	P16284	Platelet endothelial cell adhesion molecule (CD31)	Cell adhesion molecule which is required for leukocyte transendothelial migration under most inflammatory conditions	67	8	SLPDWTVQNGK	1244.62	+1
						PELESSFTHLDQGER	1744.81	+1
						EQEGEYYCTAFNR	1609.65	+1
						DTETVYSEVR	1198.55	+1

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 84-90

FEATURE

PDF

COVID-19: How the Oswaldo Cruz Foundation is Working to Address the Pandemic in Brazil

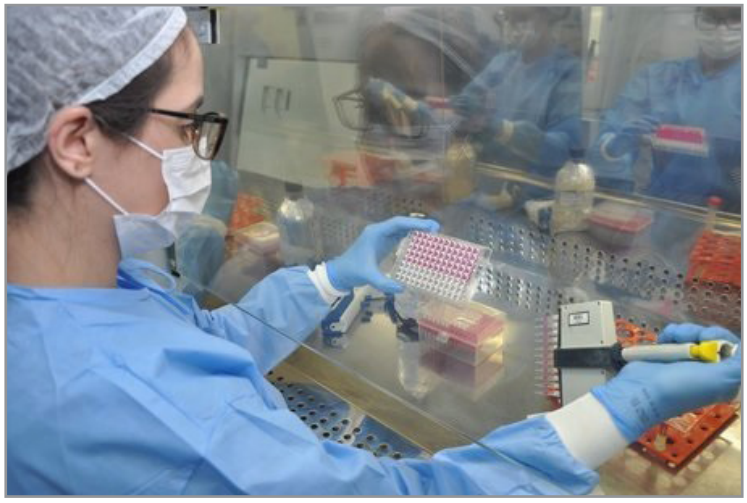
The COVID-19 pandemic caused by the new coronavirus (SARS-CoV-2) has presented itself as one of the greatest health challenges on a global scale at the beginning of this century. Insufficient scientific knowledge about the new coronavirus, its high speed of dissemination and the capacity to cause deaths in vulnerable populations has generated uncertainties about what would be the best strategies to use to face the pandemic in different parts of the world.

In Brazil, the challenges are even greater, as little is known about the transmission characteristics of the new coronavirus in the context of great social inequality, with populations living in precarious housing and sanitation conditions, without systematic access to water and in a situation of overcrowding. According to the National Council of Health Secretaries (Conass), from the beginning of the pandemic in March until October 8 2020, Brazil recorded 148,957 deaths from COVID-19, and 5,028,444 infected individuals. Given this scenario, institutions such as the Oswaldo Cruz Foundation (Fiocruz) are extremely important in the fight against COVID-19 in Brazil. One of the initiatives of this institution was the creation of the 'Observatory COVID-19 Fiocruz' whose general objective is the development of integrated analyses, technologies, proposals and solutions to face the pandemic by COVID-19 by the Brazilian Unified Health System (SUS) and by Brazilian society. The observatory is structured in a collaborative way, allowing the initiatives and work already developed in the different laboratories, research groups and sectors of Fiocruz to develop their activities in an agile way, with internal and external cooperation networks for the production and dissemination of materials to face the pandemic. Its work dynamics involves the production of information, dashboards and analyses, the development of technologies and communication.

COVID-19 is an infectious disease caused by the coronavirus of severe acute respiratory syndrome 2 (SARS-CoV-2) – Photo: Divulgação.

In addition, Fiocruz is also at the forefront of the development of one of the main vaccines in the fight against the new coronavirus. The vaccine ChAdOx-1 or AZD1222, also known as the Oxford vaccine,

is currently one of the most advanced in development in the world and, in Brazil, will be produced by Fiocruz, according to a technological order agreement with the pharmaceutical AstraZeneca, the holder of the patent. At Fiocruz, the production of this vaccine will be the responsibility of the Immunobiological Technology Institute (Bio-Manguinhos), which already has a history of manufacturing this type of product and incorporating technology. The funds for the adaptation of the Bio-Manguinhos plant and for the purchase of the first batch of vaccines will come from a national provisional measure that will allocate R$ 1.9 billion to Fiocruz. Dr. Nísia Trindade Lima, president of Fiocruz, foresees the production of up to 265 million doses of the ChAdOx-1 vaccine in 2021. The number cited by Dr. Nísia takes into account the forecast of the finalization of the initial 30 million doses of the vaccine in January 2021.



Brazilian Ministry of Health estimates production of 140 million doses of vaccine in the first half of 2021 – Photo: Divulgação/Fiocruz Minas.

The Oswaldo Cruz Foundation acts with the objective of promoting health and social development, generating and disseminating scientific and technological knowledge, and being an agent of citizenship. Fiocruz is linked to the Brazilian Ministry of Health and is the most prominent health science and technology institution in Latin America, operating in several cities in Brazil and internationally. Below, a brief history of each of Fiocruz's research centers is presented.

Oswaldo Cruz Foundation - The beginning

On May 25th 1900, the Federal Serum Therapeutic Institute was created at 'Fazenda de Manguinhos' in the North Zone of Rio de Janeiro city to manufacture sera and vaccines against the Bubonic Plague. Through the hands of the young Brazilian doctor and scientist Oswaldo Gonçalves Cruz, the institute was responsible for the sanitary reform that eradicated the Bubonic Plague and Yellow Fever epidemics from the city of Rio de Janeiro. Since then, the institution has experienced an intense trajectory, which is confused with the development of public health in Brazil.

Throughout the 20th century, Fiocruz experienced the many political transformations in Brazil. It lost its autonomy with the so-called "1930 Revolution", a coup that gave Getúlio Vargas the presidency of the Brazilian Republic. With the 1964 Brazilian civil-military coup, Fiocruz was hit by the so-called 'Massacre de Manguinhos', the impeachment of the political rights of some of its scientists. However, in the 1980s, Fiocruz experienced democracy again, and in an expanded way. Under the management of the sanitarist Sergio Arouca, from 1985 to 1988, Fiocruz had programs and structures recreated, and held its 1st Internal Congress, a landmark of modern Fiocruz; the Internal Congress is an event in which strategic issues related to the institutional macro-project, among others, are defined. The Internal Congress is held every

four years, always in the first year of each new Fiocruz presidency. In the following years, Fiocruz was the scene of major advances, such as the isolation of the HIV virus for the first time in Latin America. As a centenary institution, Fiocruz has a robust history in the early years of the 21th century. It expanded its facilities and, in 2003, its statute was finally published. It has achieved great scientific advances, with achievements such as the decoding of the bacillus Calmette–Guérin (BCG) genome, used in the vaccine for tuberculosis prevention.

Fiocruz Manguinhos, in Rio de Janeiro – Photo: César Guerra Chevrand (COC/Fiocruz).

Fiocruz has its headquarters in the neighborhood called 'Manguinhos', in the North Zone of Rio de Janeiro city, where a series of historical constructions are located that call attention for the beauty of the architecture. This group of buildings form the Manguinhos Historical Architectural Center is composed of buildings erected in the first decades of the 20th century. Fiocruz has other units installed in 10 states in the Northeast, North, Southeast and South regions of Brazil and has an office in Maputo, the capital of Mozambique, Africa.

Fiocruz Ceará

Since February 2009, Fiocruz has been present in the State of Ceará; its main initiative is the creation of the Northeastern Family Health Training Network (Renasf), carried out in partnership with teaching and research institutions in the region and with State and Municipal Health Secretariats.

Fiocruz in Ceará – Photo: Fiocruz Portal.

In 2018, the new headquarters of Fiocruz Ceará was built on a 10-hectare plot ceded by the State Government and located in the Technological and Industrial Pole of Health, in the municipality of Eusébio.

Fiocruz Amazonia

The Leônidas & Maria Deane Institute (ILMD/Fiocruz Amazonia) is Fiocruz's technical-scientific unit in the State of Amazonas. Based in Manaus, its mission is to contribute to the improvement of the living and health conditions of the Amazonian populations and to the regional scientific and technological development, integrating research, education and public health actions.

Fiocruz in Manaus – Photo: Divulgação.

For the development of its actions, ILMD/Fiocruz Amazonia establishes cooperation agreements with the other units of Fiocruz, with national and international partner institutions of Science, Technology and Innovation in Health that support multidisciplinary and inter-institutional projects, generating essential knowledge for the creation of public policies, which contribute to the improvement of society's quality of life. The production of scientific knowledge at ILMD/Fiocruz Amazonia also occurs through technical cooperation actions carried out through technical-scientific assistance developed alongside the National Unified Health System (SUS), with special focus on knowledge of the socio-sanitary and epidemiological realities of Amazonia.

Fiocruz Bahia

In 1950, the then governor of the State of Bahia created the Gonçalo Moniz Foundation in Salvador, the capital of Bahia. The purpose of this entity was to maintain and provide public health services, with focus on laboratory tests and the production of vaccines and sera required by Bahia, as well as the training of technical-specialized personnel in collaboration with universities and other research establishments.


Technological Park of Bahia will house the Fiocruz's Data and Knowledge Integration Center for Health – Photo: Nilton Souza/Divulgação.

Subsequently, in 1957, the Nucleus of Bahia Research (NEP) was created through an agreement between Fiocruz, the National Institute of Rural Endemics (INERU) and the Gonçalo Moniz Foundation. This collaboration aimed to study parasitic endemics in Bahia, at a time when collaboration between the federal and state governments was strengthened, developing a favorable environment for the establishment of activities related to science, technology and innovation in the area of health in Bahia.

In 1970, the NEP was incorporated into Fiocruz and became known as the Gonçalo Moniz Research Center (CPqGM). Ten years later, CPqGM became the Gonçalo Moniz Institute (IGM), a technical and scientific unit of Fiocruz. Currently, IGM, also known as Fiocruz Bahia, focuses on the study of infectious and parasitic diseases, in conducting anatomopathological exams, in addition to hosting two stricto sensu postgraduate courses at master and doctoral levels.

Fiocruz Brasilia

In 1976, during a time of political transformation and the constitution of the 'Sanitary Reform' movement in Brazil, the then vice-president of Fiocruz, Dr. Vinícius Fonseca, decided to establish a Fiocruz unit in Brasília, the capital of Brazil. At that time, the office was named Brasilia Regional Directorate (Direb) and was installed at the headquarters of the Pan American Health Organization (1976), and, later, in the building of the National Institute of Food and Nutrition – INAN (1977), which currently belongs to the National Ministry of Health. In the 1980s, Fiocruz Brasília changed its name to Fiocruz Regional Coordination (Coreb) and entered into strategic partnerships with the University of Brasília (UnB) and the Science and Technology Advisory of the Ministry of Health.

