Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

CHIMICA

Ciclo XXX

Settore Concorsuale: 03/A1

Settore Scientifico Disciplinare: CHIM/01

METABOLOMICS INVESTIGATIONS TOWARDS FORMULATED NATURAL COMPLEX PRODUCTS BY UNTARGETED AND TARGETED MASS SPECTROMETRY-BASED APPROACHES

Presentata da: SARA TAMIMI

Coordinatore Dottorato

Prof. Aldo Roda

Supervisore

Prof. Aldo Roda

Co-supervisore

Dott.ssa Luisa Mattoli

Esame finale anno 2018

Acknowledgments

First of all I would like to thank Aboca S.p.a. and in particular my co-supervisor Dr. Luisa Mattoli who first believed in this project, giving me the possibility to work at the Phytochemistry Research laboratories in which advanced technology and passion are the principal ingredients of every new research project. Then I would like to thank Professor Aldo Roda who gave me the opportunity to attend this PhD program with the University of Bologna.

Thanks to all the colleagues who worked with me and who contributed to the realization of this work.

Above all I would like to thank Bruno, he knows why… and obviously my mother, without her nothing would have been possible.

TABLE OF CONTENTS

List of Abbreviations

ADME	Absorption distribution metabolism excretion
APCI	Atmospheric pressure chemical ionization
API	Active pharmaceutical ingredient
ASE	Accelerated solvent extraction
CEF	Compound export format
CE-MS	Capillary electrophoresis-mass spectrometry
CID	Collision-induced decomposition
DIMS	Direct infusion mass spectrometry
DModX	Distance to the model
EI	Electron ionization
EIC	Extracted ion chromatogram
EMA	European medicines agency
EP	European pharmacopoeia
ESI	Electrospray ionization
EU	European union
FAB	Fast atom bombardment
FDA	Food and drug administration
FIA-ESI-MS	Electrospray ionization mass spectrometry flow injection analysis
FT-ICR-MS	Fourier transform-ion cyclotron resonance-mass spectrometry
GC-MS	Gas chromatography-mass spectrometry
GC-FID	Gas chromatography-flame ionization detector
GMP	Good manufacturing protocol
GS	Golden standard
HCA	Hierarchical cluster analysis
HIV	Human immunodeficiency virus
HMDB	Human metabolome database
HMPC	Committee on herbal medicinal products
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
ICP-MS	Inductively coupled plasma-mass spectrometry
IT	Ion trap
LC-DAD	Liquid chromatography-diode array detector
LC-MS	Liquid chromatography-mass spectrometry

LC-RID	Liquid chromatography-refractive index detector
LLE	Liquid-liquid extraction
LTQ-Orbitrap MS	Linear trap quadrupole-Orbitrap mass spectrometry
MAE	Microwave assisted extraction
MCBTs	Multicomponent botanical therapeutics
MFC	Median fold change
MS	Mass spectrometry
MSI	Metabolomics standard initiative
MSPC	Multivariate statistical process control
MVDA	Multivariate data analysis
NCEs	New chemical entities
NMR	Nuclear magnetic resonance
NP	Natural product
NPs	Natural products
PC	Principal component
PCA	Principal component analysis
PCDL	Personal compound database and library
PLS	Partial least square
Q	Quadrupole
QC	Quality control
QQQ	Triple quadrupole
Q-TOF	Quadrupole-time of flight
RP	Reversed-phase
SD	Standard deviation
SFE	Supercritical fluid extraction
SOP	Standard operating procedure
SPE	Solid phase extraction
TIC	Total ion chromatogram
TOF	Time of flight
UAE	Ultrasound assisted extraction
UHPLC	Ultra high performance liquid chromatography
USD	United states dollar
WHO	World Health Organization

ABSTRACT

Crude plant extracts and natural products deriving from them have been used since ancient in folk medicine for treating a broad range of human diseases and, according to the World Health Organization (WHO) reports, plant-derived traditional medicines are still playing an important role in the health care systems of developed countries. However, although the last decade has witnessed a marked growth in the market of plant-based natural products, their high complexity in terms of composition makes a challenging task the guarantee of quality, efficacy and safety requirements. In such scenario, the recently approved European regulation EU 2017/745 has created a breaking point in the regulation of medical devices forcing the manufactures to investigate the overall qualitative composition of devices made of substances and quantitative information for the main constituents or for the components responsible for the desired effect. In view of this stricter standardization of plant-based medical devices, an intense research towards advanced technologies and robust analytical protocols is required both for evaluating the composition in metabolites and for reaching the required quality standards.

Based on these considerations, in this *Philosophiae Doctor* Thesis, Grintuss® syrup was selected as a case study of formulated natural complex products for quali-quantitative metabolomics investigations. In particular, in the first part of this dissertation, an untargeted metabolomic approach for monitoring the batch-quality of Grintuss® adult syrup by electrospray ionization mass spectrometry flow injection analysis (FIA-ESI-MS) was successfully developed (**Chart I**). The proposed workflow has involved the preliminary sample preparation followed by its analysis by FIA-ESI-MS. Thus, after having assessed the validity of the model by using multivariate statistical process control (MSPC), the quality verification of new batches under investigation was achieved by comparing their profile with the fingerprinting profile of batches prepared according to the optimized manufacturing process, and validated by

traditional continuous quality verification throughout the production chain, and the profile obtained from batches deriving from pre-established production errors.

The second task of this doctoral work was aimed at profiling the metabolites contained in Grintuss® pediatric syrup by applying a targeted metabolomic LC-MS-based approach (**Chart I**). In pursuing this aim, an *in-house* database of high-purity standard reference compounds was built by collecting compounds belonging to different chemical classes. Thus, after the acquisition of MS and MS/MS spectra of standard compounds of interest, the matching between the detected metabolites from samples of Grintuss® syrup and standard reference compounds from the database was assessed by following the All Ions MS/MS targeted screening workflow from Agilent Technologies. Finally, the quantitative analysis of the recognized metabolites was performed and the results thus achieved have been combined with information deriving from the analysis of additional compounds as mono-, di- and polysaccharides, fats and minerals allowing to reach the knowledge of >99.9% of the overall composition of Grintuss® pediatric syrup.

Chart I – Targeted and untargeted mass spectrometry-based approaches for the metabolomics investigation of Grintuss® syrup.

1 Introduction

1.1 Plants as an endless source of biologically active metabolites

Nature has been a priceless source of inspiration since ancient.[1] Indeed, if on one side natural products (NPs) from plants and animals have been a rich source of biologically-active compounds, still continuing to enthral the scientific community for the impressive complexity and diversity of chemotypes, on the other side chemists have always tried to mime Mother Nature reproducing artificially NPs themselves.[2] Despite the reduced investments for NP-based drug discovery especially in the big Pharma during the last two decades,[3] nowadays the deeper understanding of the biological pathways, the unremitting advances in the "omics" disciplines and the availability of novel investigating hyphenated techniques allow the rapid identification and characterization of new biologically-active secondary metabolites as an endless source of new medicines. Plants, in particular, represent the main source of biologically-active NPs and secondary metabolites (**Box 1**).[4]

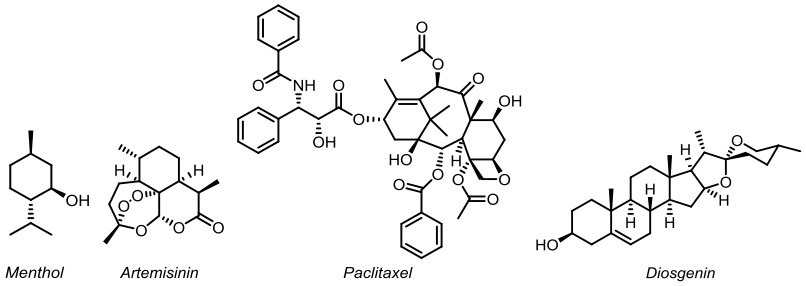
Box 1. Main classes of pharmacologically-active secondary plant metabolites.

Plant secondary metabolites are defined as molecules that do not possess recognized roles in the maintenance of fundamental life processes in the plants but are mainly involved in modulating the interaction of the plant with its environment. The production of these compounds is often low and strongly depends on the physiological and developmental stage of the plant. Secondary metabolites are characterized by enormous chemical diversity and every plant has its own characteristic set of secondary metabolites.[5] Based on their biosynthetic origins, plant secondary metabolites can be structurally divided into three major groups which include terpenoids, alkaloids and phenolics, while glycosides, tannins and saponins are part of them according to their specific structure.[6]

Terpenoids

Terpenoids are polymeric isoprene derivatives synthesized from acetate *via* the mevalonate pathway. The

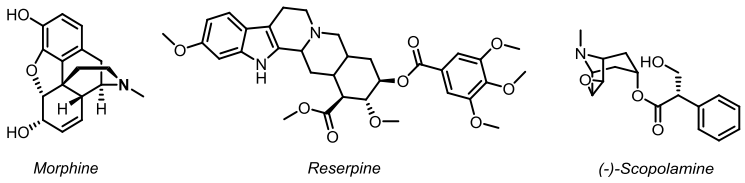
high volatility of some terpenoids provides, for sessile plants, a tool for communicating with other organisms such as neighboring plants, pollinators and foes of herbivores, *via* air-bone info-chemicals. Recent findings demonstrate that certain nitrogenous terpenes possess potent anti-hypertensive activity, while the antimicrobial and insecticidal properties of other terpenoids have led to their utilization as pesticides and fungicides in agriculture.[6]



Representative examples of biologically-active terpenoids and steroids

Alkaloids

Alkaloids represent a group of secondary metabolites characterized by the presence of basic nitrogen atoms, although some related compounds with neutral and weakly acid properties are also included in this group. In addition to carbon, hydrogen and nitrogen, this group may also contain oxygen, sulfur and rarely other elements such as chlorine, bromine and phosphorus. Most of them are toxic to other organisms and can be extracted by acid-base. They are biosynthesized from amino acids such as tyrosine and are endowed with diverse pharmacological effects. Compared with most other classes of secondary metabolites, alkaloids are characterized by a great structural diversity and there is no uniform classification of them.[6]

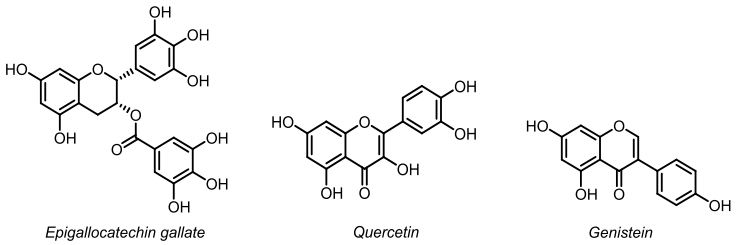


Representative examples of biologically-active alkaloids

Phenols

Phenolic compounds are one of the largest group of secondary plants constituents present in fruits, vegetables, teas, cocoa and other plants that possess certain health benefits. They are endowed with antioxidant, anti-inflammatory, anti-carcinogenic and other biological properties, and may protect from oxidative stress and some diseases. Phenolic compounds are distributed in almost all plants and are diverse in structure, while sharing the presence of aromatic rings as the common feature. Most of the phenolic compounds are polymerized into larger molecules such as proanthocyanidins and lignans. Furthermore, phenolic acids may occur in food plants as esters or glycosides conjugated with other natural compounds such as flavonoids, alcohols, hydroxyfatty acids, sterols, and glycosides. Phenols can

be classified in sub-families based on the number of hydroxyl groups (1-, 2- and polyatomic phenols), the chemical composition (mono-, di-, oligo- and polyphenols) and the number of aromatic rings present in the skeleton (one aromatic ring, two aromatic rings, quinones and polymers). Among them, polyphenolic compounds are the most common secondary metabolites in plants, including more than 8000 different compounds that share similar chemical structures but with some distinctive differences. Based on these differences, polyphenols can be classified into flavonoids, tannins and glycosides. The sub-family of *flavonoids* includes anthocyanins, flavones and flavonols, a series of water-soluble pigments found in the vacuoles of plant cells. They are widely distributed in plants, fulfilling many functions such as flower coloration, producing yellow, red or blue pigmentation in petals designed to attract pollinator animals. They have become very popular because of their health benefits which include anti-allergic, anti-cancer, antioxidant, anti-inflammatory and anti-viral activity. *Tannins* are water soluble phenolic compounds (with exception of some high molecular weight structures) generally in form of oligomers and polymers that are able to form complexes and precipitates with proteins, starch, cellulose and minerals. Tannins are synthesized *via* shikimic acid pathway, also known as the phenylpropanoid pathway. Tannins are used as astringent against diarrhea, diuretic against stomach and duodenal tumors, and anti-inflammatory agents. *Glycosides* may be phenol, alcohol or sulfur compounds. They are characterized by a sugar portion or moiety attached to one or non-sugar portions. Many plants store chemicals in the form of inactive glycosides, which can be activated by enzymatic hydrolysis. *Saponins* are largely distributed in plants and are characterized by the presence of a glycoside whose aglycone is related to a triterpenoid or steroidal sub-structure. They are endowed with particular biological (antimicrobial, antioxidant, hemolytic) and physicochemical (emulsification, solubilization, foaming) properties exploited in food, cosmetics and pharmaceutical industries.



Representative examples of biologically-active phenols

Historically, the earliest record concerning the use of plants for medical purposes dates back to 2600 B.C. on clay tablets (Mesopotamia). These include oils from *Cedrus* (cedar), *Cupressus sempervirens* (cypress), *Glycyrrhiza glabra* (licorice) and *Commiphora* (myrrh) species, and juice from *Papaver somniferum* (poppy), which are still used today to treat coughs, colds and inflammation.[7] The Ebers Papyrus (2900 B.C.) represents the best known example of pharmaceutical record which documents the use as medicine of over 700 plant-based drugs

ranging from gargles and pills to infusions. The Chinese Materia Medica (1100 B.C., 52 prescriptions)[8] and the Indian Ayurvedic system (1000 B.C., more than 500 drugs)[9] are others well documented records concerning the use of natural products. After that, the Greek physician Dioscorides (100 A.D.) recorded the collection, storage and use of medicinal herbs, while Galen, a practitioner and teacher of pharmacy and medicine in Rome, is known for his key contribute in the prescription and formulation of plant-based natural products used in therapy. During the Middle Ages, the Arabs expanded the Greco-Roman expertise by including new NPs deriving also from Chinese and Indian herbs. However, since the 18th century, medicinal plants were used without any rationalized mechanistic knowledge of their pharmacological activities. In this regard, Anton von Störck and William Withering, with their investigations towards poisonous herbs laid the basis for the rational use of medicinal plants.[10]

Therefore, at the beginning of the 19th century, numerous bioactive components from medicinal plants have been identified, isolated and extensively studied, while apothecaries, the progenitors of the modern pharmaceutical companies, specialized itself in the purification and formulation of biologically-active NPs.[11] Subsequently, several efforts were made in order to make available NPs at lower costs by chemical synthesis. After the "penicillin era", with the advancements in the field of chemistry in the mid-twentieth century, formulated single natural products superseded the crude plant extracts generally utilized by apothecaries at the beginning of 1900.[12] In recent years, plants have been extensively investigated for their therapeutic potential and have been the forefront of Ayurvedic and folk medicine to treat several human diseases.[1a] Indeed, in 1985 the World Health Organization (WHO) estimated that approximately 65% of the world population relied on plant-derived traditional medicines for their primary health care, while plant products also play an important indirect role in the health care systems of developed countries.[13] In this regard, however, we are still so far from exploring the whole chemical space of NPs since more than 90% of the world's biodiversity, estimated at 2 million species of plants, animals, fungi and micro-organisms, has not been evaluated for any potential biological activity.[1a,3c,14] Furthermore, restricting the field to the terrestrial flora, it is estimated that only 15% of the approximately 300,000-500,000 species have been phytochemically investigated, while only 6% have been systematically investigated on a pharmacologically standpoint.[1b,15] The growing interest in NP-based drug discovery can be demonstrated by analyzing the PubMed publications trend in this area in the period 1982-2012 (**Figure 1**).[11]

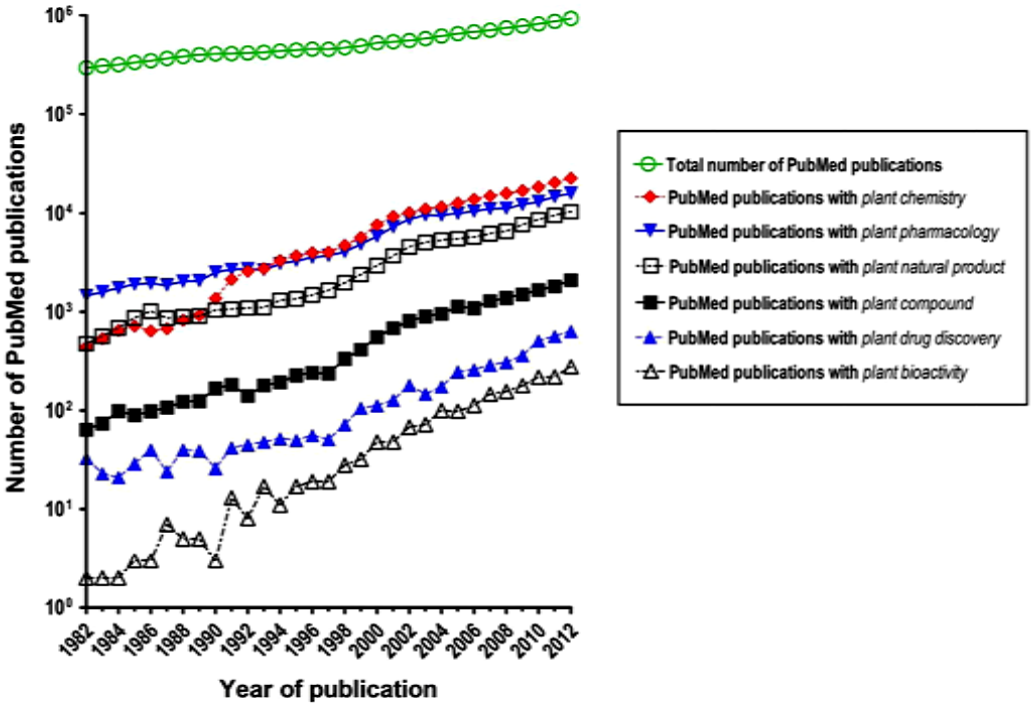


Figure 1 – PubMed publication trend analysis in plant-derived natural product pharmacology, chemistry, and drug discovery in the period 1982-2012.[11]

In another interesting study, Koehn and co-authors examined the worldwide patent trends between 1984 and 2003 in NP discovery (**Figure 2**).[3a] Interestingly, although the restricted industrial funding for NP drug discovery in the same period, the percentage of NPs-based patents has remained basically constant, with a slight increase in the patent activity through the 1980s, a slight decline from 1990 to 1999 and a novel increase between 2000 and 2003 (**Figure 2**).

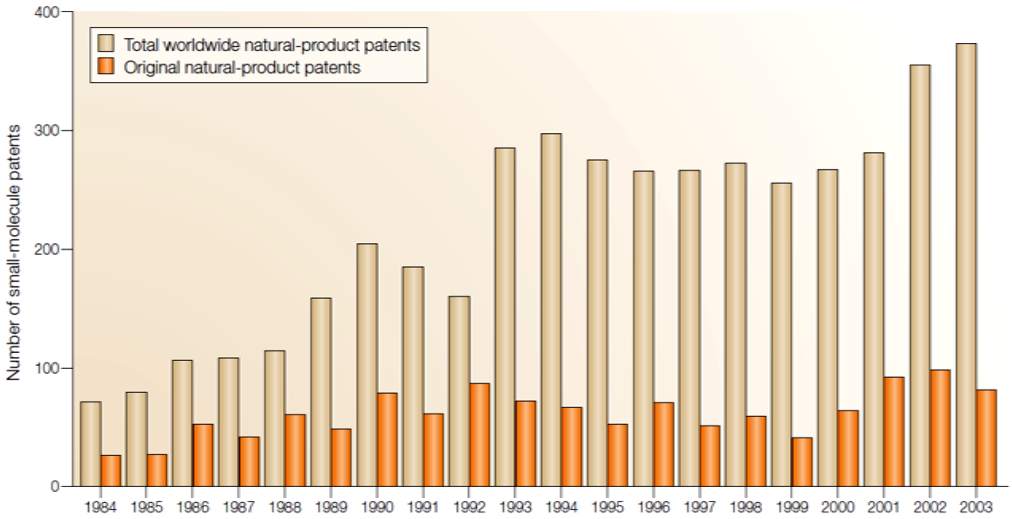


Figure 2 – Worldwide natural products-based patent between 1984 and 2003.[3a]

Cragg and Newman have extensively reviewed the state of the art of NP-derived therapeutic agents over the years.[1b,14d,16] Of the 1073 small molecules new chemical entities (NCEs) entered into the market over the period 1981-2010, 34% are NP inspired or derived (6% unmodified NPs and 28% modified NPs) (**Figure 3**). However, of the 66% of NCEs that are considered "formally synthetic", only 36% of them can be classified as truly synthetic since 16% correspond to synthetic molecules containing pharmacophores derived directly from NPs and 14% are actually modeled on a NP inhibitor of the molecular target of interest or mimic the endogenous substrate of the active site (**Figure 3**).[1b] Considering two of the main disease categories, 69% of anti-infectives and 75% of anti-tumoral agents are naturally derived or inspired.

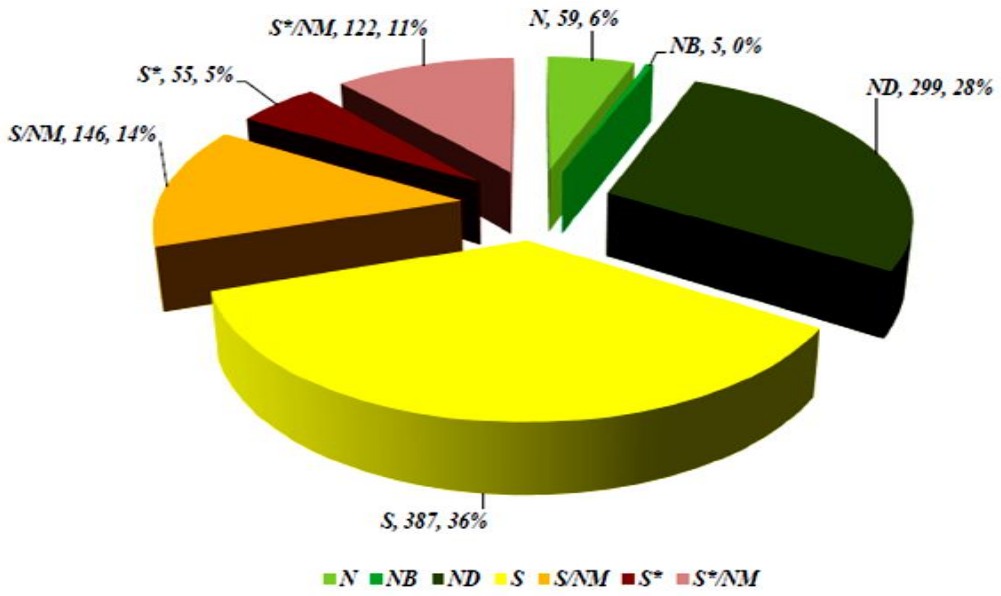


Figure 3 – Sources for the new chemical entities launched on the marked over the period 01/1981-12/2010 (n= 1073). **N**: unmodified natural products; **NB**: botanical natural products; **ND**: modified natural products; **S**: synthetic compounds with no natural product conception; **S***, **S*/NM**: synthetic compounds with a natural product pharmacophore with NM indicating competitive inhibition; **S/NM**: synthetic compounds showing competitive inhibition of the natural product substrate.[1b]

During the last century, the relative number of NP-based approved drugs varied from 33% in the 1950s to more than 50% in 1930s and 1970s, while in 1970s the relative and then absolute number of NP-based NCEs approved began to decline. Finally, in the last three decades (1984-2014), there has been a revitalized interest in NP-based drug discovery as witnessed by the average of NP-based NCEs approved per year in therapy (7.7, with a maximum of 10.3 in 1990) (**Table 1**) and investigated in clinical trials in 2014 (**Table 2**).[11,17]

Table 1 – NP-based drugs approved for therapeutic use in the last thirty years (1984-2014).[11]

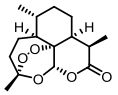
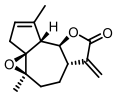
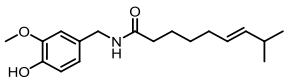
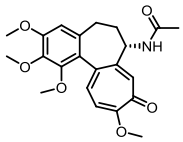
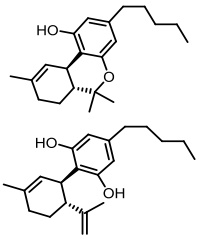
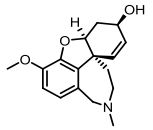
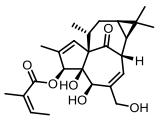
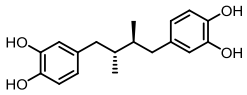
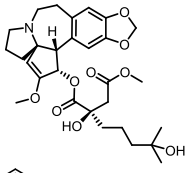
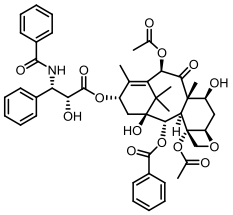
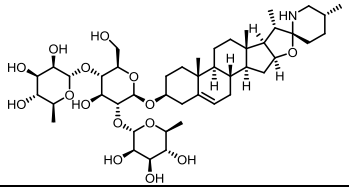
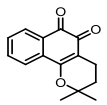
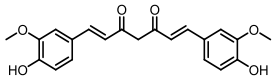
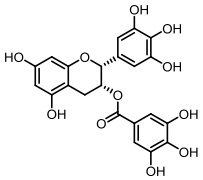
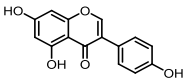
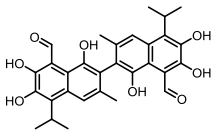
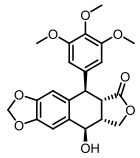
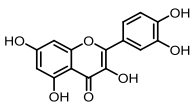
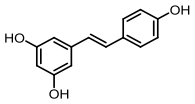
Structure	Generic name (Trade name)	Plant species	Indication (mechanism of action)
	Artemisinin (Artemisin)	*Artemisia annua* L.	Malaria treatment
	Arglabin (Arglabin)	*Artemisia globella* Kar. et Kir., *Artemisia obtusiloba* var. *glabra* Ledeb.	Cancer chemotherapy
	Capsaicin (Qutenza)	*Capsicum annum* L., or *C. minimum* Mill.	Postherpetic neuralgia
	Colchicine (Colcrys)	*Colchicum* spp.	Gout
	Dronabinol+Cannabidiol (Sativex)	*Cannabis sativa* L.	Chronic neuropathic pain
	Galanthamine (Razadyne)	*Galanthus caucasicus* (Baker) Grossh.	Dementhia associated with Alzheimer's disease
	Ingenol mebutate (Picato)	*Euphorbia peplus* L.	Actinic keratosis
	Masoprocol (Actinex)	*Larrea tridentata* (Sessè & Moc. ex DC.) Coville	Cancer chemotherapy
	Omacetaxine mepesuccinate (Synribo)	*Cephalotaxus harringtonia* (Knight ex Forbes) K. Koch	Oncology
	Paclitaxel (Taxol)	*Taxus brevifolia* Nutt.	Cancer chemotherapy
	Solamargine (Curaderm)	*Solanum* spp.	Cancer chemotherapy

Table 2 – NP-new chemical entities currently in clinical trials in 2014.[11]

Structure	Generic name	Plant species	Indications
	β-Lapachone	*Haplophragma adenophyllum* (Wall. ex G. Don) Dop	Solid tumors
	Curcumin	*Curcuma longa* L. (Turmeric)	Cognitive impairment, cancer, schizophrenia, inflammation, depression, cardiovascular disease
	Epigallocatechin-3-*O*-gallate	*Camellia sinensis* (L.) Kuntze (*Green tea*)	Epstein-Barr virus reactivation in remission patients, Alzheimer disease, cystic fibrosis, obesity
	Genistein	*Genista tinctoria* L.	Colon cancer, rectal cancer, adenocarcinoma, osteoporosis
	Gossypol	*Gossypium hirsutum* L.	B-cell chronic lymphocytic leukemia, refractory chronic lymphocytic leukemia
	Picropodophyllotoxin	*Podophyllum hexandrum* Royle	Glioblastoma, anaplastic astrocytoma
	Quercetin	*Allium cepa* L.	Chronic obstructive pulmonary disease, diabetes mellitus, heart disease
	Resveratrol	*Vitis vinifera* L.	Diabetes, lipid metabolism disorders, cardiovascular disease

As evident from these data, several blockbuster drugs are based on naturally derived or inspired products as the anti-tumoral agents taxol (Paclitaxel®), isolated from *Taxus brevifolia*, trabectedin (Yondelis®), isolated from the sea squirt, and mevastatin produced by *Penicillium citrinum* that led to the synthetic statin Rosuvastatin (Crestor®), one of the top 15 worldwide best-selling drugs in 2015,[18] just to name a few. Based on these data, it would be unbelievable to think about the patient life without statins, angiotensin antagonists and angiotensin-converting-enzyme inhibitors, most of the anticancer and antibacterial drugs, immunosuppressive agents and other naturally derived or inspired top selling drugs. However, beyond single-molecule NP-based drugs, NP extracts, also defined as multicomponent botanical therapeutics (MCBTs)[3b] or phyto-pharmaceuticals,[3c] also play a key role in modern therapy as

dietary supplements, food additives, cosmetics, medical devices and medicines.[19,20,21] In this regard, the drug discovery journey for phyto-pharmaceuticals generally starts with the knowledge concerning the use of medicinal plants in the traditional medicine and with the subsequent ethno-medicinal and ethno-botanical studies of traditional medicines that may lead to the identification and isolation of their bioactive components. After that, clinical trials on NP extracts are powerful tools for finding new targets by applying a target-based drug discovery approach known as reverse pharmacology (**Figure 4**).[22]

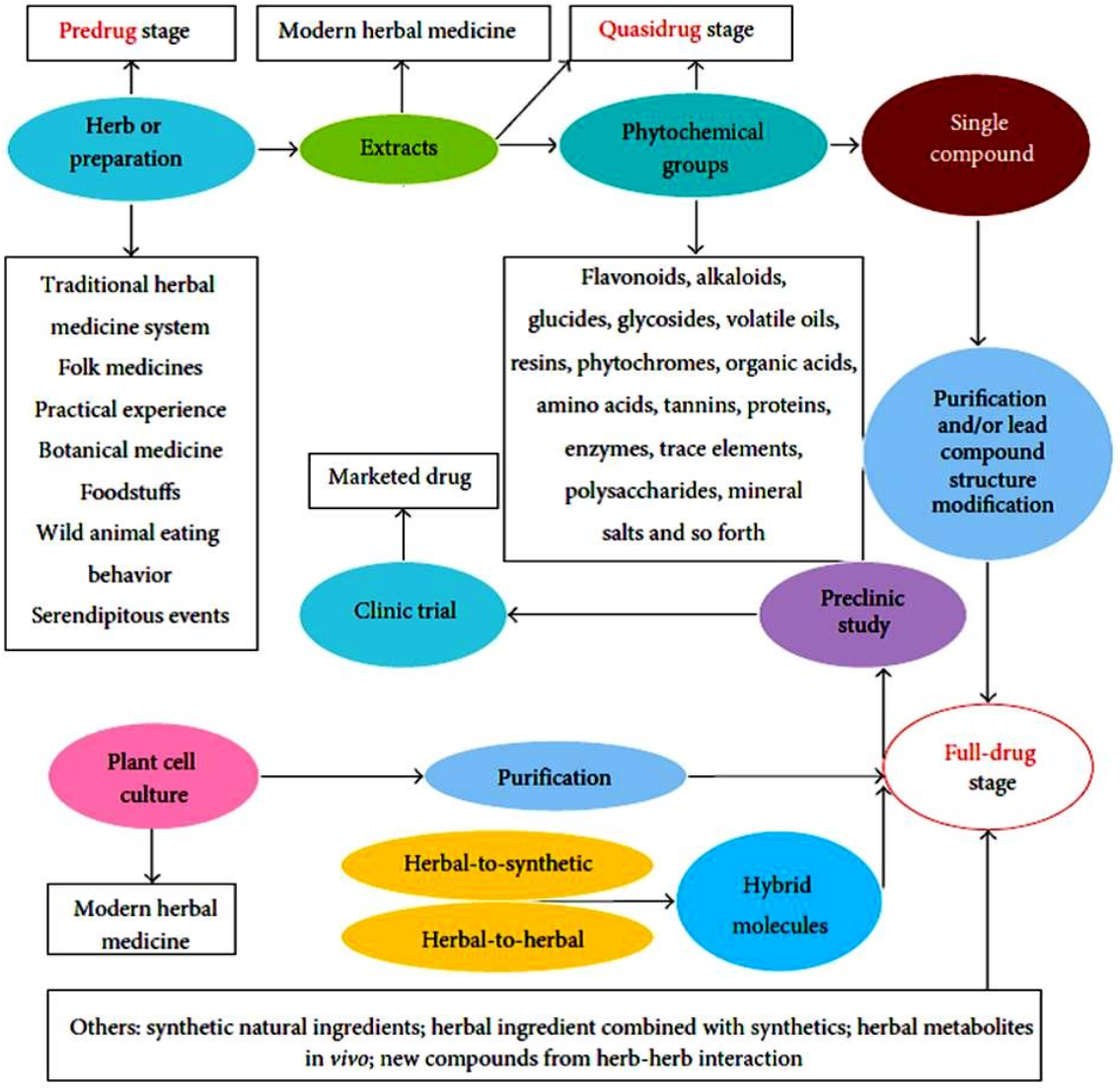


Figure 4 – Drug discovery journey for phyto-pharmaceuticals.[22]

From a pharmacological point of view, due to the concomitant presence of several bioactive components, the clinical efficacy of a typical plant extract is not always explainable by considering the effect of the single active NP since synergistic effects are likely to occur.