The 1990s began a new period of staff structuring and the expansion of activities, with regionalized actions and the formalization of institutional representation. In 1995, still as Coreb, the headquarters of Fiocruz Brasília started to occupy some rooms in the Ministry of Health, returning, in 1999, to the building of INAN. During the first decade of the 2000s, it was officially renamed Brasília Regional Directorate, later adopting the name Fiocruz Brasília, following the same strategy as other Fiocruz regional units.

Fiocruz in Brasilia – Photo: @fiocruzbrasilia.

In June 2010, the Fiocruz Brasília building was inaugurated at the UnB campus, three years after the foundation stone was laid. This building, with four independent blocks, houses the Fiocruz Brasília School of Government, the SUS Open University (UNA-SUS) and several projects and programs divided into areas of knowledge such as: Sanitary Law; Food, Nutrition and Culture; Education, Culture and Health; Health Promotion, Environment and Work; Epidemiology and Health Surveillance; Economy and Public Policies; and Bioethics and Health Diplomacy. Each of these areas develops varied activities, such as projects and research, courses, events and publications.

Fiocruz Mato Grosso do Sul

In 2007, a delegation composed of authorities from the State of Mato Grosso do Sul (MS) visited Fiocruz/RJ to express interest in hosting one of the Fiocruz's new regional units. In 2008, the first Fiocruz/MS seminar was held in the city of Bonito, MS, with representatives from teaching and research institutions and the management of the Unified Health System (SUS) in the Brazilian Midwest Region. The main objectives of the meeting were to discuss the importance of the new unit and define the main themes that should be prioritized by Fiocruz/MS: "Environment, Health and Agribusiness", "Health on the Frontiers", "Indigenous Health", "Most Relevant Diseases and Aggravates in the Midwest Region" and "Cerrado and Pantanal Formation".

The Master Plan of the Campus Fiocruz/MS foresees its implementation in stages associated with the availability of land, the budgetary disbursement and the expected implementation of the planned activities, thus assuming a commitment to economic, social and environmental sustainability. Stage I, which consisted of the construction of the administrative office building and support area, was inaugurated on December 8, 2011.

Fiocruz Minas

At the beginning of the 20th century, the Governor of the state of Minas Gerais proposed the creation of a branch of the Federal Serum Therapeutic Institute in Belo Horizonte, the capital of Minas Gerais, to Oswaldo Cruz. In 1907, the Institute René Rachou was created with its headquarters at 'Praça da Liberdade', in Belo Horizonte and, at the request of cattle ranchers in Minas Gerais, research was initiated

Fiocruz in Minas Gerais – Photo: Divulgação.

to develop a vaccine against the "Pest of Manqueira" or Carbúnculo Sintomático, epizooty that mainly attacked herds of cattle in Minas Gerais. The production and sale of this vaccine enabled the survival and expansion of the institution's activities in the following years to a large extent.

Nowadays, Fiocruz-Minas is mainly dedicated to the development of research on infectious and parasitic diseases such as Chagas disease, schistosomiasis, leishmaniasis and malaria, in addition to research into the epidemiology of aging, risky and occupational behavior.

Fiocruz Paraná

In 1999, the Carlos Chagas Institute (ICC) was born as a result of a partnership between Fiocruz – RJ and the government of the State of Paraná, through the Technological Health Park (TECPAR). Ten years later, the Carlos Chagas Institute became a regional technical-scientific unit of Fiocruz in Paraná, within the context of Fiocruz's nationalization project.

Located on the campus of the Paraná Institute of Technology, in the Industrial City of Curitiba, ICC/Fiocruz stands out as a productive and well-installed research center with nine research laboratories working in the areas of biochemistry, molecular biology and cell biology of infectious agents and their hosts, focused on studying the regulation of gene expression of microorganisms and parasites, the molecular characterization of stem cells, molecular virology and biotechnology.

Fiocruz in Paraná – Photo: Divulgação.

Fiocruz Pernambuco

Fiocruz in Recife – Photo: Diego Nigro/JC Imagem.

In 1950, the Aggeu Magalhães Institute (IAM) was opened in Recife, the capital of Pernambuco State. This institute belonged to the Sanitary Organization Division of the National Ministry of Education and Health. In the first decade of existence, the main studies carried out at the IAM were on schistosomiasis, mainly through studies of the ecology of mollusks and the action of molluskicides. Filariasis has also been the subject of studies, such as an epidemiological survey carried out in the neighborhood of 'Afogados' to raise the degree of prevalence of this disease in Recife.

In 1976, after the approval of Fiocruz's Programmatic Reorientation Plan by the Ministry of Health, the IAM became a special unit of Fiocruz. In 1987, IAM/Fiocruz started offering specialization courses in public health, through the then recently created Center for Studies in Collective Health. In 1991, the Multiprofessional Residency Program in Collective Health was established, in partnership with the State Health Secretariat.

Currently, IAM/Fiocruz offers graduate programs in academic and professional modality, as well as Lato Sensu specializations for health professionals. This institute meets the demand for highly complex diagnoses in the Chagas' disease Outpatient Clinics of the Federal University of Pernambuco (UFPE), the University of Pernambuco (UPE) and other institutions.

Fiocruz Piauí

In 2008, the project for the creation of a Fiocruz unit in the State of Piauí was born. The initiative was the result of a program called 'Mais Saúde' (more health) of the Brazilian federal government, which sought to improve public health management in locations that did not yet have basic healthcare.

In 2014, the activities of the Fiocruz office in Teresina, capital of Piauí, started with Master's courses. Currently, the Regional Technical Office of Fiocruz Piauí is located at the Regional Nucleus of the Ministry of Health, in Teresina. The consolidation of this office, which is under the coordination of the Vice-Presidency for Education, Information and Communication at Fiocruz, has been designed to contribute to the solution of the main health problems in the region and to improve the quality of life of the population, acting effectively in economic and technological developments, focused on research decentralization policies and the human resources training.

Fiocruz Rondônia

In 2009, Fiocruz's proposal to expand its presence in the national territory led to the establishment of Fiocruz's Technical Office in the city of Porto Velho, the capital of the State of Rondônia. Legally known as 'Fundação Oswaldo Cruz Noroeste', Fiocruz Rondônia incorporated the Institute for Research in Tropical Pathologies – IPEPATRO and expanded its activities in applied and epidemiological research, advanced and qualified training of human resources, and the provision of services, assistance and health surveillance. Marked by its insertion in the Western Amazon, Fiocruz Rondônia fills regional gaps, acting on infectious and parasitic pathologies, biotechnologies and vectors that transmit diseases of importance to public health.

Fiocruz Rondônia's activities will be expanded after the inauguration of its own building, which will allow Fiocruz to contribute even more to the reduction of regional inequalities in the fields of health and science and technology.

Fiocruz Africa

Fiocruz's first international office was opened on October 17, 2008, in Maputo, Mozambique. Fiocruz Africa offers Master's, doctoral and specialization courses in infectious and contagious diseases, public health, planning in health systems and attracting strategic human resources, among others. On the agenda of Fiocruz's actions on the African continent, are courses for the health area; also, exchanges in the area of vaccine production and diagnostic kits are planned.

Source: https://portal.fiocruz.br/

Brazilian Journal of Analytical Chemistry
2020, Volume 7, Issue 29, pp 91-99.

BrJAC

SPONSOR REPORT

PDF

This section is dedicated for sponsor responsibility articles.

Real-Time Search Enables a New Gold Standard for TMT Quantitation Accuracy on the Orbitrap Eclipse Tribrid Mass Spectrometer

Aaron M. Robitaille[1], Romain Huguet[1], Ryan Bomgarden[2], Jesse D. Canterbury[1], Daniel Lopez-Ferrer[1]

[1]Thermo Fisher Scientific, San Jose, CA
[2]Thermo Fisher Scientific, Rockford, IL

This report was extracted from the Thermo Scientific Technical Note 65729

Keywords: Real-Time Search for SPS MS^3, TMT & TMTpro Multiplexing, Quantitative Proteomics, Orbitrap Eclipse Tribrid MS, Mass Spectrometry

INTRODUCTION

Isobaric chemical tagging strategies using Thermo Scientific™ Tandem Mass Tags™ (TMT™) are powerful tools for studying global protein dynamics within a cell, tissue or an organism [1]. The primary advantage of the tags is their ability to simultaneously identify and quantify proteins in multiple sample sets in a single liquid chromatography coupled to mass spectrometry (LC-MS) analysis. The TMT reagent consists of an MS/MS reporter group, a spacer arm, and an amine-reactive group. The amine-reactive groups covalently bind to the peptide N-termini or to lysine residues. After labeling, the peptides are introduced into the mass spectrometer where each tag fragments during MS^2, producing unique reporter ions. Peptide quantitation is accomplished by comparing the intensities of the reporter ions. However, achieving quantitative accuracy is highly dependent on the purity of the precursor ion population selected for MS^2 analysis [2]. Innovations such as synchronous precursor selection (SPS) MS^3 technology are required to obtain accurate quantitation, and particularly important to measure subtle changes in low-abundance proteins [3]. The SPS MS^3-based methods provide higher accuracy compared to MS^2 methods for TMT quantitation. However, even with added selectivity of SPS MS^3, the accuracy can still suffer if non-specific fragments get selected for the MS^3 step. Furthermore, due to the requirement for an MS^3 scan, SPS MS^3 methods lower the scan rate of data acquisition. To improve upon the existing SPS MS^3 methods, we recently implemented a Real-Time Search step between the MS^2 and MS^3 scans on the Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer (Figure 1), which allows an MS^3 scan acquisition only if the MS^2 spectrum produces a positive peptide identification. In this application note we evaluate the benefits of Real-Time Search for SPS MS^3 implemented on the Orbitrap Eclipse Tribrid mass spectrometer to enhance TMT quantification accuracy and proteome coverage. We also apply this approach to the new Thermo Scientific™ TMTpro 16plex Label Reagent, which enable multiplexing of up to 16 samples [4]. Overall, we demonstrate that Real-Time Search sets a new gold standard for quantitative performance and throughput for isobaric tagging experiments.