Synergistic effects may include multiple aspects including multi-targets effect, mutual interaction between two components of the same extract able to improve the pharmacokinetic profile of one of the two components, enhanced efficacy of one component due to the inhibition of its metabolism by another component.[23] Although synergistic effects are difficult to prove, the experimental demonstration of a given synergistic effect can be achieved by comparing the pharmacological effects of the mono-substances *versus* the combination of substances by analyzing isobole curves based on data from several dose combinations.[23] These kind of studies allow to discriminate between simple additive effects, antagonistic interactions or real synergism with potentiated or additive effects.[24] Poly-pharmacology, systems biology and synergistic effects may explain why whole plants extracts are often considered therapeutically superior over the single isolated constituents.[25]

Herbal phyto-pharmaceuticals have reached 60 USD billion with an annual growth rates of 5-15% due to their increased usage in therapy, especially in psychosomatic, metabolic and minor disorders.[26] This remarkable sudden success is mainly related to the approval of several phyto-pharmaceutical drugs worldwide in the last ten years. For example, in 2002 the European Medicines Agency (EMA) approved and granted as orphan drug the medicinal specialty known as NexoBrid®, a concentrate of proteolytic enzymes from the pineapple plant for the removal of damaged tissue caused by severe burn wounds.[3c] Moreover, FDA approved Veregen®, a mixture of Tea catechins for the treatment of genital diseases caused by human papilloma virus,[20b] and Fulyzaq®, an extract from the red sap of the *Croton lechleri*, for the treatment of HIV diarrhea.[27] Sativex® is a well-known titrated extract containing an association between the psychoactive δ-9-tetrahydrocannabinol and the anti-inflammatory cannabidiol. Sativex® has been approved since 2005 in different countries as oro-mucosal spray for the treatment of spasticity due to multiple sclerosis and neuropathic breads.[3c] More recently, in 2012, the dry extract of *Dioscorea nipponica* was approved in Germany for the treatment of headache, muscle pain and cramps, representing the first traditional Chinese medicine approved in EU.[3c]

1.2 Legal requirements and regulatory guidelines for the quality control of herbal medicinal products and medical devices for human use in the European Union

Although the last decade has witnessed a marked growth in the market of plant-based natural products, their high complexity in terms of composition still makes a challenging task the guarantee of quality, efficacy and safety requirements.[28] Among different herbal products it is possible to distinguish: medicines or drugs, medical devices, food/dietary supplements, novel foods, functional foods and cosmetics (**Table 3**).[29,30]

Table 3 – Definitions of drug, medical device, food/dietary supplement, novel food, functional food and cosmetic.[29,30]

Type of product	Definition
Medicine or drug	″(*a*) Any substance or combination of substances presented as having properties for treating or preventing disease in human beings; (*b*) Any substance or combination of substances which may be used in or administered to human beings either with a view to restoring, correcting or modifying physiological functions by exerting a pharmacological, immunological or metabolic action, or to making a medical diagnosis.″
Medical device	″Any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of: diagnosis, prevention, monitoring, treatment, or alleviation of disease; diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap; investigation, replacement or modification of the anatomy or of a physiological process; control of conception and which does not achieve its principal intended action in or on the human body by pharmacological, immunological, or metabolic means, but which may be assisted in its function by such means.″
Food/dietary supplement	″A product intended for ingestion that contains a "dietary ingredient" intended to add further nutritional value to (supplement) the diet. A "dietary ingredient" may be one, or any combination, of the following substances: a vitamin; a mineral; an herb or other botanical; an amino acid; a dietary substance for use by people to supplement the diet by increasing the total dietary intake; a concentrate, metabolite, constituent, or extract.″
Novel food	″Food that has not been consumed to a significant degree by humans in the EU prior to 1997, when the first Regulation on novel food came into force. 'Novel Food' can be newly developed, innovative food or food produced using new technologies and production processes as well as food traditionally eaten outside of the EU.″
Functional food	″A food given an additional function (often one related to health-promotion or disease prevention) by adding new ingredients or more of existing ingredients.″
Cosmetic	″Intended to be applied to the human body for cleansing, beautifying, promoting attractiveness, or altering the appearance without affecting the body's structure or functions.″

Recently, within the EU, appropriate regulatory actions have been undertaken to regulate and harmonize the legal *status* of these various groups of plant preparations throughout the different European countries.

In case of herbal medicines, the herbal substance (i.e. the starting plant material) and/or the herbal preparation (i.e. the plant extract, tincture, exudates, fatty oil, essential oil) represents the active substance. Therefore, the quality of the herbal active substance must be assured in such a way that reliable and reproducible therapeutic success is guaranteed from batch to batch.[29] In contrast to active pharmaceutical ingredients (APIs) derived from traditional chemical synthesis, starting materials from plants are subjected to a considerable variability involving not only the content of the active fractions but also the concomitant presence of additional components and contaminants of different origin. In this regard, genetic factors and growing conditions such as climate, soil quality and pests may affect the overall spectrum of constituents and therefore the quality of the plant material. Furthermore, harvesting and drying conditions, pre- and post-harvest treatments with pesticides and fumigants, as well as comminuting and storage conditions are further factors that may affect the quality of the herbal substance. The characterization of herbal extracts involves the detailed investigation of all the botanical and phytochemical aspects of the plant, macroscopic and microscopic assays and purity tests performed according to the requirements of the European Pharmacopoeia (EP). Besides the conventional purity tests, herbal substances have to meet additional requirements due to their origin including tests for heavy metals, residues of pesticides, fumigants, residual solvents, mycotoxins and microbial contaminants.[29] On the other hand, phytochemical identification of herbal substances and preparations consists of chromatographic fingerprinting and quali-quantitative determination of selected constituents which represent the main or the most active components.

From a legislative point of view, the Directive 2004/24/EC represented the first attempt made by the European Parliament to regulate and harmonize the legal *status* of plant preparations throughout the different European countries.[31] The Directive establishes that herbal medicines released in the market need authorization by the national regulatory authorities of each European country and that these products must meet the acceptable level of safety and efficacy. The safety of herbal medicinal products is evaluated on the basis of existing scientific literature (data from clinical studies, case reports, pre-clinical studies). If data on safety are not sufficient, it will be communicated to consumers.[32] According to these

criteria, two categories of herbal medicinal products have been established based on their efficacy/safety profile: well-established use herbal medicinal products (medicinal herbs with a recognized level of safety and efficacy) and traditional use herbal medicinal products. The last category include those medicinal herbs that do not have a recognized level of efficacy but are acceptably safe.[33] Currently, after more than ten years since the introduction of the first European legislation on herbal medicines, over 1,300 plant-based products have been registered and more than 600 of them have been authorized for marketing on the basis of consolidated use in the EU Member States. EMA reports that today the sale of these medicines is accompanied by clear and harmonized information in all countries.[34] Since its establishment, the Committee on Herbal Medicinal Products (HMPC) has issued 128 scientific opinions including monographs containing information on therapeutic use, contraindications, interactions with other drugs and potential side effects, based on the revision of all scientific data and historical information, with data on quality, efficacy and safety.[34] These scientific opinions are used by the Member States of the Union to evaluate herbal medicine. The HMPC has also developed twenty guidelines to harmonize the registration and authorization processes of herbal medicines. Given their complexity, borderline cases such as plant products that fall into different product categories, are object of discussion in this context.

Within the phyto-pharmaceuticals, the specific sub-category of medical devices requires a specific discussion. As a definition, ″medical devices include any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of:

- diagnosis, prevention, monitoring, treatment, or alleviation of disease;
- diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap;
- investigation, replacement or modification of the anatomy or of a physiological process;
- control of conception;

and which does not achieve its principal intended action in or on the human body by pharmacological, immunological, or metabolic means, but which may be assisted in its function by such means (Directive 2007/47/EC)″.[35] Despite the requirement to transpose the currently

directives (90/385/EEC, 93/42/EEC, 98/79/EC) into law in each Member State, the binding laws, decrees and ordinances of Member States led to differences in levels of requirements and in some cases to different approaches between Member States. As a result a variety of interpretations could be observed across the European Union, including different approval processes for clinical investigations, safety reporting during clinical investigations and not unified criteria for the classification of products.[36] To solve these issues, two new Regulations on medical devices have been just approved on April 2017,[37] thus replacing the existing Directives and creating a breaking point in the regulation of these products:

- Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.
- Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU.

The new rules will be apply after a transitional period of 3 and 5 years after entry into force on 25 May 2017, respectively for the Regulation on medical devices and on *in vitro* diagnostic medical devices.[38] The aim of the new regulations is to ensure a better protection of public health and patient safety. Among the Regulation (EU) 2017/745, the approved text clearly defines the regulatory framework of “medical devices made of substances” and acknowledges that this product category is of utmost importance for the self-medication market. There are many aspects of the new regulation that will have a significant impact on the industrial system. These include the classification rule 21 and defining the concept of “absorption”, and clinical product evaluation. Medical devices are basically divided into four classes, ranging from low risk to high risk: I, IIa, IIb, III. According to the new rule 21, “devices that are composed of substances or of combinations of substances that are intended to be introduced into the human body *via* a body orifice or applied to the skin and that are absorbed by or locally dispersed in the human body are classified as: class III if they, or their products of metabolism, are systemically absorbed by the human body in order to achieve the intended purpose; class III if they achieve their intended purpose in the stomach or lower gastrointestinal tract and they, or their products of metabolism, are systemically absorbed by the human body; class IIa if they

are applied to the skin or if they are applied in the nasal or oral cavity as far as the pharynx, and achieve their intended purpose on those cavities; class IIb in all other cases" (Annex VIII, Chapter 3). When absorption is relevant for the intended purpose the classification is clear. However, all substance based medical devices, are somehow absorbed by the human body. In the absence of a clear criterion, a product containing substances generally recognized as safe and other active substances absorbed in concentrations not relevant for risk assessment would be classified in class III without any justification. Among others requirements of the new regulation there is the need to report on the label of devices that "are composed of substances or combinations of substances, that are intended to be introduced into the human body *via* a body orifice or applied to the skin and that are absorbed by or locally dispersed in the human body, the overall qualitative composition of the device and quantitative information on the main constituent or constituents responsible for achieving the principal intended action" (Annex I, Chapter 3). In this view, the research into new analytical methods for the study of the composition of products containing natural complex substances and, if necessary, the application of ADME studies for the evaluation of plasma concentrations are required and an important role will be played by the combination of new emerging disciplines.

1.3 Metabolomics: a valuable tool to analyze natural complex matrices

As discussed in the first paragraph, due to the concomitant presence of a multitude of different components, medicinal plants exert health effects as their whole rather than by virtue of a single component with the consequent occurrence of multi-target effects, mutual interaction among the components and synergistic phenomena. In this scenario, metabolomics is recently emerged as a valuable "omics" technology applied for the comprehensive profiling of the metabolome and aimed at identifying, quantifying and gathering detailed information about the complete set of metabolites present in a cell or organism (**Figure 5**).[39]

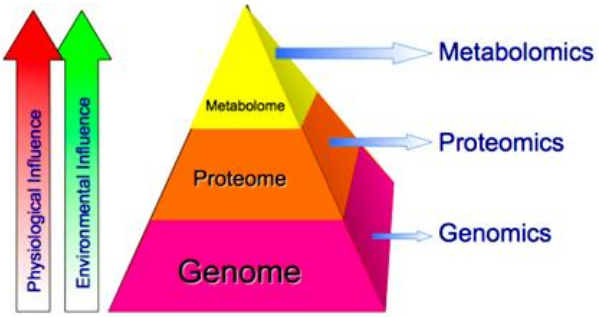


Figure 5 – The "pyramid of life" showing the varying influence of the environment and physiology on the genome, proteome and metabolome (extracted from lecture by David Wishart from University of Alberta - 2016 Informatics and Statistics for Metabolomics workshop)

Since the metabolome refers to the complete set of small-molecules found within a biological sample as cells, cellular organelles, organs, tissues, biofluids or an entire organism,[40] metabolomics investigations on medicinal plants require the simultaneous analysis of a huge spectrum of compounds with very diverse physico-chemical properties. Accordingly, powerful analytical tools, able to cover a portion as widely as possible of the chemical space of plant metabolites, are required for the quali-quantitative determination of this vast array of chemicals. Currently, two main complementary approaches can be applied for metabolomics investigations, namely metabolic profiling (or targeted profiling) and metabolic fingerprinting (or non-targeted fingerprinting).[41] Metabolic profiling focuses on the quantitative analysis of a group of metabolites related to a specific metabolic pathway or a class of compounds. While the metabolic profiling involves target analysis based on the direct measurements of selected analytes, the metabolic fingerprinting aims at comparing, rather than identifying, specific patterns or fingerprints of metabolites that change in response to disease, environmental or genetic perturbations.[41] However, due to the high complexity and chemical diversity of the

metabolome, the simultaneous and comprehensive qualitative and quantitative analysis of all metabolites within a cell, tissue or organism still represents a fascinating challenge to face.[42]

In this regard, a single analytical approach is generally not enough to provide a comprehensive overview of the metabolome but the combination of multiple and hyphenated techniques would be highly desirable.[43] Accordingly, many analytical technologies have been employed to profile complex plant-derived matrices and the choice of the most suitable tool represents a compromise among speed, selectivity and sensitivity.[44] Mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy are the most commonly adopted analytical techniques to pursue this goal. NMR is one of the most efficient techniques for structure elucidation based on the measurement of specific molecular properties that can be associated with the molecular structure of known or unknown chemical entities.[45] NMR is considered as non-biased technique that theoretically provides a unique signal for each chemically distinct nucleus responsive to the magnetic field. Despite its non-destructive feature and the possibility to easily set-up metabolic profiling platforms in high-throughput fashion, NMR-based approaches are characterized by a low sensitivity.[46] Therefore, MS is by far the technology of choice for plant metabolomics studies,[47] especially when coupled to powerful chromatographic techniques such as liquid chromatography (LC),[48] gas chromatography (GC)[49] and capillary electrophoresis (CE).[50] In particular, this paragraph will be mainly focused on the application of MS-based targeted and untargeted approaches for plant metabolomics studies through a typical workflow which includes the design of analytical experiments, followed by sample preparation, data acquisition, data processing and data analysis (**Figure 6**).[51]

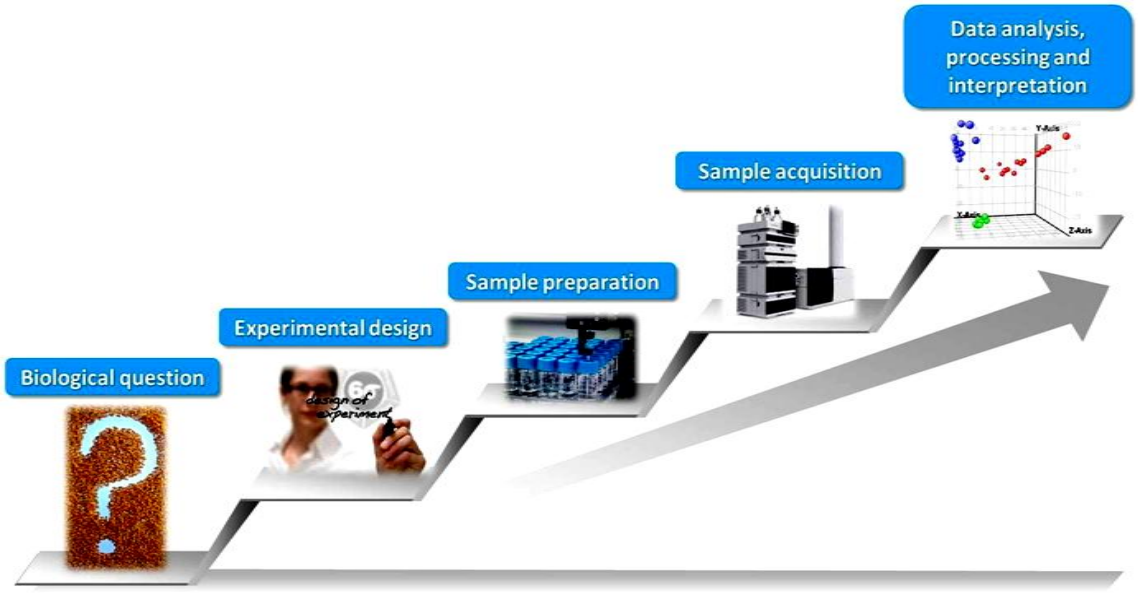


Figure 6 – Typical MS-based metabolomics workflow.

Experimental design

To ensure scientific soundness and robust conclusions to the metabolomic investigation, a rational design of the analytical experiment is required in order to prevent, or at least minimize, the occurrence of misleading factors of analytical origin that can garble the investigated variables. In other words, experimental design means providing the appropriate foresight to plan a study in order to guarantee that the variation observed in the study and related to the biological observations are significantly greater than process variability artificially introduced in the study. Thus, many experimental considerations are required, especially for the design of a large-scale study.[52] These include, first of all, the reproducibility of sample collection, process and storage since differences in how samples are collected, processed and stored can strongly affect the observed metabolic profile. The design of multiple analytical experiments is often required, therefore large biological studies are generally divided into smaller analytical experiments to achieve an appropriate analysis of samples by controlling different factors. Among the most crucial factors to control, the number of injections within a single analytical session and the build-up of matrix and metabolites on LC and GC columns may cause variability in retention time data.[52] Furthermore, the mass spectrometric performance represents an additional crucial factor which can lead to a different response of the analytes and *m/z* variability due to the aggregation of matrix components on the MS source or injector, thus affecting the ion transmission from atmospheric to *vacuum* regions of the mass spectrometer. Another expedient is based on the randomization during sample preparation and data acquisition to ensure that no bias is introduced. Moreover, between two analytical sessions it is important to return the instrument to the optimal performance by cleaning the LC column and the MS source to remove contaminants and tuning the mass spectrometer. Due to the numerous sources of analytical variability in MS-based analytical methods, a pragmatic approach for monitoring data quality is based on the use of quality control (QC) sample, prepared from the sample under analysis and analyzed at regular intervals throughout the analysis.[52] Being representative of the qualitative and quantitative composition of the samples under investigation, QC samples can be applied to chromatographic conditions and MS instruments following maintenance, in order to correct small levels of variation, to quantitatively measure technical reproducibility and to integrate data from different analytical experiments. Internal standards spiked into samples is another method that can be applied to reduce sample variability, in particular when an isotopic analogue of the metabolite under

investigation is used. However, the use of an internal standard for each metabolite to be detected in a complex plant extract is difficult to apply because of the large number of standards required, the cost of purchasing isotopic analogues for each metabolite and the limited qualitative knowledge related to metabolome composition before sample preparation and data acquisition. In this regard, the use of a single internal standard can be applied to correct analytical variation for a group of metabolites that are either chemically related or unrelated.[52]

Sample preparation

Sampling and sample preparation are fundamental steps to prevent any biochemical process able to degrade the target metabolites or to generate new artifacts, thus avoiding any change in the metabolites concentration and identity.[53] The main methods used for quenching the enzymatic activity in plant matrices are represented by freeze clamping, immediate freezing in liquid nitrogen and acidic treatments. Concerning the sample preparation, the ideal protocol must allow the analytes extraction from complex biological matrices, making them compatible with the analytical technique while removing at the same time those components able to interfere with the analysis. However, the adopted protocol for the sample preparation can be different depending on the type of metabolomic analysis. For example, in target metabolic analysis, since the metabolites are known, the sample preparation can be specifically addressed for the target analytes, while the ideal sample preparation protocol for untargeted approaches should be as aspecific and universal as possible with minimum analytes losses.[53,54] The main techniques for sample preparation include:

- *Solid-phase extraction (SPE)*: This technique, widely used in targeted and profiling analysis to eliminate the interfering matrix, aims to extract analytes from biological fluids. SPE is based on the temporary retention of the target metabolites from a flowing liquid sample on a solid sorbent. The retained analytes are then eluted from the sorbent by using a solvent of suitable elution strength.[55]

- *Liquid-liquid extraction (LLE)*: Mainly used for tissue extraction by shaking or vortexing the matrix with organic solvents after the fine grinding of the frozen sample.[56] The solvents

used are isopropanol, ethanol, methanol, acetonitrile, water, water/methanol for polar compounds, while more lipophilic metabolites can be extracted with chloroform or ethyl acetate. The simultaneous use of polar and apolar solvents allows the fractionated sorting of metabolites into polar aqueous and lipophilic organic fractions which can be separately analyzed.[57] Recent variants of this technique include ultrasound-assisted extraction (UAE), microwave-assisted extraction (MAE), extraction with ionic liquids, supercritical fluid extraction (SFE) and accelerated or pressurized solvent extraction (ASE).[58]

- *Direct injection*: This approach involves the simple dilution and the direct injection of the sample without any prior manipulation.[59] Although is a fast analytical approach in which complex mixtures are often resolved into components differing in ion mass, avoiding time-consuming chromatographic separation, direct infusion mass spectrometry (DIMS) suffers from drawbacks such as ion suppression and adduct formation phenomena due to the high-salt content and the contamination of the non-volatile residues with rapid reduction of the instrument performance.[60] The resulting ion intensity profiles are typically analyzed by multivariate statistics to identify key ions discriminating classes of samples.[59]

Sample analysis

Modern MS offers a vast set of technologies that differ in operational principles and performance, including different ionization techniques, mass analyzers, resolving power and mass accuracy.[61] The most common ionization techniques in metabolomics are electron ionization (EI), electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI).[61] ESI and APCI are the most common employed in LC-MS-based metabolomics: the former is well suited for the ionization of a wide range of metabolites including drug compounds, aminoacids, organic acids, sugars, steroids, phospholipids and fatty acids; the latter is used especially for the ionization of non-polar compounds such as phospholipids, fatty acids, steroids and essential oils. Mass analyzers with different resolving power are also employed in metabolomics. These include quadrupole (Q) and triple quadrupole (QQQ), the time of flight (TOF), quadrupole-time of flight (Q-TOF) and ion trap (IT) (**Table 4** and **Figure 7**). Each of these has different advantages and disadvantages and range of applicability that depend on the specific metabolomics area.[51]

Table 4 – Overview of the most commonly used mass spectrometers for metabolomics applications.[51]

Mass analyzer type	Resolving power (FWHM)	Mass accuracy (ppm)	Scan rate	Mass range	Dynamic range
Low resolution instruments					
QqQ	Up to 7,500 (at *m/z* 508)	5-500 ppm	Up to ~ 5,000Da/s	Up to 3,000 *m/z*	10^5-10^6
IT/LIT	Up to 10,000	50-500 ppm	Up to ~ 33,000Da/s	Up to 4,000 *m/z*	10^4
Qq-LIT	Up to 9,200 (at *m/z* 922)	50-500 ppm	Up to ~ 20,000Da/s	Up to 2,000 *m/z*	10^5-10^6
High resolution instruments					
TOF	Up to ~20,000 (at ~*m/z* 1,000)	< 1-2 ppm[a]	Up to 40 Hz	Up to 20,000 *m/z*	10^4-10^5
Qq-TOF	Up to 60,000 (at *m/z* 1,222)	< 1-2 ppm[a]	Up to 100 Hz	Up to 40,000 *m/z*	10^4-10^5
IT-TOF	10,000 (at *m/z* 1,000)	< 2 ppm[a]	10 Hz	Up to 5,000 *m/z*	10^3
Orbitrap-Exactive	Up to 140,000 (at *m/z* 200)	< 1 ppm[a] < 3 ppm[b]	Up to 12 Hz with a mass resolution of 17,500 (at *m/z* 200)	Up to 6,000 *m/z*	10^3-10^4
LTQ-Orbitrap	Up to 240,000 (at *m/z* 400)	< 1 ppm[a] < 3 ppm[b]	Up to ~8 Hz with a mass resolution of 15,000 (at *m/z* 400)	Up to 4,000 *m/z*	10^3-10^4
Q-Orbitrap	Up to 140,000 (at *m/z* 200)	< 1 ppm[a] < 5 ppm[b]	Up to 12 Hz with a mass resolution of 17,500 (at *m/z* 200)	Up to 4,000 *m/z*	10^3-10^4
LTQ-FT 7T	>750,000 (at *m/z* 400)	< 1 ppm[a] < 1.2 ppm[b]	1 Hz with a mass resolution of 100,000 (at *m/z* 400)	Up to 4,000 *m/z*	10^3-10^4
Qq-FT 7T	> 1,000,000 (at *m/z* 400)	< 1 ppm[a] < 1.5 ppm[b]	1 Hz with a mass resolution of 250,000 (at *m/z* 400)	Up to 10,000 *m/z*	10^3-10^4

[a] Internal calibration. [b] External calibration.

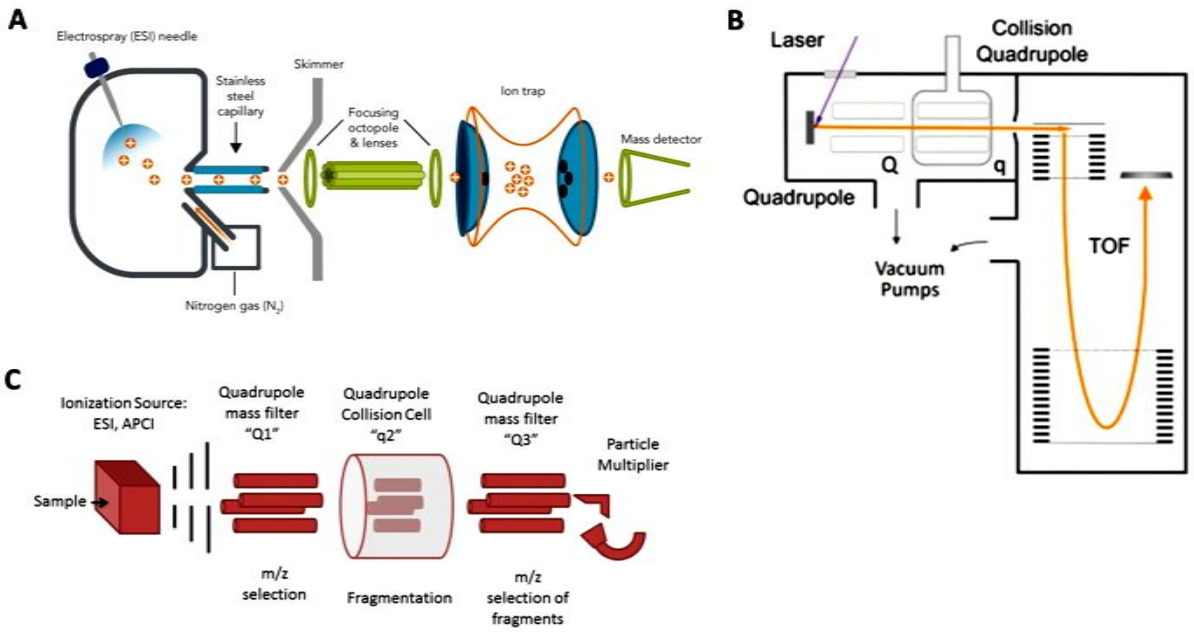


Figure 7 – Schematic representation of some mass spectrometry analyzers: ion trap (**A**), Q-TOF (**B**) and triple quadrupole (**C**).

For example, quadrupole-based analyzers are particularly appropriate for quantitative analysis, while TOF analyzers have greater scan speed, dynamic range and sensitivity, allowing the quantitative profiling over a wider range of metabolite concentrations and the detection of lower concentrations. For the analysis of extremely low concentrated metabolites, an IT-based instrument would be the most suited, allowing the trapping of specific ions. MS/MS or MS^n-analyzers are required for compounds identification. Q-TOF, QQQ and conventional IT have been commonly employed to perform MS/MS within plant metabolomics for the identification of a broad range of phytochemical species, although conventional MS^n-analyzers lack mass accuracy if compared with the more recent technologies such as Fourier transform ion cyclotron resonance MS (FT-ICR-MS) and linear trap quadrupole orbitrap MS (LTQ-Orbitrap MS). Recently, Q-TOF technology has prompted a great deal of interest thanks to the development of higher sensitivity analyzers able for accurate quantitative profiling together with the possibility to filter specific ranges of mass ions within the quadrupole and collision-induced decomposition (CID) for structural analysis *via* fragmentation. Modern Q-TOF/MS instruments also allow the simultaneous collection of two functions of data within a single run: one is performed at low energy in order to extrapolate the exact precursor ion mass spectra while the other is acquired at higher energy to provide the exact mass of the fragment ions. Thus, for each chromatographic peak, both MS quantitative profiling data and MS/MS

fragmentation data for compounds identification can be acquired.[62] Among the different MS-based metabolomics techniques, we can distinguish direct MS analysis and MS coupled to chromatography.[61] MS coupled with chromatographic techniques represents an excellent solution for complex mixture analyses with several advantages including the reduction of matrix effect and ion suppression phenomena, isomers separation, more accurate quantification of single metabolites.[61] Currently, two main chromatographic techniques have been coupled with MS for plant metabolomics investigations:

- *Gas chromatography coupled to MS (GC-MS)*: This technique is ideally suited for the analysis of volatile and thermally stable compounds. These include ketones, aldehydes, alcohols, esters, furan, pyrrole derivatives and other heterocyclic compounds. If the analytes under analysis are not sufficiently volatile, a derivatization step is previously required. GC-MS is used in metabolome analysis thanks to its high separation efficiency due to the use of modern capillary columns that enable the separation of more than 100 compounds in a single analysis by an accurate selection of the capillary column, the stationary phase and the film thickness. In addition, the standardized MS electron ionization energy of 70 eV leads to reproducible mass spectra and highly transferable electron ionization MS spectral libraries that allows compounds identification through mass spectral library matching.[61] GC is often coupled to single quadrupole MS detector which has the advantages of high sensitivity and good dynamic range although suffers from slower scan rates and lower mass accuracy. GC-triple quadrupole MS/MS is capable of multiple reaction monitoring of analytes, which can overcome the identification and quantification problems associated with co-eluting compounds in complex matrices. GC-TOF-MS technology offers high mass resolution, high mass accuracy and fast scan speeds.[61]

- *Liquid chromatography coupled to MS (LC-MS)*: This technique is generally used for molecular identification and quantification of polar, less-polar and neutral metabolites. It is particularly valuable for target analysis thanks to its resolving power, sensitivity and specificity, allowing the identification of target metabolites in very complex matrices. Reversed-phase (RP) columns such as C18 and C8 are the most utilized. LC-MS^n represents the most important tool for the identification of metabolites in crude plant extracts, providing not only molecular mass information but also structure information deduced

from fragmentation patterns deriving from the collision-induced dissociation. LC-MS represents also an important technique for metabolic fingerprinting thanks to the possibility to detect known and unknown compounds which can be identify without any prior knowledge of their exact chemical structure.[63] In particular, ESI, APCI and fast atom bombardment (FAB) have been used as ionization techniques for the MS analysis of plant metabolites in combination with RP-LC.[63]

Data analysis

Metabolomics raw data are a source of information and, in turn, a great source of knowledge after suitable data processing. In this regard, the ability to mine big data and to undertake reliable comparative analysis governs the success of metabolomic-based experiments.[46,64] Depending on the approach used, metabolomic data can be submitted to different purposes.

When the goal of the metabolomic analysis is to determine the relative differences between the metabolomes of two or more systems, data analysis is largely carried out by using chemometric tools.[65] Accordingly to Goodacre *et al.*,[65] after the acquisition of mass spectra, the data sets require a pre-processing step, which includes the reduction of background noise, adjusting for baseline shifts and machine drift, peak alignment, peak detection and mass spectra deconvolution (the separation of overlapping signals into individual chemical peaks).[66] Statistical analysis is then applied to the data set to extract biological relevant information by using an appropriate algorithm or a set of algorithms. Univariate analysis is the simplest statistical method and is carried out with only one variable at a time (ANOVA, *t*-tests, etc).

Although metabolomics experiments generate multivariate data, univariate methods can be employed to test individually metabolites that are increased or decreased significantly between different groups. Multivariate data analysis (MVDA) treats, instead, the analysis of multiple variables simultaneously and includes both unsupervised classification methods, such as principal component analysis (PCA) and hierarchical cluster analysis (HCA), and supervised approaches such as partial least square (PLS) for metabolites classification.[67] In particular, PCA provides a rapid means for the visualization and comparison of data sets and can be used to classify different sample groups and to provide information about the variance in the data, with minimal loss of information. In fact, the concept behind PCA is to describe the variance of a set of multivariate data in terms of a set of underlying orthogonal variables, namely principal

components (PCs). The original variables can be expressed as a particular linear combination of PCs, with each PC accounting for a portion of the total variance of the data set. HCA is a method of grouping samples in a data set by their similarity. On the contrary, PLS is a MVDA technique that allows sample discrimination by reduction of dimensionality and maximizing correlation between variables.

Thousands of metabolites can be detected by applying untargeted approaches, with limited prior knowledge of the sample composition.[68] In this regard, the Metabolomics Standard Initiative (MSI)[69] has published several guidelines on metabolomics experiments and analysis including several levels of identification with various levels of confidence:

1. *Identified compounds*: known compounds identified by comparing retention time and mass spectrum, or retention time and NMR spectrum, or exact mass and MS/MS, or exact mass and isotope pattern with authentic compounds under identical analytical condition.
2. *Putatively annotated compounds*: identified without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries.
3. *Putatively characterized compounds*: the identification is based on the characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class.
4. *Unknown compounds*: although unidentified or unclassified, these metabolites can still be differentiated and quantified based upon spectral data.

However, despite the relative high level of information available for a specific compound, its accurate identification may be sometimes difficult. For example stereoisomers appear very similar based on their chromatographic or mass spectrometric profile, especially in non-optimized analytical methods commonly applied for untargeted metabolomic studies: thus, in such cases, the development of chromatographic methods for the unequivocally resolution of the isomeric mixture is mandatory. The first step for compounds identification using MS platforms is based on the accurate measurement of the *m/z*. After the identification of one or more possible molecular formulas by using available compound libraries such as KEGG, PubChem or ChemSpider, the matching to known metabolites is generally performed.[68,70] Thanks to the availability of fragments from MS^2 or MS^n, the possible elemental composition

can be restricted since fragmentation lead to characteristic fragment ions for each molecular formula, helping to generate hypothesis about completely unknown compounds.[71] In LC-MS, *in silico* fragmentation software are available to enable the matching of *in silico* derived mass spectra (instead of mass spectra derived from authentic chemical standards) to the experimentally derived mass spectra. The comparison of tandem mass spectra against measurements from literature or database allows to keep the putatively annotated compounds. Once the MS^2 or MS^n mass spectra are available, they can be compared with reference spectra available in several public libraries.