MATERIAL AND METHODS

Materials used

- Thermo Scientific™ Pierce™ TMT11plex Yeast Digest Standard, P/N A40939
- TMTpro 16plex Label Reagent Set, P/N A44520
- Thermo Scientific™ EASY-Spray™ LC columns, 50 cm × 75 µm diameter, P/N ES803A
- Thermo Scientific™ EASY-Spray™ ion source
- Thermo Scientific™ EASY-nLC™ 1200 system
- Orbitrap Eclipse Tribrid mass spectrometer
- Thermo Scientific™ Proteome Discoverer™ 2.3 software

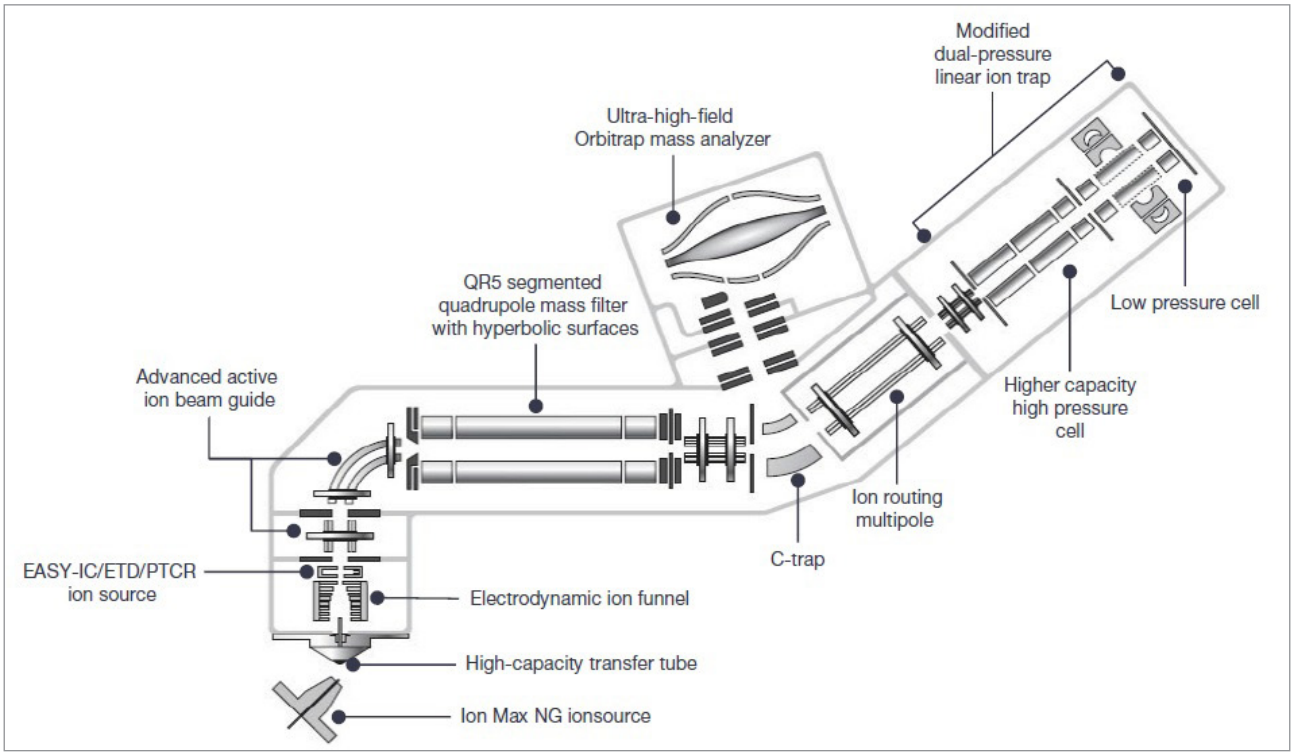


Figure 1. Schematic representation of the Orbitrap Eclipse Tribrid mass spectrometer that enables Real-Time Search.

Glossary of key terms

- Tandem Mass Tags (TMT): isobaric tags used for the precise quantitation of multiplexed samples
- Liquid Chromatography -Mass Spectrometry (LC-MS): combines physical separation and mass measurement
- Automatic Gain Control (AGC) target: specifies the maximum number of charges to accumulate for a given analysis
- Peptide Spectrum Matches (PSM): search engine match between theoretical and observed peptide spectra
- Synchronous Precursor Selection (SPS): MS^3 based method that increases TMT quantification accuracy
- Phase-Constrained Spectrum Deconvolution (ΦSDM): an advanced processing method for Fourier transform MS
- Quality Control (QC): standardized sample, LC/MS method, and data analysis workflow to assess system readiness
- Triple Gene Knockout (TKO): met6Δ, his4Δ, or ura2Δ yeast strains used to access ratio distortion for TMT quantitation
- Interference Free Index (IFI): calculated as average scaled protein abundance of (BY4741 - met6Δ, his4Δ, or ura2Δ)

Instrumentation settings and experimental setup

To assess sensitivity, accuracy, and precision of the Orbitrap Eclipse Tribrid mass spectrometer for TMT-based quantitation, we utilized the Pierce TMT11plex Yeast Digest Standard. This standardized sample provides users with a tool to measure the accuracy, precision, and proteome depth of TMT methods across different instrumentation. For a detailed user guide, please refer to "A quality control standard for Tandem Mass Tags (TMT) proteomic workflows" Technical Note 72968 [5]. In brief, for liquid chromatography (LC) conditions, we used an analytical gradient from 8% to 32% acetonitrile (vol/vol) with 0.1% (vol/vol) formic acid in 50 min with a column heater set to 45 ºC, unless otherwise indicated. Experiments were run with an EASY-nLC 1200 system in combination with an EASY-Spray C18 50 cm long column coupled to an EASY-Spray ion source. Samples were analyzed on an Orbitrap Eclipse Tribrid mass spectrometer. Settings for Thermo Scientific™ Orbitrap™ Tribrid™ Series Instrument Control Software Version 3.3 are found in Table 1. Raw data files were processed using Proteome Discoverer 2.3 software using the SEQUEST® HT search engine with a 10 ppm MS^1 and 0.5 Da MS^2 mass tolerance, TMT6plex (229.163 Da) or TMTpro16plex (304.207) set as a static modification, and a 1% false discovery rate (FDR).

Table 1. Mass spectrometer data acquisition settings

Acquisition settings	**MS^2**	**SPS MS^3 (with or without Real-Time Search)**
Top speed (s)	2.5	2.5
RF lens	30%	30%
Orbitrap MS^1 resolution	120,000	120,000
Scan range (*m/z*)	400–1400	400–1400
Standardized MS^1 AGC target	100%	100%
MS^1 max IT (mode)	Auto	Auto
Charge state	2-5	2-5
Dynamic exclusion (s)	45	45
MS^2 resolution	45,000	Turbo
MS^2 scan range (*m/z*)	First mass 110	200–1200
MS^2 isolation window	0.7 *m/z*	0.7 *m/z*
Standardized MS^2 AGC target	500%	100%
MS^2 max IT (mode)	Auto	Auto
MS^2 HCD NCE%	36	36
SPS MS^3 resolution	—	50,000
MS^2 scan range (*m/z*)	—	100–500
SPS MS^3 isolation window	—	0.7 *m/z*
Standardized SPS MS^3 AGC target	—	500%
SPS MS^3 max IT (mode)	—	Auto
SPS MS^3 HCD NCE%	—	55
SPS MS^3 notches	—	10

RESULTS

Deeper proteome coverage with the highest depth and quantitative accuracy for TMT-based quantitation

To improve upon existing SPS methods, we im666plemented a Real-Time Search filter between the MS^2 and MS^3 scans. Real-Time Search compares the raw data collected by the mass spectrometer to the information in the user provided FASTA database and identifies peptides from the mass spectra using the open source Comet search engine [6]. Real-Time Search benefits TMT SPS MS^3 methods in two distinct ways [7,8]. First, MS^3 scans are only triggered if a peptide-spectrum match (PSM) is identified from the preceding MS^2 spectrum. This increases the number of peptides identified with SPS MS^3. Secondly, Real-Time Search identifies precursor fragments ions that are specific to the identified peptide sequence, and then passes their masses to be selected for SPS MS^3 event. This adds extra specificity to the SPS MS^3 analysis, as it eliminates the possibility of a non-specific fragment to be included in the SPS MS^3 event. To assess the accuracy and sensitivity of TMT quantification methods, we utilized the Pierce TMT11plex Yeast Digest Standard (Figure 2) [5,9] analyzed with Proteome Discoverer software. TMT SPS MS^3 quantitation with Real-Time Search improves the quantitation accuracy, enabling up to 95% interference free measurements [10]. Real-Time Search enables on-thefly MS^2 data processing, matching the MS^2 spectrum acquired in the linear ion trap to the user-defined FASTA database. Advantageously, this results in near immediate determination of PSM, taking less than 5 ms per MS^2 spectrum when searching the UniProt human reference proteome, containing 74,449 proteins, with one variable modification and one missed cleavage. The search process is conducted concomitantly with the acquisition of the next MS^2 spectrum, in this way it does not impact the cycle times of the MS experiment. If a peptide spectrum match is made, the peptide is selected for subsequent SPS MS^3 based quantitation. With Real-Time Search, MS^3 scans are only triggered if a PSM is identified from the preceding MS^2 spectrum. This increases the number of peptides identified with Real-Time Search for SPS MS^3 to nearly the level that could be obtained in MS^2-only experiments while maintaining a significantly higher interference free index (IFI).

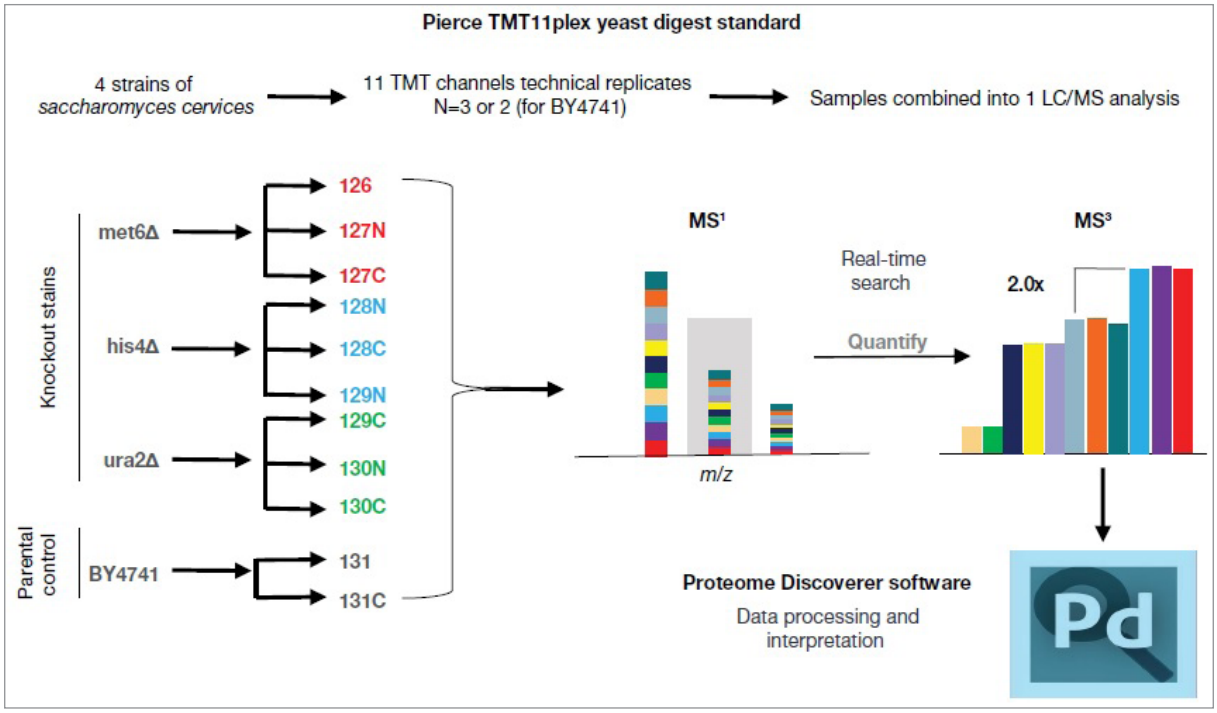


Figure 2. Schematic representation of Pierce TMT11plex Yeast Digest Standard method workflow. The standard is composed of four Saccharomyces cerevisiae strains with the three lines respectively lacking the non-essential proteins Met6, His4, or Ura2, and the parental strain BY4741 for reference channels.

A schematic representation of Real-Time Search for TMT SPS MS³ is shown in Figure 3.

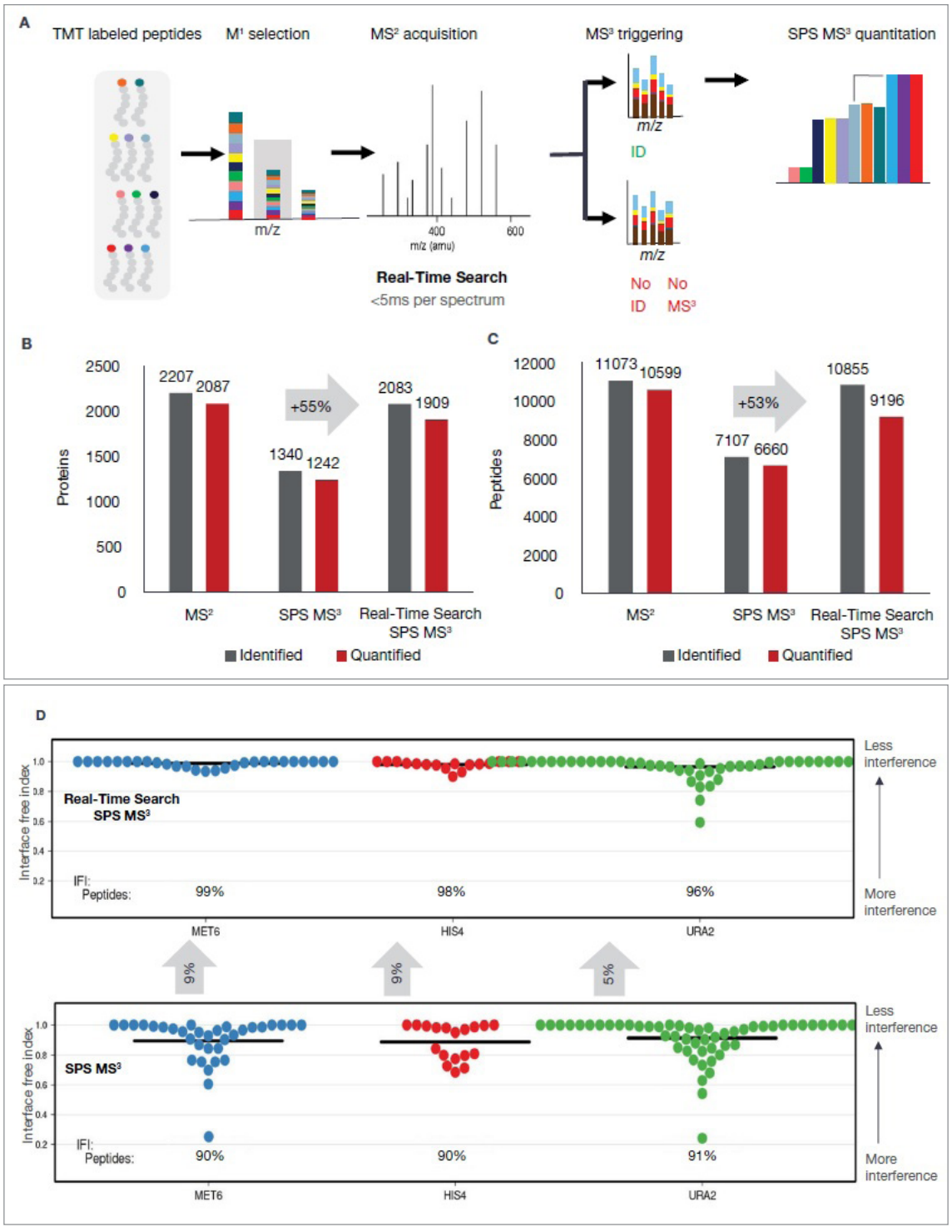


Figure 3. Achieving improved proteome depth and accuracy for SPS MS³ quantitation. (A) Schematic representation of TMT workflow on the Orbitrap Eclipse Tribrid mass spectrometer. If an MS² spectrum results in a PSM, a corresponding MS³ scan is triggered for quantitation. We evaluated how Real-Time Search influenced TMT identification rates for (B) proteins and (C) peptides. (D), the IFI was measured at the peptide level and visualized by TKOmics.com. 500 ng of TMT11plex Yeast Digest Standard was measured using a 50 min gradient using MS², SPS MS³, or Real-Time Search SPS MS³ methods.

We initially evaluated how Real-Time Search on the Orbitrap Eclipse Tribrid mass spectrometer influenced TMT peptide and protein identification rates. A 500 ng sample of Pierce TMT 11plex Yeast Digest Standard was analyzed on a 50 min gradient using MS^2, SPS MS^3, or Real-Time Search for SPS MS^3 methods. With Real-Time Search for SPS MS^3, 53% more peptides and 55% more proteins were identified than in the classic SPS MS^3 experiment, approaching the results of the MS^2 experiment. Additionally, Erickson et al. demonstrated that Real-Time Search for SPS-MS^3 could achieve the same proteome coverage as classical SPSMS^3, but in half the analysis time, thus further increasing sample throughput [7]. Importantly, Real-Time Search for SPS MS^3 results produced the highest IFI, i.e., the most accurate results, far surpassing what can be achieved in an MS^2 experiment. The IFI is calculated based upon the Pierce TMT11plex Yeast Digest Standard. This standard contains three yeast strains with gene deletions for met6Δ, his4Δ, or ura2Δ. In the channels where the genes have been deleted, any signal observed for the three proteins is due to co-isolation interference which can mask true differences in biological changes. An IFI of 0.75 indicates that twenty five percent of the observed signal came from co-isolated ions, incorrectly contributing to the reporter ion signal. This negatively affects quantitation and can produce a significant alteration if the interfering ions are expressed at higher abundance than the selected precursor. Thus, the IFI provides a method to measure if new approaches can reduce co-isolation interference and improve TMT quantitation accuracy [9].

Real-Time Search improves quantitative accuracy

Real-Time Search utilizes the open source search engine Comet to identify PSM in real time [6]. Importantly, post-acquisition data processing is required to accurately control the FDR. The results of searching the same dataset in an offline version of Comet with the online Real-Time Search implementation of Comet demonstrate a near perfect correlation. However, due to subtle differences in how search algorithms function, the choice of post-acquisition search engine other than Comet will influence both the identification and the assigned Xcorr (Figure 4A-B). An optional LC-MS survey experiment can be run with wide tolerances for Xcorr and dCn to further refine the MS^3 trigger thresholds. In addition to the doubled rate of peptide identification and quantitation, the quantitative accuracy is improved with Real-Time Search. In the traditional SPS MS^3 method, where fragment ions are selected based on the intensity, an n-notch experiment would select the n most abundant fragment ions from an MS^2 spectrum. In contrast, the Real-Time Search filter would select up to a maximum of n notches, which correspond to fragment ions from the identified PSM (Figure 4C-D), as well as specific b- and y-type ions that also contain the TMT label. For example, in the case of a peptide that is only N-terminally labeled, the y-type ions will not contain a TMT label; therefore, these fragments will not contribute to the overall TMT reporter ion signal and will be excluded from SPS by Real-Time Search. A portion of the SPS MS^3 signal is derived from the ions that do not belong to the peptide of interest (Figure 4E), since the fragment ions are selected based on intensity, and therefore lowered the accuracy of the experiment. When Real-Time Search is employed, 91% of the PSM quantified used only fragment ions from the correct peptide of interest (Figure 4F). The remainder of PSMs with non-matching fragment ions were likely due to subtle differences in the Real-Time Search engine (Comet) and the post-acquisition database search engine (SEQUEST HD). Overall, the added selectivity afforded by the Real-Time Search significantly improves quantitative accuracy of the experiment.

Real-Time Search can also be used with custom amino acid modifications and supports the analysis of commonly found post-translation modifications (PTMs) such as phosphorylation. The Orbitrap Eclipse Tribrid MS enables additional functionalities that improve the results of the quantitative experiment. These include Precursor Fit Filter, which allows for selecting ions based on precursor ion specificity of the isolation window, TurboTMT powered by the ΦSDM algorithm that increases scan acquisition rates [11,12], and a QR5 segmented quadrupole mass filter with its improved precursor ion selectivity and transmission, making it possible to use narrower isolation widths to further improve the precursor ion specificity and with that, the TMT quantitation accuracy.