The Human Metabolome Database (HMDB) contains information about 8,500 compounds and 10% of them have MS^2 data available.[72] The Metlin database contains information about 44,000 metabolites and for 10% of them high-resolution MS^2 spectra are available.[73] Finally, MassBank represents the first public mass spectral database that accept spectral data from the community. Both HMDB and Metlin do not possess a good coverage of plant-specific metabolites, while on the other side, MassBank provides a much better coverage of secondary plant metabolites.[74] The comparison of experimental data (accurate *m/z*, retention time, fragmentation mass spectrum) for each metabolite to mass spectral libraries constructed with authentic chemical standards is the ideal process to provide definitive identification. This is currently more successfully achieved for data acquired on GC-MS platforms compared to LC-MS platforms, but the growing number of spectra of a variety of instruments will help to overcome the old problem that soft ionization spectra are less reproducible than those generated by GC-EI-MS. Without metabolites identification it is impossible to base biological reasoning on the datasets.[68]

AIMS OF THE WORK

Aboca's research is focused on the investigation of novel natural biologically-active substances, transformed by physical technological processes that use water and ethanol as extraction solvents while avoiding any unnecessary additives.[75] The application of hyphenated techniques, modern tools in molecular biology and analytical platforms for the quali-quantitative and biological characterization of natural complex products is the basis of the scientific research in Aboca S.p.a., which is aimed not only to enable the use of natural products to treat symptoms of common human minor disorders but also to gain new pharmacological knowledge leading to the primary objective of restoring the body's equilibrium. Furthermore, the unrelenting progress in the "omics" disciplines will help, even more in the near future, in evaluating the correspondence between the patterns of highly complex natural-derived APIs and the physio-phatological pathways in human body. Aboca's philosophy believes that the scientific knowledge available today allows us to rationalize a more systemic vision of humankind and its interactions, even with complex substances such as those found in nature. In other words, according to the so-called holistic vision, the sum of the parts is more functional than the sum of the performance of the individual parts.

When applied to the phytochemical field, the holistic model allows to investigate natural complex systems by an inter- and multidisciplinary method and to understand how the activity of a plant complex could be different, and sometimes greater, with respect to the single isolated components.

Based on these considerations and according to this "mission", more than ten years ago the Phytochemical Research Unit was established in Aboca S.p.a. (*Sansepolcro, Italy*) with the aim to investigate formulated natural complex products by using hyphenated analytical techniques including LC-MS, GC-MS, ICP-MS, LC-DAD, LC-RID.

As a part of this research interest,[76] the following chapter of this *Philosophiae Doctor* Thesis describes the application of two complementary mass spectrometry-based approaches for the metabolomics investigation of Grintuss® syrup, a medical device based on honey and functional vegetable substances (resins, polysaccharides and flavonoids from *Grindelia robusta*, *Plantago lanceolata* and *Helychrisum italicum*) which creates a protective film with a "barrier effect" that soothes the cough and protects the upper respiratory tract.[77]

The first approach, based on an untargeted fingerprinting, was aimed to obtain a broad picture of the whole metabolome of Grintuss® syrup by detecting as many groups of metabolites as possible, without identifying nor quantifying. Thus, focussing on the whole formulation rather than on the single components, this approach allowed the generation of a sort of “identity card”, namely multivariate control chart, useful to assess the quality end-points and to compare the biological uniformity among different batches of the production supply chain by finding statistically significant biomarkers.

On the other hand, a LC-MS-based metabolomic profiling approach have been successfully developed and applied for the quali-quantitative analysis of GrinTuss® syrup by matching the detected metabolites with standard reference compounds deriving from an *in-house* database specifically built.

Although the work presented in this PhD Dissertation has been mainly focussed on the application of LC-MS protocols, it is a part of a larger multidisciplinary project aimed at achieving the comprehensive knowledge of the whole quali-quantitative profile of natural complex products by combining different analytical facilities. Thus, the investigation disclosed as well as the results achieved within this Doctoral work should be considered as a preliminary exploratory attempt to face the fascinating challenge to guarantee the quality, efficacy and safety of phytochemicals rather than an exhaustive dissertation on the topic.

2 Results and Discussion

2.1 On the way to a validated untargeted metabolomics approach for monitoring the batch-quality compliance of Grintuss® syrup

According to the holistic approach and starting from the general assumption that several metabolites may contribute to the whole activity of natural complex products, in the first part of this Thesis the biological uniformity among different batches within the production supply chain of formulated natural complex products, has been investigated by focussing on the whole formulation rather than on the single components. In particular, with the aim to overcome reproducibility issues associated with the traditional untargeted MS-based metabolomics analysis, in this study a multivariate statistical process control (MSPC) for monitoring the batch-quality of Grintuss® adult syrup has been successfully adopted. As previously discussed, the purpose of non-targeted metabolomics studies is to find statistically significant biomarkers through unbiased differential analysis of as many signals as possible deriving from the biological sample under investigation. In particular, within phytochemical untargeted investigations of complex natural products, the main task is to highlight those attributes accounting for the standard quality manufacturing of the formulated product without the need to identify at which metabolites these biomarkers correspond. In both cases, the comparison will be statistically valid only whether all the samples are analyzed under the same experimental conditions using a robust analytical method and whether the change observed in the metabolomic profile can be unquestionably related to variations in the concentration of specific biomarkers object of the investigation.[78] Although exhaustive guidelines for the validation of non-targeted approaches are still not available, the main validation strategies currently adopted in such investigations involve the use of quality control samples (QCs),[52,79] spiked standards[80] and statistical model validation.[64,65,81] Besides the efforts to set up a unique validation protocol, the MSI is defining the criteria for the standardization of data reporting in

order to make the data readily available to others for evaluation, support an extension or repeat of the work as desired.[82] In pursuing this aim, the workflow adopted involved the preliminary sample preparation followed by its direct infusion through flow injection into a mass spectrometer endowed with an ESI interface and an ion trap as analyzer. In this regard, within the untargeted approach, the direct infusion *via* flow injection electrospray ionization mass spectrometry analysis (FIA-ESI-MS) of whole formulated natural complex products allows to avoid any prior chromatographic separation thus providing a huge set of data in high-throughput fashion. On the other hand, especially for large scale multi-batches experiments, analytical variability in automated direct infusion mass spectrometry (DIMS) metabolomics analysis, as FIA-ESI-MS, is a well-known issue since hundred of metabolites are simultaneously measured in a single experiment.[83] For this reason, the use of a metabolomics analysis for quality evaluation of complex natural products batches through different years of production is a challenging task. To face this challenge, a specific data processing procedure, based on the building of multivariate control charts, was applied. Thus, after having assessed the validity of the model using the multivariate control charts, the quality verification of a new batch under investigation was achieved by comparing its profile with the fingerprinting profile of batches prepared according to the optimized manufacturing process and validated by the traditional continuous quality control throughout the production chain (selected as model of "good" quality batches) and the profile of batches deriving from pre-established production errors (selected as model of batches of "poor" quality). In order to correctly compare samples analyzed within different experimental sessions, the use of the so called golden standard (GS) sample, a sample being endowed with the same quali-quantitative composition of the samples to be analyzed, was required together with a robust method for data alignment.

Experimental design

As the first step, one year of production (from April 2014 to May 2015) of Grintuss® adult syrup formulation was monitored to design the experiment. In particular, a total of 24 samples, each one corresponding to a different batch of production, was collected and of these, 6 batches were used as the test set while 18 were used to build the control chart. Moreover, within the test set, 3 samples, indicated as t4-t6, were specifically produced in laboratory by using a deliberately wrong composition in order to reproduce the profile of batches deriving from production errors and were therefore used as model of "poor" quality products.

In particular, "poor" quality batches were obtained by preparing Grintuss® adult syrup formulation with a wrong relative ratio of both the sugary component and the active freeze-dried extracts (**Table 5**, **entry 1**, sample t4) and without the active freeze-dried extracts (**Table 5**, **entries 2** and **3**, samples t5 and t6). On the other hand, the training set (18 samples) was used to reproduce the profile of "good" quality batches prepared according to the optimized manufacturing protocol (according to the GMP) and submitted to the standardized quality control operations. All the samples were prepared, preliminary analyzed and stored at -80 °C before performing the analysis. Thus, a preliminary data analysis of the training set based on the PCA allowed to identify the most suitable GS to be used in the experimental sessions. In particular, the GS was selected as the closest sample to the centre of the PCA score space for being the most representative sample of the training set. Thus GS, prepared according to the other samples and stored at -80 °C, was analyzed together with samples under investigation in different analytical sessions to monitor the instrumental variability.

Table 5 – Quali-quantitative composition of "poor" quality batches of Grintuss® adult syrup.

Entry	Sample	Composition
1	t4	Deionized water (81%), honey (15.57%), lemon juice (1.25%), essential oils (0.67%), Plantain mother tincture (0.67%), Curry plant (*Helichrysum italicum*) freeze dried extract (0.53%), Gumweed (*Grindelia robusta*) freeze dried extract (0.17%), Plantain (*Plantago lanceolata*) freeze dried extract (0.15%).
2	t5	Honey (81%), deionized water (19%)
3	t6	Brown sugar (50%), deionized water (25%), honey (25%)

Sample preparation

0.5 g of each sample were solubilised in ethanol/water (100 mL, 50:50 w/w). The solution obtained was analyzed, without further dilution, to achieve a signal intensity of 10^5-10^6. After the preliminary analysis, for the choose of the GS, an aliquot of each solution was stored at -80 °C.

FIA-ESI-MS analysis

The collected samples were thus analyzed in 11 different analytical sessions, with the test set corresponding to the last session. In all the sessions, experiments were run in triplicate with randomized run order while GS samples were run in regular positions within each analytical session. FIA-ESI-MS analysis was carried out in a HPLC system and the outcome was analyzed by an ion trap mass spectrometer equipped with an ESI interface operating in negative ion mode. Different mobile phases were tested (*data not shown*) and the mixture methanol/water 50:50 (v/v) was finally chosen as the best mobile phase in terms of analytical performance. Thus, the mobile phase consisted of methanol (A) and ultrapure water (B) according the following elution: 0-5 min A/B 50:50 (v/v). Before injecting a new sample, the system was equilibrated for 20 minutes by pumping the mixture A/B (50:50, v/v) as washing solvent. The flow rate was set-up to 0.2 mL min^{-1} with an injector volume of 5 µL. In absence of the column, the injection of the sample using the HPLC pump generated a peak that eluted between 0.05 min and 5.0 min, and this delay corresponds to the time required to the signal to return to the baseline.

Data processing

Different sources of variability may affect the data sets produced by FIA-ESI-MS with the result to cover the biological diversity of the collected samples. In particular, while the biological variability represents the key point for comparing the quality of different samples, the instrumental variability represents a critical aspect to be controlled and minimized. In this regard, in order to produce fingerprints and multivariate control charts able to bring out only the biological variability among the samples while minimizing the instrumental variability, specific procedures were applied. These include the randomized run sequences with GS and blank samples, calibration procedures for the detector and the use of suitable tools for data pre-processing and pre-treatment. Firstly, the data acquired were peak picked and, within each experimental session, the background was subtracted to the corresponding sample using *Subtract View Spectrum as Background* routine of the Data Analysis software (trap version 3.3, Bruker Daltonik GmbH). A typical Grintuss® adult syrup fingerprint mass spectrum background subtracted is depicted in **Figure 8**. As a result, for each sample a peak list with *m/z* values and the corresponding peak intensity was generated.

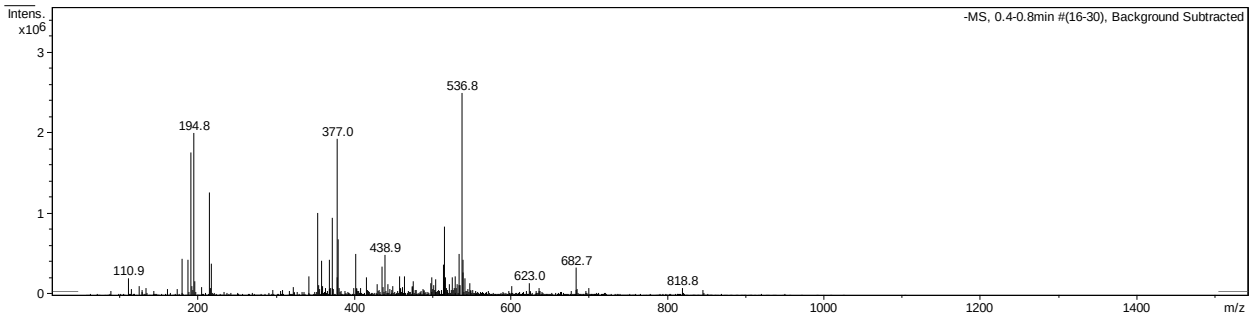


Figure 8 – Grintuss® adult syrup fingerprint mass spectrum background subtracted.

Since the analysis of the collected samples generate peak lists in which the same metabolite may exert different *m/z* values depending on the mass accuracy and the resolution of the mass spectrometer, a data matching procedure was required to match the corresponding peaks in the obtained peak lists (*data pre-treatment*). The result of the peak list matching was a data table having the sample representation in terms of peak intensity in rows and the corresponding *m/z* in columns for each detected metabolite. Peaks detected in different samples were matched through the calculation of a suitable match score (**Equation 1**).

$$\text{score}_{ij} = 1 - \left| \text{m/z}_i - \text{m/z}_{\text{master } j} \right| / d \quad \textbf{(Equation 1)}$$

where m/z_i is the *m/z* value for the peak *i* to be matched, $m/z_{\text{master}j}$ is the *m/z* value of the peak *j* of the master peak list and *d* the *m/z* tolerance.

The master peak list was defined as the list containing all the detected peaks represented by the mean of *m/z* values calculated for the samples showing that peak. Due to the *m/z* resolution of the mass detector used in this study, *d*= 0.8 was considered. The algorithm used for matching the peaks takes one peak from a peak list at time and matches the peak to either the best matching peak of the master peak list, or appends a new peak to the master peak list if matching is not found for the peak. The best matching is calculated on the basis of the match score while the *m/z* value of the master peak list is updated by mean calculation if the matching is found. The algorithm can be summarized as the following 4 steps:

1. The master peak list is initially set to the peak list of the first sample;
2. The peak list of the second sample is compared with the master peak list. For each peak of the peak list the match score is calculated and, on the basis of the best score, each peak is matched to a peak of the master peak list or added as new peak of the master peak list if the score is zero;
3. The *m/z* values of the master peak list are updated on the basis of the matched peaks;
4. Steps 2 and 3 are iterated thus including the other samples one at time.

This procedure generates a master peak list and a data table containing all the matched samples. When a new sample requires to be matched, step 2 of the algorithm is iteratively repeated without the updating step of the master peak list. Peaks showing match score equal to zero are automatically excluded from the data table and separately investigated to highlight the presence of potential contaminants. Since the master peak list could reflect the order of the samples used during the peak list matching procedure, a specific strategy was adopted to define the order of the samples to be used. In particular, a first data matching was performed by using a *m/z* tolerance equal to 2. The obtained data table was then centered on the mean and submitted to PCA considering 3 PCs. As a result, a ranking for the samples was defined according to the increasing distance to the centre of the PCA score space. The sample closest to the centre was used as the starting point for peak list matching (step 1) while the obtained ranking defined the order to use. The procedure was repeated until the convergence of the list to the ranking. Furthermore, if a matching peak is not found in some of the peak lists, the resulting data table may contain empty gaps. In this case, to reduce the number of gaps in the resulting data table, and then to minimize the influence of missing data in the control charts, only metabolites detected in more than 90% of the samples in the calculation of the master peak list were taken into account. The master peak list was built considering only the GS included in the training set. As a result, a master list composed of 529 peaks was obtained. For data normalization, the following expression for the intensity $\hat{I}_{ikJ}$ in the matched peak list was considered (**Equation 2**).

$$\hat{I}_{ikJ} = \alpha_k I_{iJ} + \delta_{ikJ} \quad \textbf{(Equation 2)}$$

where I_{iJ} is the intensity of the metabolite *i* in the sample *J*, α_k is a factor depending on the experimental run *k* and δ_{ikJ} is the background random noise.

The effect of the random noise (δ_{ikJ}) was reduced by calculating the median of the 3 replicates of each sample within the same experimental session while the effect of α_k was taken into account applying the Median Fold Change (MFC) normalization on the GS injections. In this regard, since the GS was run in all the different experimental sessions and different times within the same experimental session, it was possible to estimate α_k and therefore to normalize the data following a two-steps procedure. Thus, after having select a run of the GS as a reference, the ratio between the intensity measured for a specific run of the GS and the intensity of the reference was calculated for each variable of the master peak list and, finally, the median of the all fold changes was obtained. The median was then used to normalize all the samples nearest to the considered GS. The procedure was repeated for each run of the GS. It is worth of note that the GS selected as the reference does not affect the performance of the obtained control charts since different references produce data sets with a different factor scale without modifying the topology of the latent space.