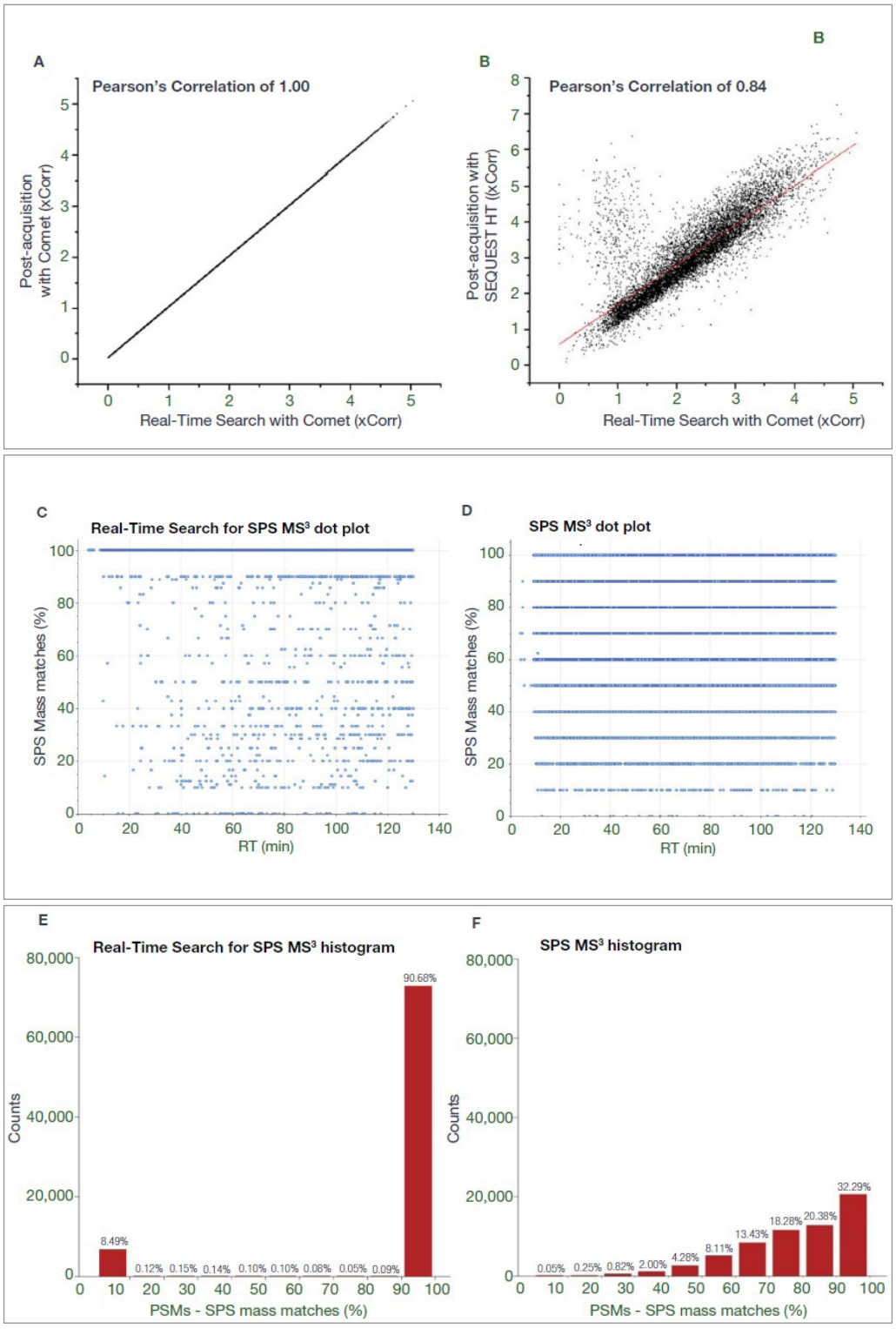


Figure 4. Real-Time peptide identification for TMT quantitation. Real-Time Search utilizes the open source search engine Comet. We evaluated the correlation of Xcorr scores between (A) Real Time Search with Comet and post-acquisition search with Comet, or (B) Proteome Discoverer software using SEQUEST HD. Secondly, we evaluated the number of fragment ions selected for a 10-notch SPS MS³ experiment that correctly matched the post-acquisition identification in Real-Time Search for SPS MS³ (C and E) or SPS MS³ (D and F). Selection of a fragment ion that does not correspond to the correct peptide negatively impacts quantitation in classical SPS-MS³.

TMTpro 16plex isobaric label reagent enables up to 16 samples to be quantified simultaneously

Increasingly, biological experiments are requiring higher multiplexing capability and sample throughput. Applications such as thermal shift assays, time courses, drug dose curves, large scale translational cohort studies, hyperLOPIT subcellular localization studies, or whole cell interactomics approaches such as QMIX [13] require an extensive number of comparisons to be able to measure statistically meaningful differences. TMTpro 16plex isobaric label reagents are next generation reagents that increase the level of sample multiplexing up to sixteen without compromising identification or quantitation. The TMTpro tag is isobaric and amine reactive, similar to the classic TMT tags, but differ by the incorporation of a longer spacer region and isobutyl proline mass reporter region. After fragmentation within the mass spectrometer, the TMTpro reagents generate unique reporter ions having masses in the range of 126 to 134 Da that are used for relative quantitation of the isobaric precursor ions. The TMTpro 16plex isobaric label reagent enables increased sample throughput and results in fewer missing values for quantitation when analyzing up to 16 samples in a single LC-MS analysis or across multiplex experiments.

To test the performance of the new tags, we labeled the Thermo Scientific™ Pierce™ 6 Protein Digest Standard with the TMTpro 16plex isobaric label reagent. The sample was then spiked into the Thermo Scientific™ Pierce™ HeLa Protein Digest Standard at a ratio of 1:8 as shown in Figure 5A. The combined mixture was then analyzed on the Orbitrap Eclipse Tribrid MS using the SPS MS^3 method with and without utilizing Real-Time Search. We observed that Real-Time Search increased the numbers of quantified proteins by 26%, while also improving both the precision and accuracy of quantitation (Figure 5D). This improvement in quantitation accuracy for large ratios is particularly useful for single cell proteomics approaches that utilize TMT-based quantitation [14].

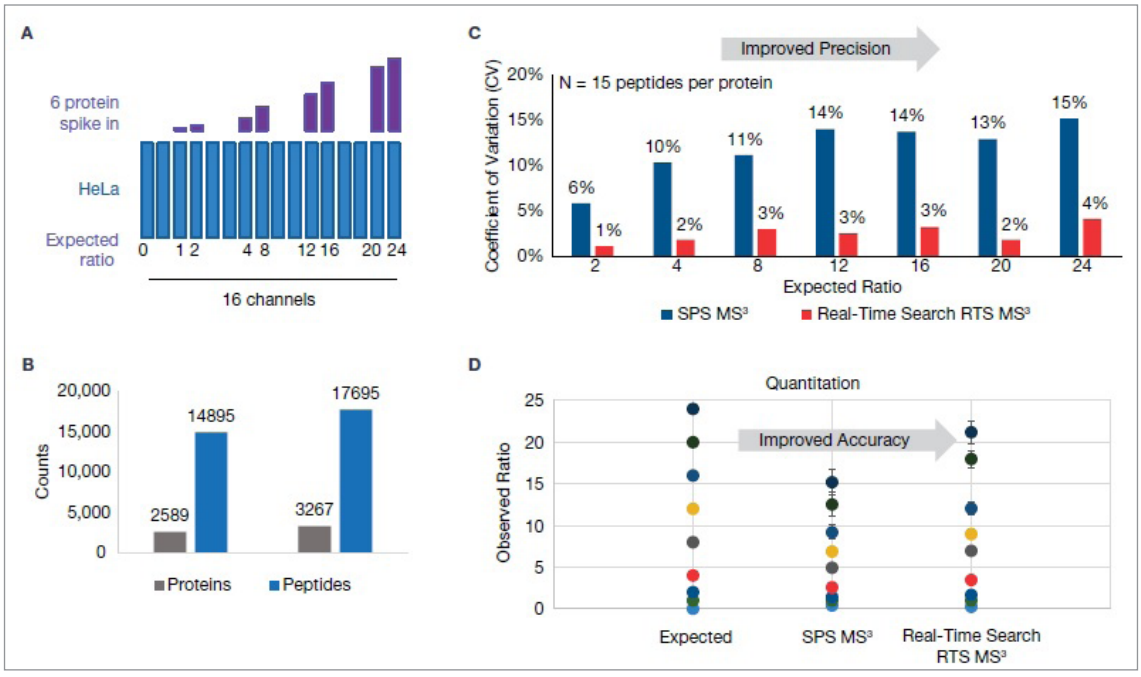


Figure 5. Next generation TMTpro isobaric tags increase sample multiplexing. We evaluated the potential for Real-Time Search to use custom modifications using next generation TMTpro isobaric tags (A). An equimolar mixture of Pierce 6 Protein Digest (red) was mixed in various ratios into the Pierce HeLa Protein Digest Standard and labeled (B). 1 µg of sample was then analyzed on a 120 min gradient using Real-Time Search for SPS MS^3 or SPS MS^3 on the Orbitrap Eclipse Tribrid MS. Data was analyzed using Proteome Discover 2.3 software. We analyzed how Real-Time Search effected TMT SPS MS^3 quantitation precision (C) and accuracy (D). This work is for research purposes only.

CONCLUSION

We evaluated the utility of Real-Time Search for TMT SPS MS^3-based quantitation, which, in addition to other new features including Precursor Fit Filter and TurboTMT, allow for intelligent acquisition methods that improve quantitation accuracy, precision, and proteome coverage. The combination of isobaric tag reagents, high resolution accurate mass instrumentation, and advanced software enable the identification of thousands of proteins multiplexed from up to 16 samples in a single run while achieving accurate and precise quantitation including of low level precursors and/or small ratios. Together, the TMT and TMTpro multiplexing works enables the investigation of novel biological questions, such as the heterogeneity of single cell protein expression across a tissue.

Acknowledgements

We would like to thank Jimmy Eng from the University of Washington for assistance implementing Comet, as well as Devin Schweppe, Qing Yu, and Steven P. Gygi from the Department of Cell Biology at Harvard Medical School in Boston, USA for their collaboration and input.

REFERENCES

1. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. *Anal. Chem*., **2003**, *75* (8), pp 1895–1904 (https://doi.org/10.1021/ac0262560).
2. Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B. *Mol. Cell Proteomics*, **2008**, *7* (9), pp 1702–1713 (https://doi.org/10.1074/mcp.M800029-MCP200).
3. McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. *Anal. Chem*., **2014**, *86* (14), pp 7150–7158 (https://doi.org/10.1021/ac502040v).
4. Li, J.; Van Vranken, J. G.; Pontano Vaites, L. et al. *Nat. Methods*, **2020**, *17* (4), pp 399–404 (https://doi.org/10.1038/s41592-020-0781-4).
5. Robitaille, A. M.; Choi, J.; Viner, R.; Huhmer, A.; Opperman, K.; Rogers, J. C. Thermo Fisher Scientific, Technical Note 72968, **2019** [available from:https://planetorbitrap.com/library?t=QTIzMzBlODg0OWE1NWYzYw%3D%3D&keywords=A2330#tab:keywords].
6. Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. *Proteomics*, **2013**, *13* (1), pp 22–24 (https://doi.org/10.1002/pmic.201200439).
7. Erickson, B. K.; Mintseris, J.; Schweppe, D. K.; Navarrete-Perea, J.; Erickson, A. R.; Nusinow, D. P.; Paulo, J. A.; Gygi, S. P. *J Proteome Res*., **2019**, *18* (3), pp 1299–1306 (https://doi.org/10.1021/acs.jproteome.8b00899).
8. Schweppe, D. K.; Eng, J. K.; Bailey, D.; Rad, R.; Yu, Q.; Navarrete-Perea, J.; Huttlin, E. L.; Erickson, B. K.; Paulo, J. A.; Gygi, S. P. *J. Proteome Res*., **2020**, *19* (5), pp 2026–2034 (https://doi.org/10.1021/acs.jproteome.9b00860).
9. Paulo, J. A.; O'Connell, J. D.; Gygi, S. P. *J. Am. Soc. Mass Spectrom*., **2016**, *27* (10), pp 1620–1625 (https://doi.org/10.1021/jasms.8b05143).
10. Gygi, J. P.; Yu, Q.; Navarrete-Perea, J.; Rad, R.; Gygi, S. P.; Paulo, J. A. *J. Proteome Res*., **2019**, *18* (2), pp 687–693 (https://doi.org/10.1021/acs.jproteome.8b00737).
11. Kelstrup, C. D.; Aizikov, K.; Batth, T. S.; Kreutzman, A.; Grinfeld, D.; Lange, O.; Mourad, D.; Makarov, A.; Olsen J. V. *J. Proteome Res*. **2018**, *17* (11), pp 4008–4016 (https://doi.org/10.1021/acs.jproteome.8b00381).
12. Grinfeld, D.; Aizikov, K.; Kreutzmann, A.; Damoc, E.; Makarov, A. Phase-constrained spectrum deconvolution for fourier transform mass spectrometry. *Anal. Chem*., **2017**, *89* (2), pp 1202–1211 (https://doi.org/10.1021/acs.analchem.6b03636).
13. Yu, C.; Huszagh, A.; Viner, R.; Novitsky, E.J.; Rychnovsky, S.D.; Huang, L. *Anal. Chem*., **2016**, *88* (20), pp 10301–10308 (https://doi.org/10.1021/acs.analchem.6b03148).
14. Dou, M.; Clair, G.; Tsai, C. F.; Xu, K.; Chrisler, W. B.; Sontag, R. L.; Zhao, R.; Moore, R. J.; Liu, T.; Pasa-Tolic, L.; et al. *Anal. Chem*., **2019**, *91* (20), pp 13119–13127 (https://doi.org/10.1021/acs.analchem.9b03349).

This Sponsor Report is the responsibility of Thermo Fisher Scientific.

RELEASE

Orbitrap Eclipse Tribrid Mass Spectrometer

Obtain high confidence insights for very complex molecules

The Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer combines industry-leading versatility and usability with the exceptional performance afforded by its built-in intelligence and revolutionary Thermo Scientific™ Tribrid™ architecture to deliver an instrument designed to address the most difficult analytical challenges.

This newest Tribrid platform includes Advanced Ion Management Technology (AIM+) with the new QR5 segmented quadrupole mass filter, Real-Time Search, Enhanced Vacuum Technology, optional Proton Transfer Charge Reduction (PTCR), and optional High Mass Range MS^n (HMR^n) mode. Collectively, these features make this instrument uniquely suited for accurate and high-throughput full-proteome quantitation, characterization of complex mixtures of protein or small molecule-based pharmaceuticals, and deciphering higher-order protein structures. The Orbitrap Eclipse Tribrid MS makes it easy to tackle these difficult analyses and collect the high-quality data required to drive the right decisions, pushing your science beyond today's discovery.

Features and Benefits of the Orbitrap Eclipse Tribrid MS

Power high impact science with performance advancements

Address complex challenges more effectively with the latest improvements in calibration, ion transmission efficiency, mass range, data acquisition speed and much more!

Gain ultimate flexibility with the ability to expand your experimental scope

You need LC-MS workflows to push today's limits. The Orbitrap Eclipse Tribrid MS comes with enhanced sensitivity, selectivity and more MS^n techniques, scan modes and intelligence. It's now possible to go beyond your scientific pursuits with the ability to expand your experimental scope.

Obtain highest quality MS data with built-in intelligence

The Orbitrap Eclipse Tribrid MS eliminates convoluted optimization procedures by providing pre-defined data acquisition methods, pre-set instrument parameters and built-in intelligence for workflows including TMT. No compromises on data quality needed.

Improve your laboratory productivity with a single system that provides maximum insights

The ability to go beyond today's discovery is now possible with one mass spectrometer built to solve analytical challenges for proteomics, structural biology, small molecule research and biopharmaceutical services.

Access more resources at: planetorbitrap.com

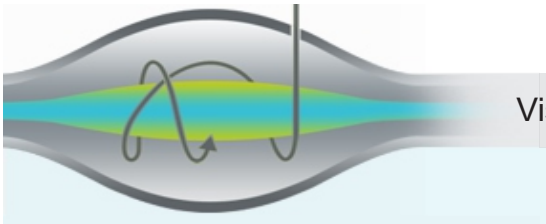
thermo scientific

Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ Mass Spectrometer

Obtain high confidence insights for very complex molecules with the Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer. It combines industry-leading versatility and usability with the exceptional performance afforded by its built-in intelligence and revolutionary Thermo Scientific™ Tribrid™ architecture to deliver an instrument designed to address the most difficult analytical challenges. This newest Tribrid platform includes Advanced Ion Management Technology (AIM+) with the new QR5 segmented quadrupole mass filter, Real-Time Search, Enhanced Vacuum Technology, optional Proton Transfer Charge Reduction (PTCR), and optional High Mass Range MSn (HMRn) mode. Collectively, these features make this instrument uniquely suited for accurate and high-throughput full-proteome quantitation, characterization of complex mixtures of protein or small molecule-based pharmaceuticals, and deciphering higher-order protein structures. The Orbitrap Eclipse Tribrid MS makes it easy to tackle these difficult analyses and collect the high-quality data required to drive the right decisions, pushing your science beyond today's discovery.

VIDEO

WEBSITE



Visit the Planet Orbitrap: planetorbitrap.com/

ThermoFisher SCIENTIFIC

RELEASE

Pittcon 2021 Conference & Expo

Don't Miss the World's Leading Conference and Expo on Laboratory Science

Science is a dynamic and constantly evolving discipline. Every day, all over the world, scientists are making new discoveries; formulating new medicines to enhance the quality of life and solving global environmental, energy and food related challenges.

Pittcon, a vital resource for knowledge, happens yearly to help keep you informed of, connected to and up-to-date on these significant ongoing findings and new instrumentation.

Cutting Edge Research

The World's Leading Technical Program

With more than 2,000 technical sessions, Pittcon makes it easy for you to get connected to the latest research and developments from world-renowned scientists from around the globe.

Co-Programming Partners

Co-programming offers a mutually beneficial venture which permits organizations to combine ongoing efforts with Pittcon to help achieve the common goal of advancing the chemical and laboratory sciences on a global level.

Some of the partner organizations are: American Chemical Society, Coblentz Society, International Association of Environmental Analytical Chemistry, Japan Analytical Instruments Manufacturers Association, Society for Applied Spectroscopy, Society for Electroanalytical Chemistry, and others.

Courses are offered at beginner, intermediate & advanced levels.

With more than 100 from which to choose, there are a wide variety of classes covering relevant analytical topics in food science, water/wastewater, environmental, life science, pharmaceutical. Courses for broad-based application and general lab functions include lab management, quality control, technical writing, statistics, data management, and lab safety.

Who Should Attend?

Anyone who develops, buys, or sells laboratory equipment, performs physical or chemical analyses, develops analysis methods, or manages these scientists.

WHEN: March 6 – 10, 2021
WHERE: New Orleans, Louisiana, USA

For more information visit https://pittcon.org

The robust technical program offers the latest research in more than 2,000 technical presentations covering a diverse selection of methodologies and applications.

Pittcon also offers more than 100 skill -building short courses in a wide range of topics

CALL FOR PROPOSALS
Symposia proposals being accepted

PITTCON 2021

March 6-10, 2021
New Orleans,
Louisiana, USA

What is Pittcon?

Pittcon is the world's leading annual conference and exposition on laboratory science. Pittcon attracts attendees from industry, academia and government from over 90 countries worldwide.

PITTCON™
CONFERENCE & EXPO

RELEASE

SelectScience® Pioneers online Communication and Promotes Scientific Success since 1998

SelectScience® promotes scientists and their work, accelerating the communication of successful science. SelectScience® informs scientists about the best products and applications through online peer-to-peer information and product reviews. Scientists can make better decisions using independent, expert information and gain easy access to manufacturers.

SelectScience® informs the global community through Editorial, Q&A and Application Articles, Featured Topics, Event Coverage, Video and Webinar programs.

Some recent contributions from SelecScience® to the scientific community

Editorial Article

How Nanotrap particles could help COVID-19 monitoring and outbreak containment

University of Connecticut (UConn) microbial ecologist Dr. Kendra Maas reveals how Ceres Nanosciences' Nanotrap® particles have saved her invaluable time as she builds an ambitious COVID-19 tracking system. In an exclusive interview, Select Science spoke with Dr. Kendra Maas to find out how she is developing a monitoring tool to analyze SARS-CoV-2 levels in wastewater. Access **here**

Webinar

Essential technology for sub-ppm mass accuracy and high-resolution untargeted metabolomics

Dr. Chris Petucci provides an overview of the University of Pennsylvania Metabolomics Core and how it performs targeted and untargeted LC-MS metabolomics for internal and external research investigators and industry. Attend this webinar **here**

Expert Insight

Ambient ionization mass spectrometry in space and time: From high throughput analysis to molecular imaging

Explore the applications and benefits of ambient ionization for biomedical research and its future potential for pharma and food industries. Professor Zoltan Takats, of Imperial College London, will discuss the utility of the rapid evaporative ionization mass spectrometry (REIMS) in the pharmaceutical or food industry as well as synthetic biology. Access **here**

SelectScience® is the leading independent online publisher connecting scientists to the best laboratory products and applications.