Data analysis

After peak list matching, median of the replicates and MFC normalization, each sample resulted to be described by the same set of variables. The obtained representation was submitted to a multivariate data analysis by PCA to compare the features characterizing batches of “good” quality with respect to batches of "poor" quality.[84] To gain a deeper understanding of these differences, latent variables derived by PCA, the residuals of the model or suitable combinations of scores and residuals such as DModX+ (i.e. the distance to the model augmented with a term that measures how far projection of the observation falls outside with respect to the acceptable model domain) can be displayed by control charts.[85] This enables the specification of multivariate control limits, rather than control limits connected to univariate control charts of the single measured variables. Data were mean centered and Pareto scaled before performing data analysis. Multivariate statistical data analysis was performed by SIMCA13 (Umetrics, Umea, Sweden) while data pre-treatment was applied by R-functions developed within the platform R 3.0.2 (R Foundation for Statistical Computing).

Results

In order to reduce the analytical variability within different analytical sessions and to obtain control charts where the biological source of variation among the samples can be easily highlighted, in this study we introduced the use of GS samples in each experimental session and a suitable data pre-treatment based on the information collected in different analytical sessions. In particular, with the aim to reduce the performance decay of the ion source during the experiments, locked master peak list was used for matching the samples while the median of the replicates was applied to remove the effect of the random noise. Finally, data normalization based on MFC was performed on the GS to estimate the factor α_k specifying the response of the detector for each run.

As a result, the data set obtained was composed by 116 analytical runs (24 samples × 3 replicates + 44 GS runs) and 529 *m/z* variables. To highlight the effects of the proposed procedure for data pre-treatment, we firstly reported the investigation of the data set before calculating the master peak list by considering only the GS, the median of the replicates and applying data normalization. The obtained PCA model showed 2 PCs (R^2= 0.75, $Q^2_{CV7\text{-}folds}$= 0.68) (**Figure 9**).

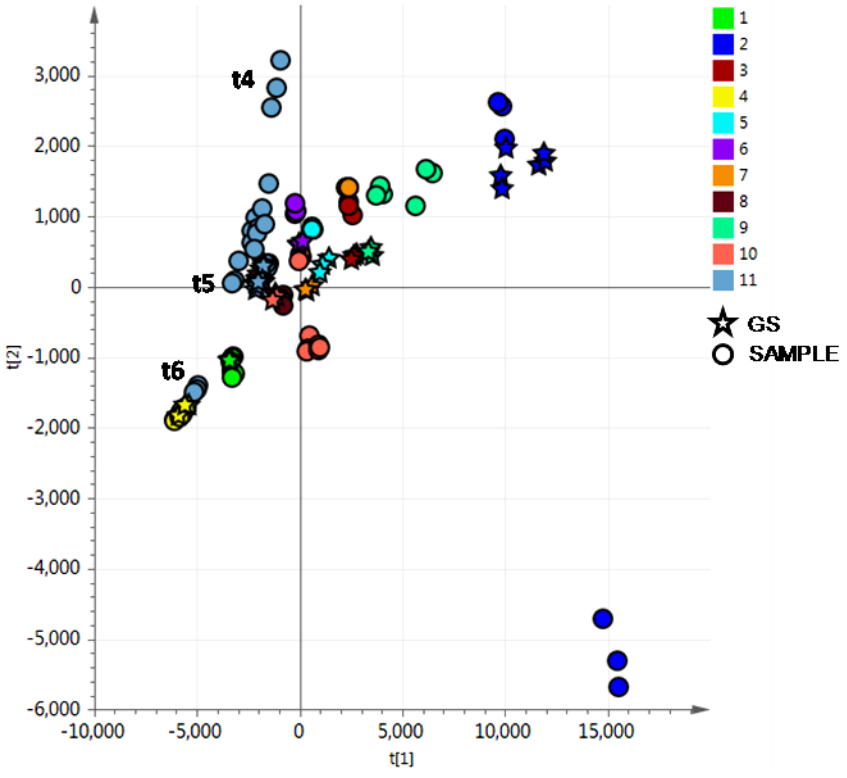


Figure 9 – PCA score scatter plot of the collected samples coloured according to the experimental session (R^2= 0.75). Stars indicate the GSs while samples are depicted as circles.

From the inspection of the PCA score plot it is possible to observe how the GS sample runs span a larger space with respect to the regions occupied by other samples as indicated by the fact that the variance of the X-block related to GS is 0.85 times than the variance of the other samples (**Figure 9**). In other words, since the GS sample runs correspond to the same sample, the variability produced in the data by the different experimental sessions has the same magnitude of the variability arising from the biological diversity among the samples. To reduce the effects of the experimental session, the pre-treatment procedure described above was applied. After the calculation of the master peak list considering only the GS, the median of the replicates and MFC normalization on the GS, the PCA model showed 2 PCs (R^2= 0.52, $Q^2_{CV7\text{-}folds}$= -0.0077) (**Figure 10**).

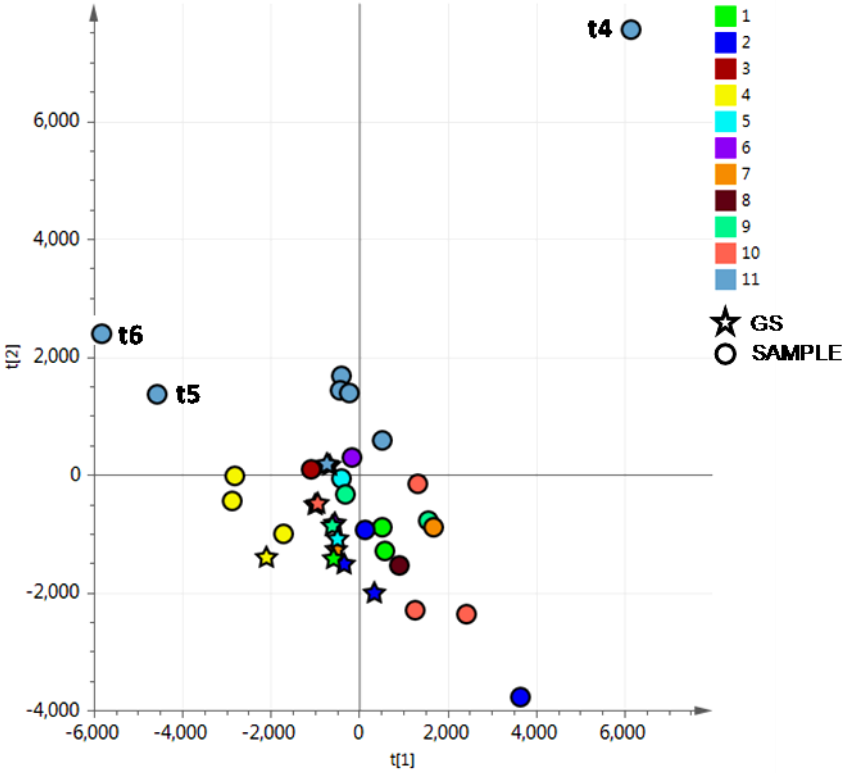


Figure 10 – PCA score scatter plot of the collected samples after data processing coloured according to the experimental session (R^2= 0.52). Stars indicate the GSs while samples are depicted as circles.

As depicted in **Figure 10**, after data processing the GS runs are more closed to each other with respect to the previous PCA score plot (**Figure 9**) and the effects due to the experimental session are strongly suppressed as demonstrated by the fact that the variance of the X-block related to GS is 0.15 times than the variance of the other samples. Thus, the control chart for MSPC was built on the basis of the PCA model of the data set composed of the 18 samples

representing batches of "good" quality. The PCA model showed 3 PCs (R^2= 0.63, $Q^2_{CV7\text{-}folds}$= 0.26). The test set composed of the last 6 produced batches was projected on the PCA model in order to extrapolate useful parameters for evaluating the compliance of the samples under investigation with respect to the batches defined of "good" quality. On the basis of the DModX+ control chart with a control limit equal to 99%, 3 batches resulted compliant while 3 batches resulted outlier. Interestingly, since the batches that resulted not compliant correspond to those specifically produced in laboratory by using a deliberately wrong formulation composition, we can conclude that the use of multivariate control charts allowed to endorse "good" quality batches while discriminating batches of "poor" quality (**Figure 11**).

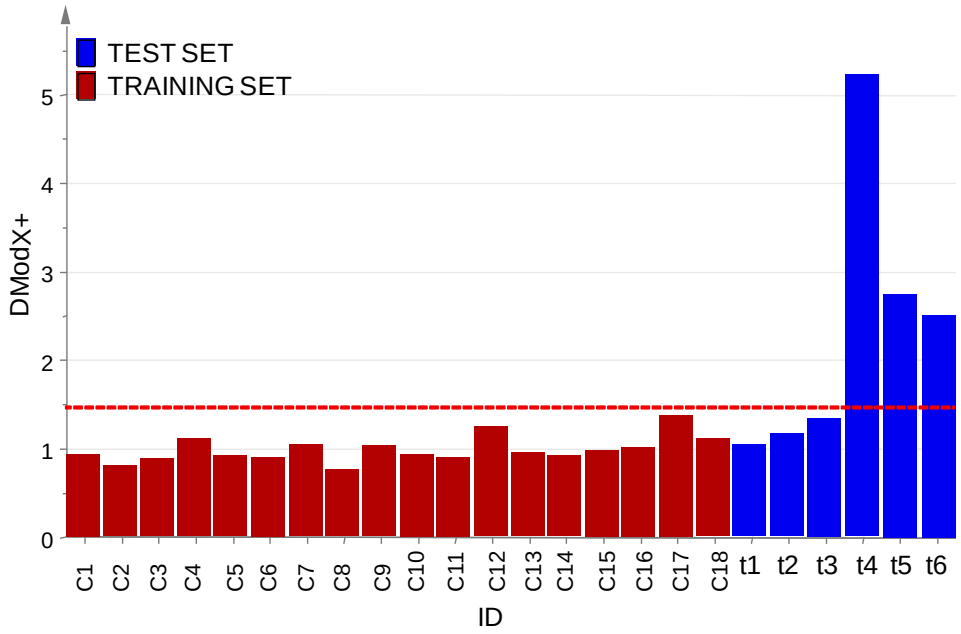


Figure 11 – DModX+ control chart (control limit equal to 99%). In red are reported the DModX+ of the samples of the training set (from C1 to C18) while in blue the DModX+ of the predicted samples (last 6 batches of production from t1 to t6).

The multivariate control chart used in this study allowed to investigate complex correlation structures within the samples that cannot be highlighted by classical univariate approach. As an additional goal, univariate data analysis based on the traditional control charts was further implemented in order to disclose significant variations in each single variable. Thus, although univariate control charts confirmed the conclusion of the multivariate control charts, it should be noted that in the multivariate case the DModX+ control chart is able to summarize the effects included in more than 500 univariate control charts. Moreover, also the part of the peak

list of the new samples to be investigated not included in the master peak list must be analyzed to evaluate the risk of poor quality batches. Usually, the number of these peaks is limited making the manual check not time consuming. In our study, peaks corresponding to the metabolites not included in the master peak list were around 15, accounting for only 2% of the total number of recorded signals, while in more than 80% of the samples metabolites not included in the master peak list were not found. In this study, a multinormal distribution was applied to describe the distribution of the "good" batches in the latent space of the PCA model. In case of more complex distributions, it is possible to apply the kernel density estimation method to evaluate the distribution of the training set samples in the latent space and the distribution of the residuals. The proposed method, based on the GS and on the master peak list used for peak matching, depends on the sample selected as GS. However, since the collected samples correspond to the same product, in principle it is possible to replace the old GS with a new one having approximately the same number of peaks of the master peak list. Therefore, if one want to re-use all the samples collected in the past, the master peak list should be updated by excluding the peaks not found in the new GS, while data normalization can be recalculated based on the new GS, once the new GS is run together the old GS.

In a routine quality control metabolomics analysis, the following workflow can be used as Standard Operating Procedure (SOP) for monitoring batches quality: acquisition of the new sample in triplicate with related GS and blank samples randomizing the experimental runs; matching of the replicates to the master peak list; normalize the data; calculation of the median of the replicates; projection of the sample on the control charts; evaluation of the compliance on the basis of the control charts (PCA scores, DModX+); update the control charts with the compliant samples. If the sample is not compliant to the "good" quality batches, it is necessary to proceed to checkout and only after more insights if the batch meets the quality requirements the control charts can be updated. Remarkably, this approach is currently under investigation at the Quality Control Department of Aboca S.p.a. as a new, alternative and/or complementary method to the traditional quality control process, for monitoring the quality of formulated natural complex products. In this regard, since the goal of the proposed approach was to define statistically significant criteria to evaluate the batch quality of formulated natural products without the need to identify the single metabolites, its main limitation at this stage is related to the impossibility to understand and define, within the obtained metabolomic fingerprinting comparison, the metabolites that, changing in their quali-quantitative profile, are

responsible for the incompliance to the quality requirements. Thus, as a future perspective this study will be further integrated with a complementary targeted analysis to recognize those metabolites acting as biomarkers for the batch quality and this preliminary investigation will be object of the following paragraph of this PhD Thesis.

Materials and methods

Absolute ethanol 99.8% and pure methanol 99.9% were purchased from Sigma-Aldrich (*St. Louis, MO*). Ultrahigh purified water was prepared in a PURELAB® Ultra water purification system (ELGA, UK). Grintuss® adult syrup was produced by Aboca S.p.a. (*Sansepolcro, Italy*). Samples preparation consisted in the solubilisation of 0.5 g of Grintuss® adult syrup in a 100 mL glass volumetric flask with ethanol/water (50:50, w/w) in ultrasonic bath (Branson 5800 Ultrasonic Cleaner) for 20 minutes setting temperature at 35 °C, then make up to the mark at room temperature. The solution obtained was then homogenized and, after filtration on 0.45 µm Millipore cellulose acetate syringe filter, was transferred in a vial and analyzed without further dilution. An aliquot of each filtered solution was also stored at -80 °C. FIA-ESI-MS analysis was carried out in a HPLC 1100 Series (*Agilent Technologies INC., Santa Clara, CA*) system equipped with a *vacuum* degasser, a binary pump, a Peltier thermostated autosampler at 10 °C and the outcome was analyzed by an Ion Trap Mass Spectrometer SL series equipped with an ESI interface (*Agilent Technologies INC., Santa Clara, CA*). The column was replaced by an adapter connecting the capillaries between the injector and the interface and the flow rate was 0.2 mL min^{-1}, with an injector volume of 5 µL. The mobile phase consisted of methanol (A) and ultrapure water (B) according the following elution: 0-5 min A/B 50:50 (v/v). Before injecting a new sample, a system equilibration of 20 minutes was performed, using A/B 50:50 (v/v) as eluting solvent. The MS parameters were optimized in wide mode, with a width from 50 to 1,500 *m/z*. The ion trap mass spectrometer optimized parameters were the following: entrance capillary voltage: +4,500 V; end plate offset: -500 V; dry gas flow: 8 L min^{-1}; dry temperature: 350 °C; skimmer: -33.1 V; capillary exit: -111.5 V; oct 1 DC: -12.00 V; oct 2 DC: -1.70 V; otc RF: 300.0 Vpp; lens 1: 5.5 V; lens 2: 57.2 V; trap drive: 45.2; nebulizer gas pressure: 20 psi. Multivariate statistical data analysis was performed by SIMCA13 (*Umetrics, Umea, Sweden*) while data pre-treatment was applied by R-functions developed within the platform R 3.0.2 (*R Foundation for Statistical Computing*).

2.2 Metabolites profiling of Grintuss® syrup by a targeted metabolomics approach

Advances in phytochemical investigations have always been related to innovations in analytical technology. Thanks to the improved performance in terms of pressure limits, extra-column dispersion and shorter time of analysis with respect to traditional HPLC, UHPLC-based analytical platforms have been becoming a standard technique for quali-quantitative plant metabolomics studies.[86] On the other hand, the high specificity, accuracy and selectivity associated with the relatively short time for methods development have thrust MS techniques coupled with UHPLC systems into the limelight as valuable tools for quali-quantitative plant metabolomics studies. Obviously, the ability to enable simultaneous qualitative and quantitative profiling in a single run is highly desirable and, to this aim, UHPLC-MS platforms must be assisted by powerful software for data handling. However, despite the unremitting advances in terms of mass resolution, mass accuracy, mass range and sensitivity, the identification and quantification of metabolites in comprehensive plant metabolomics experiments with a single analytical LC-MS platform is still not possible due to impossibility to cover the whole chemical space of plant metabolome. As an additional challenge, metabolomics studies of formulated natural products in a single analysis imply the detection of thousands of components, some of them are unknown or still not commercially available as reference standards, within a wide range of chemical structures, molecular masses, dynamic concentration range and MS responses. This requires to solve issues related with both the chromatographic method and the MS parameters, including the coelution of highly polar metabolites at the solvent front, the competition in the use of the energy available for ionization at the source of the MS, mutual ion suppression phenomena and matrix effects. Thus, many factors need to be considered to set-up a LC-MS platform for natural products analysis: sample preparation, which should be as simple as possible with the minimal number of steps required; selection of LC-MS instruments; optimization of LC conditions and MS parameters for high resolution and sensitivity, as well as exploitation of structure information based on fragmentation patterns for the confirmation of known compounds or the identification of unknown components.[86]

Based on the properties of the phytochemical constituents to be analyzed in a qualitative LC-MS analysis, the most appropriate ionization mode (positive or negative) should be chosen accordingly. In this regard, the elucidation of the fragmentation patterns based on MS^n or MS/MS spectra in LC-MS is still far from reach the level of knowledge obtained for GC-MS over

the years. Thus, LC-MS-based qualitative plant metabolomics studies require the construction of standard reference compounds databases useful to understand the relationship between structural features and fragmentation patterns, allowing the matching of target metabolites in natural complex products.[86]

On the other hand, quantitative analysis of phytochemical constituents are currently performed using MS instruments capable of MS/MS ion transition like QqQ or Q-Trap. The combination of the specific parent mass and the unique fragment ion is used to selectively monitor the compound of interest, which highly improves the sensitivity and specificity of analysis. Generally, the limit of quantification using LC-MS/MS is between 0.01 and 100 ng mL^{-1}. However, for the quantitative analysis of whole formulated natural products, matrix effects are likely to occur, thus resulting in a significant reduction of the accuracy and precision of the measurement using LC-MS. In this regard, the sample preparation, as well as the optimization of LC and MS parameters, are crucial to solve ion suppression phenomena and to reach an acceptable compromise between resolution and analysis time.[86]

The last years have witnessed a marked growth in the applications of UHPLC-MS for natural products analysis both in companies and academia. Encouraged by the promising results achieved towards the batch quality evaluation of whole formulated natural products and with the aim to provide a deeper knowledge of their quali-quantitative composition to comply the forthcoming recast regulation on medical devices (Regulation EU 2017/745), the second part of this PhD Thesis was focused on the development of a targeted metabolomics approach by means of liquid chromatography-high resolution mass spectrometry (LC-HRMS) technique for the metabolites profiling of whole formulated natural products.

In particular, choosing again Grintuss® syrup as the model of natural complex matrices, we have attempted to go beyond the traditional analytical studies by developing an unique UHPLC-ESI-Q-TOF method for its quali-quantitative metabolic profiling by avoiding any prior sample manipulation, controlling the matrix effect by simple dilution and using only one internal standard for normalization. The project has involved firstly the construction of an *in-house* database obtained by acquiring the MS and MS/MS spectra of high purity standard reference compounds purchased from different suppliers. The next step has been represented by the acquisition of the mass spectra of the precursor and fragment ions of the detected metabolites in the samples under investigation with the same LC method adopted for the standard compounds, followed by the matching assessment between the fragment ions mass spectra of

the detected metabolites in the sample and the *in-house* fragment database. Finally, quantitative analysis of the recognized metabolites in the product was performed (**Figure 12**).

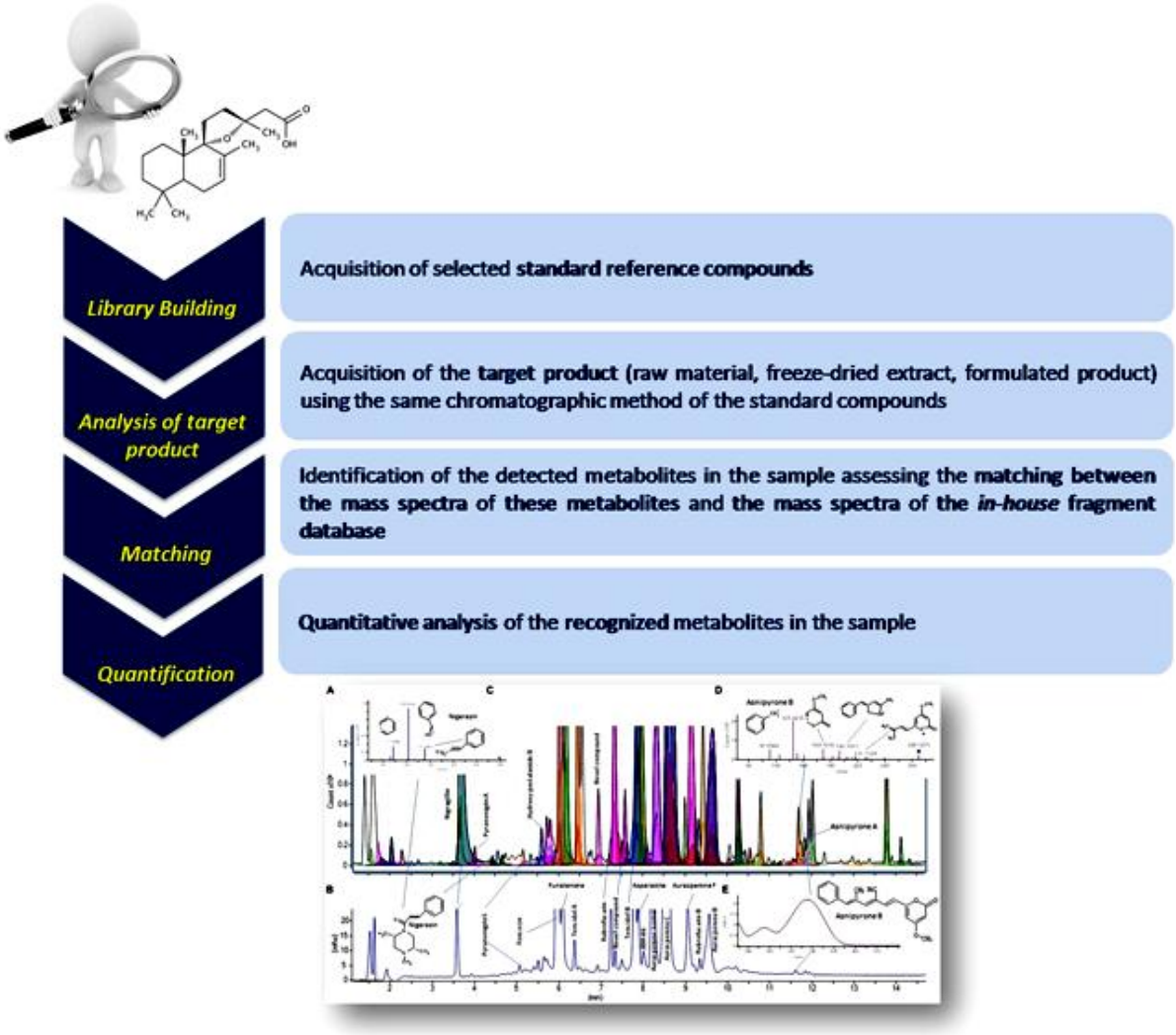


Figure 12 – Workflow adopted for the quali-quantitative metabolites profiling of natural complex products.

In-house database of high purity standard reference compounds

Multi-mode or simultaneous ion sources, as well as pulsed and parallel use of different ionization techniques are generally required to obtain detailed structural information from the wide variety of metabolites contained in complex natural matrices (**Figure 13**).[87]

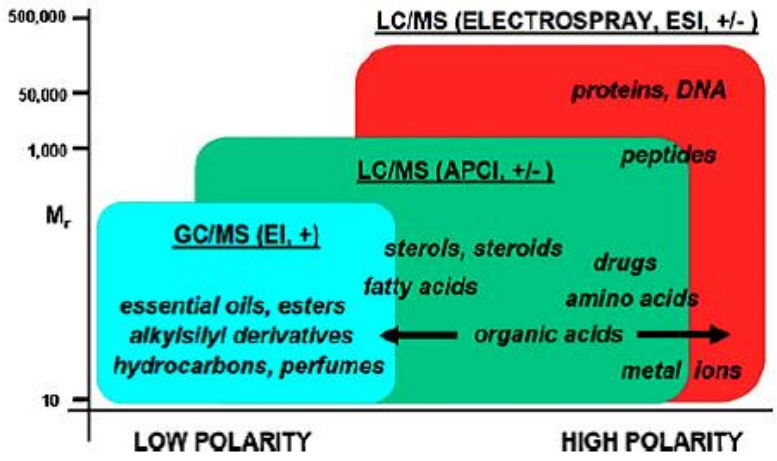


Figure 13 – Coverage of molecule classes with different ionization methods.[88]

In this regard, free chemical databases based on EI techniques are useful for mass spectral library matching since the hard ionization energy used affords highly reproducible fragmentations patterns. On the other hand, free chemical databases based on electrospray tandem mass spectra and other soft ionization techniques are more detailed but less reproducible since the resulting fragmentation patterns depend on a large number of factors, including the type of the instrument, ion source designs, ion source potentials, mobile phase effects, and many others which are not properly understood.

Thus, initial efforts have been devoted to build an *in-house* library containing the fragmentation patterns of reference compounds of interest. In pursuing this aim, almost 1,000 high-purity reference standards, belonging to different chemical classes, including phenols, terpens, organic acids, sulphur and nitrogen compounds and water-soluble vitamins, were purchased from Extrasynthese, Sigma-Aldrich, Phytolab and Chromadex. Accordingly, with the aim to set-up a general and unique UHPLC-ESI-Q-TOF protocol endowed with a broad applicability on the diverse classes of chemical components belonging to different complex natural matrices, a suitable LC-MS method for the acquisition of standard compounds and for the samples analysis was developed. The instrumental platform used was an UHPLC coupled to a hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer (**Figure 14**). For the chromatographic separation, the last generation C18 Cortecs® (Waters) was selected as highly efficient column. Characterized by a non-porous solid-core of silica in the center of the particle that is surrounded by an outer layer of porous silica, where the chromatographic separation takes place, this column is designed to give maximum efficiency, with the narrowest peak widths and highest resolution.[89]

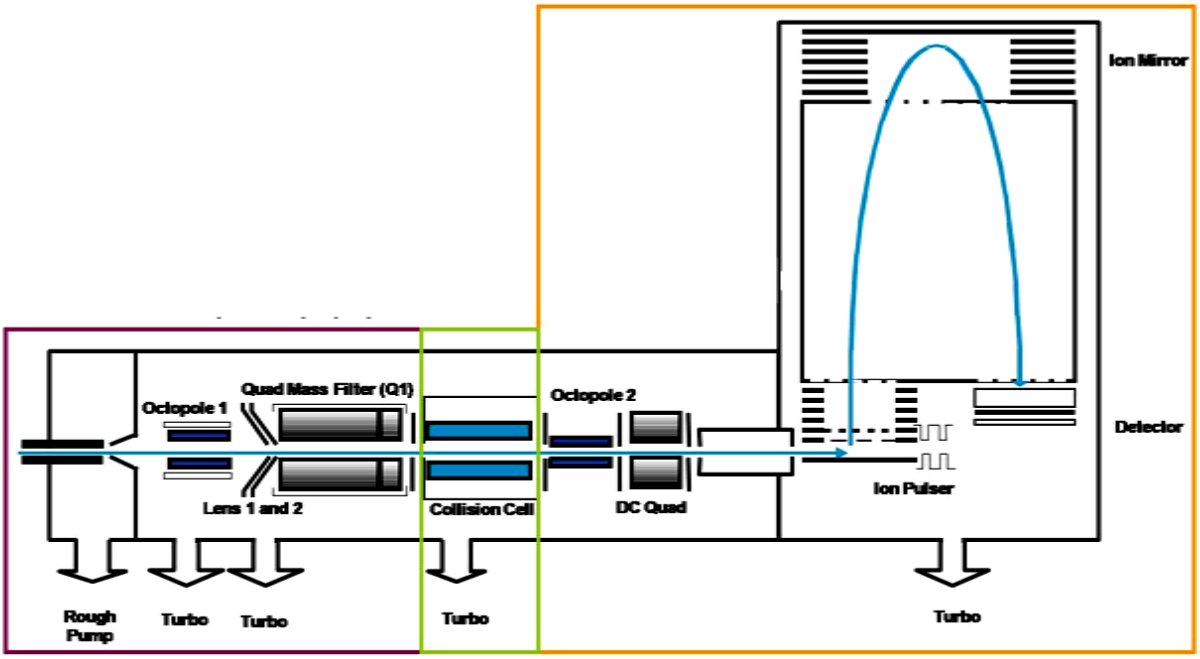


Figure 14 – Schematic representation of a hybrid quadrupole time-of-flight (Q-TOF) used for this study (adapted from Agilent Technologies).

The best compromise in terms of separation of the main constituents, method length and peaks resolution was obtained using aqueous HCO_2H (0.1%, v/v) and HCO_2H (0.1%, v/v in MeOH) as mobile phase with a gradient elution, from 1% to 99% of the organic phase in 24 minutes. So far, almost 800 standard reference compounds solutions were acquired at concentrations ranging from 100 to 500 ppb. In this regard, the acquisition in both ESI+ and ESI- mode by data dependent mode (Auto MS/MS modality in Agilent Technologies) allowed the isolation of the *pseudo*-molecular ion and its subsequent fragmentation to obtain a characteristic pattern of product ions for each reference standard. On the other hand, the use of three collision energy values (20, 30 and 40 eV) was crucial to ensure the dissociation of all the reference compounds allowing to obtain more detailed fragmentation patterns. The information acquired for each standard compound (molecular formula, monoisotopic mass value, structure, retention time value and MS/MS spectra) were collected in the *in-house* database using Agilent MassHunter Personal Compound Database and Library (PCDL) software (**Figure 15**).

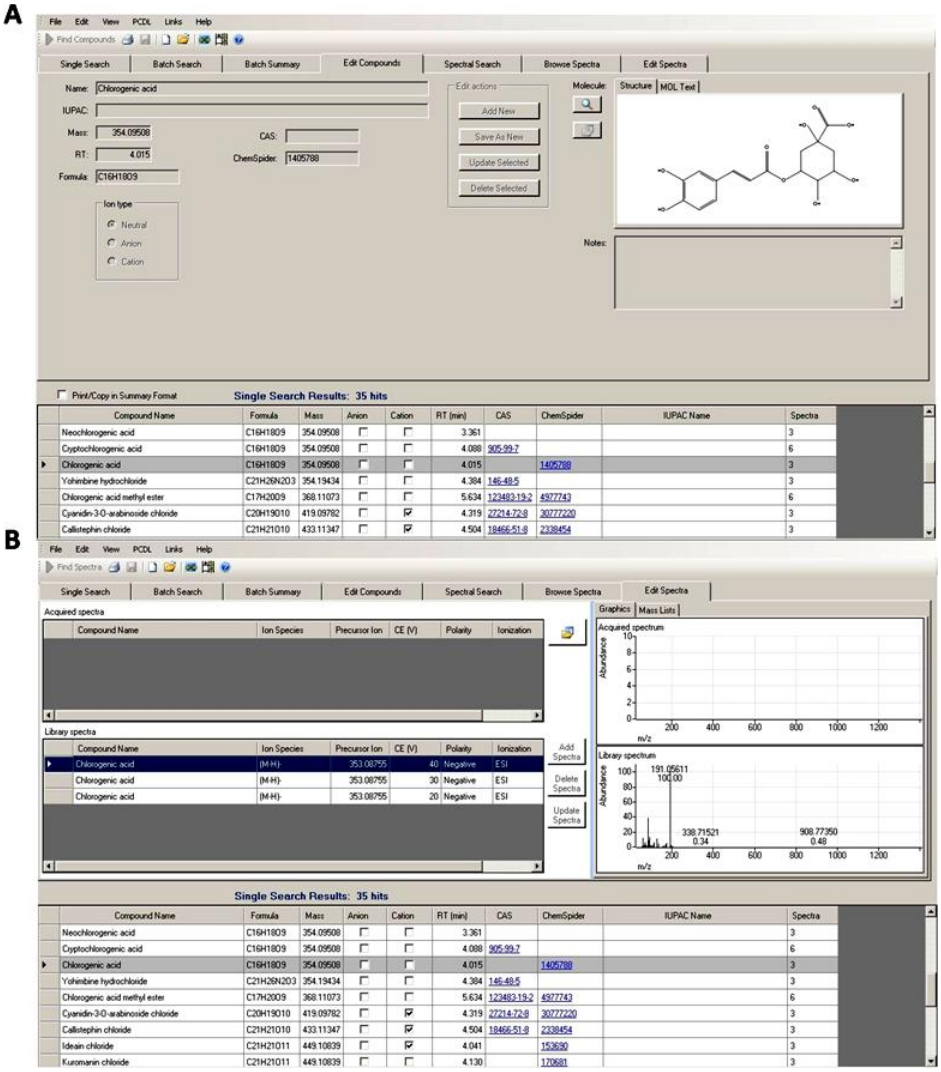


Figure 15 – Screenshot of the *in-house* database interface showing molecular formula, monoisotopic mass value, structure, retention time value (**A**) and MS/MS spectra acquired at 20, 30 and 40 eV (**B**) of a selected reference standard.