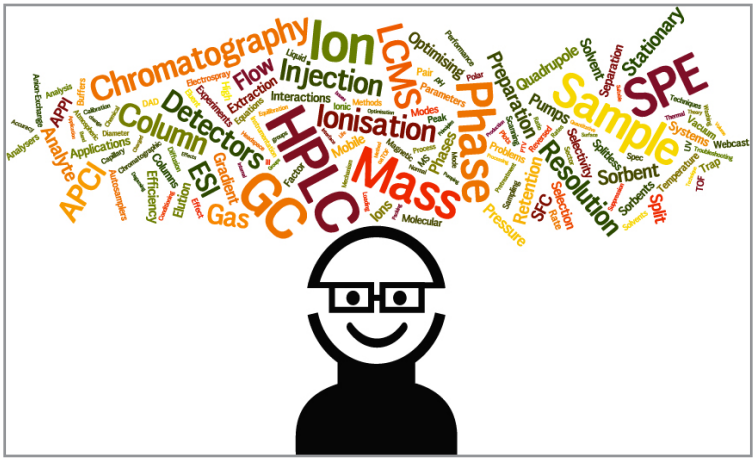
Access 2 Million+ Decision Makers

- Working with Scientists to Make the Future Healthier.
- Informing scientists about the best products and applications.
- Connecting manufacturers with their customers to develop, promote and sell technologies.

SelectScience®
The Fastest Way to Expert Opinion

RELEASE

CHROMacademy helps Increase your Knowledge, Efficiency and Productivity in the Lab



CHROMacademy is the world's largest eLearning website for analytical scientists. With a vast library of high-quality animated and interactive eLearning topics, webcasts, tutorials, practical information and troubleshooting tools CHROMacademy helps you refresh your chromatography skills or learn something completely new.

A subscription to CHROMacademy provides you with complete access to all content including:

- **Thousands of eLearning topics covering HPLC / GC / Sample Prep / Mass Spec / Infrared / Basic Lab Skills / Biochromatography**
- **Video Training Courses**
- **Ask the Expert – 24-hour Chromatography Support**
- **Assessments**
- **Full archive of Essential Guide Webcasts and Tutorials**
- **Application Notes and *LCGC* Articles**
- **Troubleshooting and Virtual Lab Tools**
- **User Forum**

Lite members have access to less than 5% of CHROMacademy content.
Premier members get so much more!
For more information, please visit www.chromacademy.com/subscription.html

LC GC's CHROMacademy
powered by crawford scientific

CHROMacademy Lite members have access to less than 5% of our content. Premier members get so much more !

Video Training courses

Fundamental HPLC
Fundamental GC
Fundamental LCMS
Fundamental GCMS
HPLC Method Development
GC Method Development

Ask the Expert

We are always on hand to help fix your instrument and chromatographic problems, offer advice on method development, help select a column for your application and more.

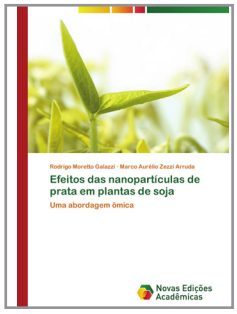
To find out more about Premier Membership contact:

Glen Murry: +1 732.346.3056 | Glen.Murry@ubm.com
Peter Romillo: +1 732.346.3074 | Peter.Romillo@ubm.com

www.chromacademy.com

The worlds largest e-Learning website for analytical scientists

NOTICES OF BOOKS

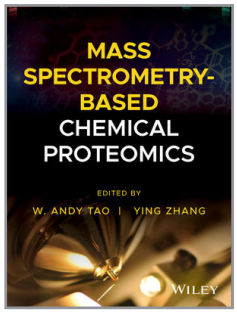


Efeitos das Nanopartículas de Prata em Plantas de Soja: Uma Abordagem Ômica

Rodrigo Moretto Galazzi, Marco Aurélio Zezzi Arruda, Authors

July 2020. Publisher: *Novas Edições Acadêmicas*

This book presents a compilation of studies related to the interaction of nanoparticles with plants and some results on the effects of silver nanoparticles on soybean culture, which has high global demand and economic value. The exposure of these plants to silver nanoparticles results in changes in the content of proteins, enzymes, nutrients, among others, alerting to the possible adverse effects of nanoparticles to living organisms. Read more …

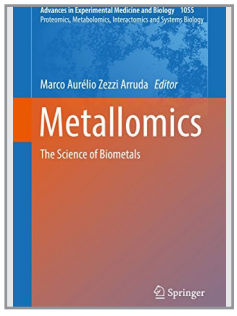


Mass Spectrometry-Based Chemical Proteomics

W. Andy Tao, Ying Zhang, Editors

July 2019. Publisher: John Wiley & Sons

This book provides strategies and concepts for understanding chemical proteomics, and analyzing protein functions, modifications, and interactions – emphasizing mass spectrometry throughout. Chapters cover those many technical advances and applications in drug discovery, from target identification to validation and potential treatments. Read more …

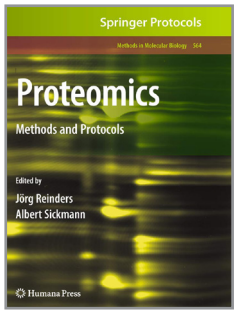


Metallomics - The Science of Biometals

Marco Aurélio Zezzi Arruda, Editor

June 2018. Publisher: Springer International Publishing

As Metallomics is intrinsically a transdisciplinary area, this book is authored by experts in Environmental, Nuclear, and Human Metallomics fields. Within these topics metals play important role, as being part of biomolecules, controlling different biochemical process, being signaling agents, being catalyst of biochemical reactions, among others. This volume demonstrates the importance of more investigation about metals and their interactions with biomolecules. Read more …

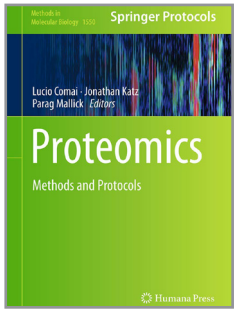


Proteomics: Methods and Protocols / *Methods in Molecular Biology Series*

Jörg Reinders, Albert Sickmann, Editors

2009. Publisher: Springer

In this book, experts in proteomic research have contributed detailed techniques, which present an overview of the contemporary challenges and possibilities in the various areas of proteomics, as well as cutting-edge protocols that provide concrete examples for successful proteome analysis. Authoritative and easy-to-use, this book will serve as a valuable reference book for laboratories around the world and raise interest in the complex and vital field of proteomics research. Read more …



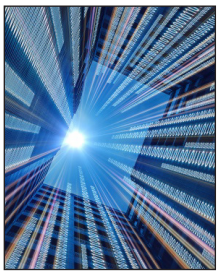
Proteomics: Methods and Protocols / *Methods in Molecular Biology Series*

Lucio Comai, Jonathan Katz, Parag Mallick, Editors

2017. Publisher: Springer

This volume aims to provide protocols on a wide range of biochemical methods, analytical approaches, and bioinformatics tools developed to analyze the proteome.
Authoritative and cutting-edge, this book aims to ensure successful results in the further study of this vital field. Read more …

PERIODICALS & WEBSITES

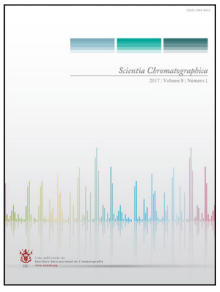


American Laboratory

The American Laboratory® is a platform that provides comprehensive technology coverage for lab professionals at all stages of their careers. Unlike single-channel publications, American Laboratory® is a multidisciplinary resource that engages scientists through digital, mobile, multimedia, and social channels to provide practical information and solutions for cutting-edge results. Addressing basic research, clinical diagnostics, drug discovery, environmental, food and beverage, forensics, and other markets, American Laboratory combines in-depth articles, news, and video to deliver the latest advances in their fields. Read more

LCGC

Chromatographyonline.com is the premier global resource for unbiased, peer-reviewed technical information on the field of chromatography and the separation sciences. Combining all of the resources from the regional editions (LCGC North America, LCGC Europe, and LCGC Asia-Pacific) of award winning magazines, Chromatographyonline delivers practical, nuts-and-bolts information to help scientists and lab managers become more proficient in the use of chromatographic techniques and instrumentation, thereby making laboratories more productive and businesses around the world more successful. Read more



Scientia Chromatographica

Scientia Chromatographica is the first and to date the only Latin American scientific journal dedicated exclusively to Chromatographic and Related Techniques (Mass Spectrometry, Sample Preparation, Electrophoresis, etc.). With a highly qualified and internationally recognized Editorial Board, it covers all chromatography topics (HPLC, GC, SFC) in all their formats, in addition to discussing related topics such as “The Pillars of Chromatography”, Quality Management, Troubleshooting, Hyphenation (GC-MS, LC-MS, SPE-LC-MS/MS) and others. It also provides columns containing general information, such as: calendar, meeting report, bookstore, etc. Read more

Select Science

SelectScience® promotes scientists and their work, accelerating the communication of successful science. SelectScience® informs scientists about the best products and applications through online peer-to-peer information and product reviews. Scientists can make better decisions using independent, expert information and gain easy access to manufacturers. SelectScience® informs the global community through Editorial, Features, Video and Webinar programs. Read more

Spectroscopy

Spectroscopy’s mission is to enhance productivity, efficiency, and the overall value of spectroscopic instruments and methods as a practical analytical technology across a variety of fields. Scientists, technicians, and laboratory managers gain proficiency and competitive advantage for the real-world issues they face through unbiased, peer-reviewed technical articles, trusted troubleshooting advice, and best-practice application solutions. Read more

EVENTS – It is suggested to consult the event's official website for updates.

Octorber 5 – 16, 2020
VIRTUAL 43rd Annual Meeting of the Brazilian Chemical Society (43rd RASBQ)
http://www.sbq.org.br/43ra/

October 7 – 8, 2020
FCE Pharma Virtual Sessions
https://www.fcepharma.com.br/sessions/

October 26 – 29, 2020
1th Online Brazilian Congress on Analytical and Environmental Chemistry (CONQUIAMB)
https://congresse.me/eventos/conquiamb/

November 9 – 11, 2020
Integrated Online National Meeting on Forensic Chemistry (7th EnqFor) & 4th Meeting of the Brazilian Society of Forensic Sciences (SBCF)
https://www.en.enqfor2020.sbcf.org.br/

November 24 – 27, 2020
Digital Analitica Latin America Congress (7th Analitica Congress)
https://www.analiticanet.com.br/

November 30 – December 2, 2020
17th Congress on Quality in Metrology Virtual Sessions (ENQUALAB 2020)
https://www.enqualab.net/

December 10 – 11, 2020
International Conference on Metrology, Measurement and Inspection (ICMMI 2020)
New York City, USA
https://waset.org/metrology-measurement-and-inspection-conference-in-december-2020-in-new-york

---

October 11 – 15, 2021
34th Latin American Congress of Chemistry – CLAQ 2020; 18th Latin American Congress of Chromatography – COLACRO; 10th Colombian Congress of Chromatography – COCOCRO; 4th Colombian Congress of Biochemistry and Molecular Biology - C2B2
Convention Center, Cartagena de Indias, Colombia
https://claq2020.com/en/bienvenida/

October 24 – 27, 2021
20th National Meeting on Analytical Chemistry (20th ENQA) & 8th Ibero-American Congress of Analytical Chemistry (8th CIAQA)
Dall'Onder Grande Hotel, Bento Gonçalves, RS, Brazil
https://enqa2021.com.br/

December 12 – 16, 2021
XXIII International Mass Spectrometry Conference (IMSC 2021)
Windsor Oceânico Hotel, Rio de Janeiro, RJ, Brazil
https://www.imsc2020.com/

1th CONQUIAMB

BRAZILIAN CONGRESS ON ANALYTICAL CHEMISTRY AND ENVIRONMENTAL

26 – 29 OCTOBER, 2020 — ONLINE

FREE REGISTRATION

GUIDELINES FOR AUTHORS

Scope

The *Brazilian Journal of Analytical Chemistry* (BrJAC) is dedicated to the diffusion of significant and original knowledge in all branches of Analytical Chemistry and Bioanalytics. BrJAC is addressed to professionals involved in science, technology and innovation projects in Analytical Chemistry, at universities, research centers and in industry.

Professional Ethics

Manuscripts submitted for publication in BrJAC cannot have been previously published or be currently submitted for publication in another journal. BrJAC publishes original, unpublished scientific articles and technical notes that are peer reviewed in the double-blind way.

Review process

BrJAC's review process begins with an initial screening of the manuscripts by the editor-in-chief, who evaluates the adequacy of the study to the journal scope, and an analysis of similarities. Manuscripts accepted in this screening are then forwarded to at least two reviewers. A larger number of reviewers may be used at the discretion of the Editor. As evaluation criteria, the reviewers employ originality, scientific quality, contribution to knowledge in the field of Analytical Chemistry, the theoretical foundation and bibliography, the presentation of relevant and consistent results, compliance to the BrJAC's guidelines, and the clarity of writing and presentation, and use of grammatically correct English.

Note: In case the Editors and Reviewers consider that the manuscript requires an English revision, the authors will be required to send an English proofreading certificate, by the ProofReading Service or equivalent service, before the final approval of the manuscript by BrJAC.

BrJAC is a quarterly journal that, in addition to scientific articles and technical notes, also publishes reviews, interviews, points of view, letters, sponsor reports, and features related to analytical chemistry.

Brief description of the documents that can be submitted by the authors

- **Articles:** Full descriptions of an original research finding in Analytical Chemistry. Articles undergo double-blind full peer review.
- **Reviews:** Articles on well-established subjects, including a critical analysis of the bibliographic references and conclusions. Manuscripts submitted for publication as Reviews must be original and unpublished. Reviews undergo double-blind full peer review.
- **Technical Notes:** Concise descriptions of a development in analytical method, new technique, procedure or equipment falling within the scope of BrJAC. Technical notes also undergo double-blind full peer review.
- **Letters:** Discussions, comments, suggestions on issues related to Analytical Chemistry, and consultations to authors. Letters are welcome and will be published at the discretion of the BrJAC editor-in-chief.

Manuscript submission

BrJAC uses the online Manuscript Manager System for the submission of manuscripts. This system guides authors stepwise through all the submission process. The texts must be uploaded in Word files, which will be converted to a single PDF file used in the double-blind peer review process. All correspondence, including notification of the Editor's decision and requests for revision, is sent by e-mail. Please submit your manuscript at **www.manuscriptmanager.net/brjac**

The mandatory files to be uploaded are: **Cover letter**, **Title Page, Manuscript** and **Novelty statement**. **Templates** for these files are available at http://brjac.com.br/guidelines-download.php

Manuscript preparation

- **Language**: **English** is the language adopted by BrJAC. The correct use of English is of utmost importance. In case the Editors and Reviewers consider that the manuscript requires an English revision, the authors will be required to send an English proofreading certificate before the final approval of the manuscript by BrJAC.
- **Required items**: the manuscript must include a title, an abstract, keywords, and the following sections: Introduction, Materials and Methods, Results and Discussion, Conclusion, and References.
- **Identification of authors**: as BrJAC adopts double-blind review, the manuscript file must **NOT** contain the authors' names, affiliations nor acknowledgments. Full details of the authors and their acknowledgements should be on the **Title Page**.
- **Layout**: the lines in the manuscript must be numbered consecutively and double-spaced.
- **Graphics and Tables**: must appear close to the discussion about them in the manuscript. For **figures** use **Arabic** numbers, and for **tables** use **Roman** numbers.
- **Permission to use content already published**: for figures, graphs, diagrams, tables, etc. identical to others previously published in the literature, the author must ask for publication permission from the company or scientific society holding the copyrights, and send this permission to the BrJAC editor-in-chief with the final version of the manuscript.
- **Chemical nomenclature**: should conform to the rules of the International Union of Pure and Applied Chemistry (IUPAC) and Chemical Abstracts Service. It is recommended that, whenever possible, authors follow the International System of Units, the International Vocabulary of Metrology (VIM) and the NIST General Table of Units of Measurement. Abbreviations are not recommended except those recognized by the International Bureau of Weights and Measures or those recorded and established in scientific publications.
- **References throughout the manuscript:** the references must be cited by **numbers in square brackets [ ]**. It is recommended that references older than 5 (five) years be avoided, except in relevant cases.
- **References item:** include a DOI (as a hyperlink) whenever it is available. **References should be thoroughly checked for errors by the authors before submission.**
  See how to format the **References item** below.

Examples of reference formatting

Journals

1. Orlando, R. M.; Nascentes, C. C.; Botelho, B. G.; Moreira, J. S.; Costa, K. A.; Boratto, V. H. M. *Anal. Chem*., **2019**, *91* (10), pp 6471-6478 (https://doi.org/10.1021/acs.analchem.8b04943).
   - *Publications with more than 10 authors, list the first 10 authors followed by a semicolon and et al.*
   - *Titles of journals must be abbreviated as defined by the Chemical Abstracts Service Source Index (http://cassi.cas.org/search.jsp).*

Electronic journals

2. Sapozhnikova, Y.; Hoh, E. *LCGC North Am*., **2019**, *37* (1), pp 52-65. Available from: http://www.chromatographyonline.com/suspect-screening-chemicals-food-packaging-plastic-film-comprehensive-two-dimensional-gas-chromatogr [Accessed 20 January 2019].

Books

3. Burgot, J.-L. *Ionic Equilibria in Analytical Chemistry*. Springer Science & Business Media, New York, **2012**, Chapter 11, p 181.
4. Griffiths, W. J.; Ogundare, M.; Meljon, A.; Wang, Y. Mass Spectrometry for Steroid Analysis. In: Mike, S. L. (Ed.). *Mass Spectrometry Handbook*, v. 7 of Wiley Series on Pharmaceutical Science and Biotechnology: Practices, Applications and Methods. John Wiley & Sons, Hoboken, N.J., **2012**, pp 297-338.

Standard methods

5. International Organization for Standardization. ISO 26603. Plastics — *Aromatic isocyanates for use in the production of polyurethanes — Determination of total chlorine*. Geneva, CH: ISO, **2017**.

Master's and doctoral theses or other academic literature

6. Dantas, W. F. C. *Application of multivariate curve resolution methods and optical spectroscopy in forensic and photochemical analysis*. Doctoral thesis, **2019**, Institute of Chemistry, University of Campinas, Campinas, SP, Brazil.

Patents

7. Trygve, R.; Perelman, G. US 9053915 B2, June 9, **2015**, Agilent Technologies Inc., Santa Clara, CA, US.

Web pages

8. http://www.chromedia.org/chromedia [Accessed 10 January 2019].

Unpublished source

9. Viner, R.; Horn, D. M.; Damoc, E.; Konijnenberg, A. *Integrative Structural Proteomics Analysis of the 20S Proteasome Complex* (WP-25). Poster presented at the XXII International Mass Spectrometry Conference (IMSC 2018) / August 26-31, **2018**, Florence, IT.
10. Author, A. A. *J. Braz. Chem. Soc*., in press.
11. Author, B. B., **2019**, submitted for publication.
12. Author, C. C., **2019**, unpublished manuscript.

Note: *Unpublished results may be mentioned only with express authorization of the author(s). Personal communications can be accepted exceptionally.*

Revised manuscript submission

Based on the comments and suggestions of the reviewers and editors a revision of the manuscript may be requested to the authors. A revised manuscript should be submitted by the authors, containing the changes made in the manuscript clearly highlighted. Responses to the reviewers must also be submitted answering in detail to the questions made by them, and describing the changes made in the manuscript.

Copyright

When submitting their manuscript for publication, the authors agree that the copyright will become the property of the Brazilian Journal of Analytical Chemistry, if and when accepted for publication. The copyright comprises exclusive rights of reproduction and distribution of the articles, including reprints, photographic reproductions, microfilms or any other reproductions similar in nature, including translations.

Final Considerations

Whatever the nature of the submitted manuscript, it must be original in terms of methodology, information, interpretation or criticism. As to the contents of published articles, the sole responsibility belongs to the authors, and *Braz. J. Anal. Chem*. and its editors, editorial board, editorial office and collaborators are fully exempt from any responsibility for the data, opinions or unfounded statements.

BrJAC reserves the right to make, whenever necessary, small alterations to the manuscripts in order to adapt them to the journal rules or make them clearer in style, while respecting the original contents. The article will be sent to the authors for approval prior publication.