Calibration curves for the reference standard compounds

After having collected the retention time and the fragmentation pattern of each standard compound, the second part of the work was dedicated at the construction of calibration curves for the subsequent quantitative analysis. In this regard, given the high number of reference standard compounds acquired for the *in-house* library, which is still under continuous implementation, the building of a calibration curve for each compound would be extremely time-consuming, expensive and low throughput. Thus, 6 stock solutions were prepared by mixing 50-150 standard reference compounds for each mix stock solution. Moreover, in order to simplify as much as possible the process by preventing incorrect detections, compounds

endowed with identical monoisotopic mass and similar retention time were dissolved in different mix stock solutions, while no further consideration concerning the potential mutual chemical reactivity among the components dissolved in the same stock solution has been made. As well known, a quantitative determination requires the acquisition of reference standard compounds dissolved in the same matrix of the sample to be analyzed. In other words, a typical quantitative determination is matrix-dependent and requires the construction of a different calibration curve for each matrix under investigation. In this regard, the unavailability of certified blank matrices for natural complex products as well as the need to develop a general targeted metabolomics approach applicable to different natural complex matrices, led us to built the calibration curves in pure solvent and therefore to consider a semi-quantitative rather than a properly called quantitative approach. Thus, each mix stock solution was prepared in H_2O/MeOH (50:50, v/v) and opportunely diluted to obtain from 5 to 10 calibration standards. Moreover, an aliquot of sulfadimethoxine-d^6 stock solution (5 ppm, 20 μL per mL of calibration standard), selected as the internal standard, was added to each mix calibration standard. All these operations were conducted in a high-throughput fashion by using Agilent Sample Prep WorkBench (**Figure 16**).

Figure 16 – Agilent Sample Prep WorkBench.

The mix calibration standards were acquired in triplicate using the same LC method previously adopted for building the *in-house* library but with a different MS modality. In particular, data were acquired both in ESI+ and ESI- mode with a “low channel” and “high channel” using Collision Energy (All Ions mode in Agilent Technologies). In other words, the

acquisition of each mix calibration standard was performed using two simultaneous scanning functions: the first acquisition was performed at low energy in order to extrapolate the exact mass precursor ion spectra, while the second was acquired at higher energy to provide the exact mass of the fragment ions. Then, for each mix stock solution, the data file of a selected mix calibration standard (for convention the most concentrated point of the curve was chosen) was loaded into MassHunter Qualitative analysis software and the algorithm Find-by-Formula was run on the low channel against the *in-house* MS/MS library. Thus, the algorithm allowed the matching between possible precursor formulas and their product ions found in the library. Using the list of product ions, the software extrapolated extracted ion chromatograms (EICs) from the high channel and aligned them with an EIC of the precursor. A coelution score was calculated and compounds which passed the user-set threshold were retained. Then, the compounds were exported as a Compound Export Format (.CEF) file and loaded into the MassHunter Quantitative software. The list of precursors and product ions with their retention times was used to create the quantitative method. Following this approach, linear calibration curves were generated over the range 10-2,000 ppb with values for the correlation coefficient R^2 ≥0.98, according to the Agilent All Ions MS/MS Targeted Screening Workflow (**Figure 17**).

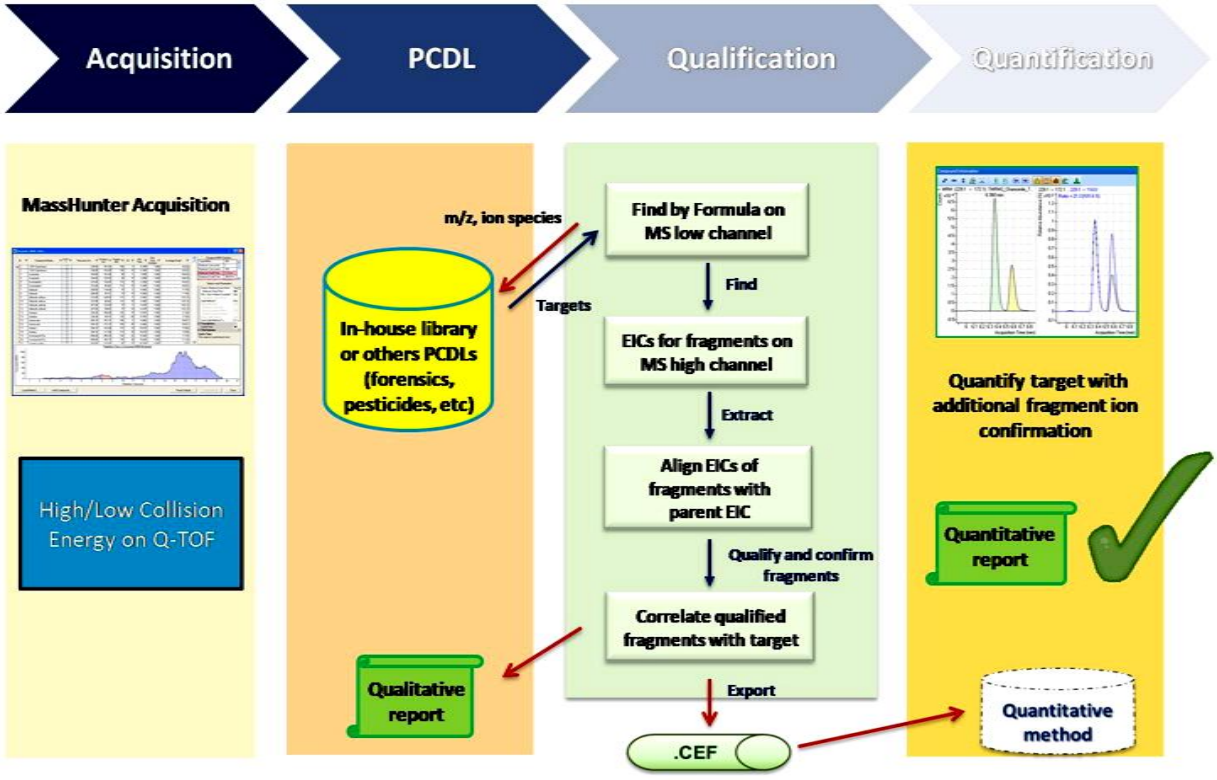


Figure 17 – All Ions MS/MS targeted screening workflow (adapted from Agilent Technologies).

Quali-quantitative metabolic profiling of Grintuss® pediatric syrup

The targeted metabolomics platform thus developed was applied for the quali-quantitative metabolic profiling of Grintuss® pediatric syrup. With this aim, three samples, for each of the four batches of Grintuss® pediatric syrup under investigation, were prepared by dissolving 0.5 g of the syrup in 50 mL of H_2O/MeOH (50:50, v/v) (10 mg mL^{-1}). The samples were then filtered and acquired both in negative and positive mode at three different concentrations (as it is, 1:10 and 1:50 in H_2O/MeOH, 50:50, v/v) in presence of sulfadimethoxine-d^6 as the internal standard (20 μL per mL of solution; 5 ppm concentration), using the same LC-MS method adopted for the calibration curves. Moreover, one calibration standard for each mix stock solution was acquired together with the samples and the relative accuracy on the curve previously obtained was calculated. As an example, the total ion chromatograms (TICs) in ESI- and ESI+ relative to batch n° 2 of Grintuss® pediatric syrup for the sample injected at 10 mg mL^{-1} are shown in **Figure 18**.

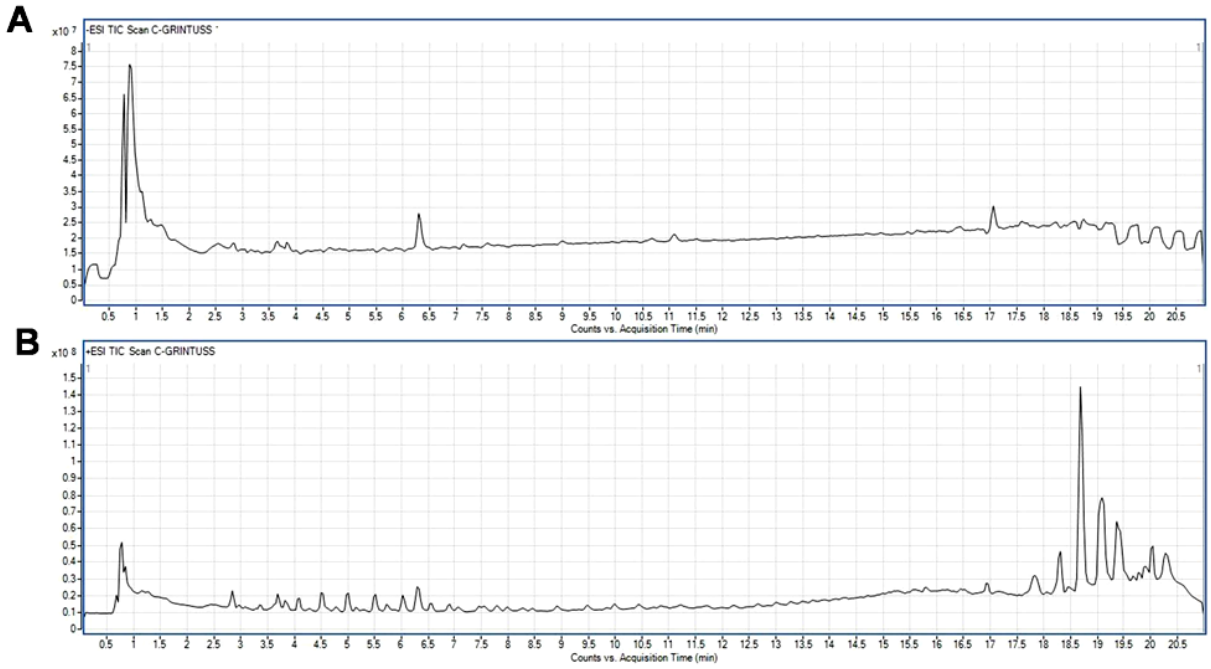


Figure 18 – Total ion chromatograms (TICs) in ESI- (**A**) and ESI+ (**B**) relative to batch n° 2 of Grintuss® pediatric syrup for the sample injected at 10 mg mL^{-1}.

The qualitative analysis was performed according to the previously described workflow. The data files for the different batches of Grintuss® pediatric syrup acquired as it is was loaded into MassHunter Qualitative analysis software and the algorithm Find-by-Formula was run to evaluate the presence of specified compounds by searching the chromatogram for each of the formula selected in the Formula Source. In particular, in this work the *in-house* database

natural_compound.cdb was selected as source of formula (**Figure 19A**). The match tolerance was set as ±10 ppm for masses and ±0.7 minutes for retention times and, concerning the scoring, the contribution (%) to overall score was set as 100 for mass score, 60 for isotope abundance score, 50 for isotope spacing score and 100 for retention time score. Moreover, the option *"Confirm with fragment ions"* in Fragment Confirmation tab was marked, thus the software selects known fragment ions from the MS/MS spectra, qualifies those fragment ions that can be seen with appropriate signal-to-noise, and then qualifies those fragment ions that have the same elution profile as the precursor ion (**Figure 19B**).

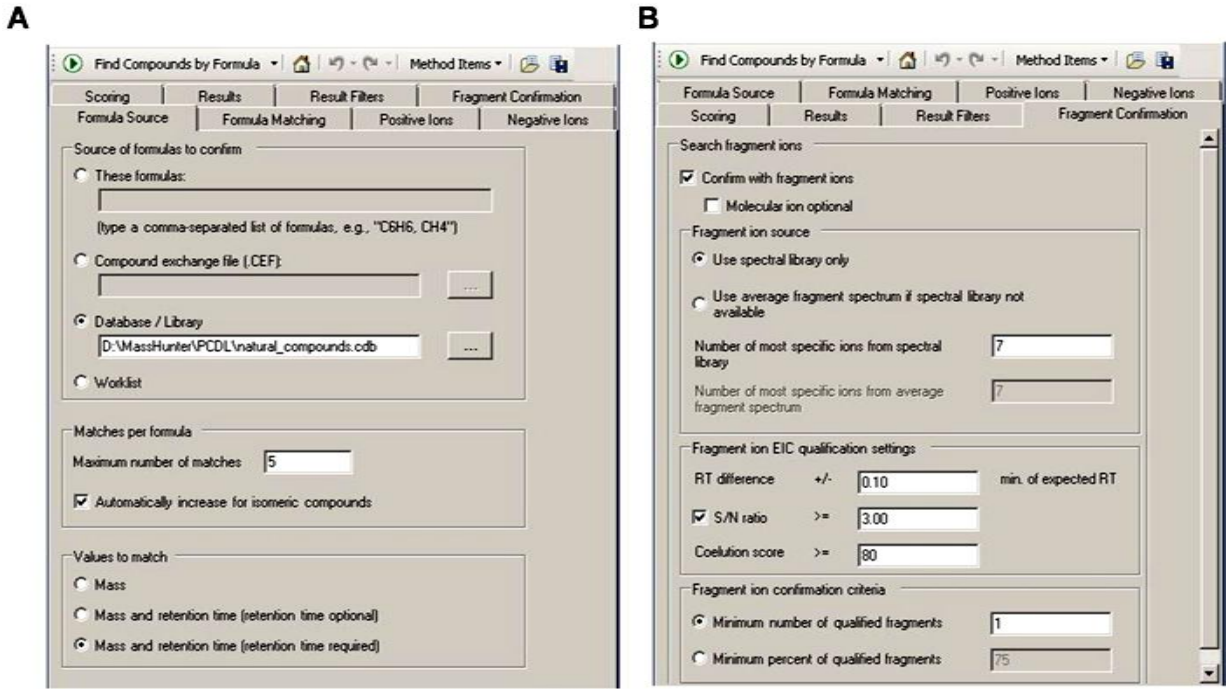


Figure 19 – Screenshot of the Find-by-Formula algorithm of MassHunter Qualitative analysis software showing the formula source tab (**A**) and the fragment confirmation tab (**B**) containing the fragment ion EIC qualification settings (retention time difference, signal to noise ratio and coelution score).

The identified compounds of the negative acquisition after the Find-by-Formula are shown in **Table 6**, while compounds qualified in the positive acquisition are shown in **Table 7**.

Table 6 – Qualified compounds of Grintuss® pediatric syrup in ESI-.

RT (min)	*m/z*	Formula	Compound name	Score	Flags (Tgt)
0.781	193.0370	$C_6H_{10}O_7$	*D*-Glucuronic acid[a] Galacturonic acid[a]	83.17 83.11	Qualified (multiple IDs)
0.849	191.0567	$C_7H_{12}O_6$	(-)-Quinic acid	87.62	Qualified
1.255	191.0206	$C_6H_8O_7$	Citric acid	94.67	Qualified[b]
1.458	225.0623	$C_6H_{12}O_6$	Fructose[a]	79.07	Qualified
1.594	187.0255	$C_6H_6O_4$	Kojic acid[a]	86.82	Qualified
2.338	407.1199	$C_{15}H_{22}O_{10}$	Catalpol	91.74	Qualified
2.846	391.1248	$C_{15}H_{22}O_9$	Aucubin	98.42	Qualified
3.049	153.0202	$C_7H_6O_4$	α-Resorcylic acid Protocatechuic acid	91.60 90.91	Qualified (multiple IDs)
3.185	353.0893	$C_{16}H_{18}O_9$	Neochlorogenic acid[a]	87.09	Qualified
3.354	339.0715	$C_{15}H_{16}O_9$	Aesculin (Vitamin C2)	74.16	Qualified (low score)
3.455	375.1293	$C_{16}H_{24}O_{10}$	Loganic acid	80.59	Qualified[b]
3.658	353.0879	$C_{16}H_{18}O_9$	Neochlorogenic acid Chlorogenic acid Cryptochlorogenic acid	93.91 91.87 89.12	Qualified (multiple IDs)
3.692	459.1146	$C_{18}H_{22}O_{11}$	Asperuloside[a]	95.24	Qualified
3.895	177.0204	$C_9H_6O_4$	Esculetin	55.43	Qualified (low score)[b]
4.572	639.1547	$C_{27}H_{30}O_{15}$	Vicenin-2[a] Saponarin	69.76 67.82	Qualified (multiple IDs)
5.013	479.0834	$C_{21}H_{20}O_{13}$	Quercetagetin 7-*O*-glucoside	78.03	Qualified
5.656	639.1923	$C_{29}H_{36}O_{16}$	Plantamajoside	86.66	Qualified
6.333	639.1896	$C_{29}H_{36}O_{16}$	Plantamajoside	60.62	Qualified (low score)
6.333	623.1983	$C_{29}H_{36}O_{15}$	Verbascoside	97.30	Qualified
6.468	447.0931	$C_{21}H_{20}O_{11}$	Luteolin 7-*O*-beta-*D*-glucoside Maritimein Homoorientin	92.74 92.27 80.15	Qualified (multiple IDs)
6.367	515.1192	$C_{25}H_{24}O_{12}$	3,4-Dicaffeoylquinic acid 3,5-Dicaffeoylquinic acid	92.39 88.71	Qualified (multiple IDs)
6.773	137.0244	$C_7H_6O_3$	Salicylic acid	88.32	Qualified
6.942	816.2622	$C_{35}H_{46}O_{19}$	Poliumoside[c]	51.28	Qualified (low score)

7.145	623.1972	$C_{29}H_{36}O_{15}$	Verbascoside[c]	79.39	Qualified
7.518	445.0777	$C_{21}H_{18}O_{11}$	Apigenin 7-*O*-glucuronide	97.63	Qualified
7.585	515.1192	$C_{25}H_{24}O_{12}$	4,5-Dicaffeoylquinic acid	95.59	Qualified
8.127	187.0975	$C_9H_{16}O_4$	Azelaic acid	71.88	Qualified (low score)
8.262	623.1607	$C_{28}H_{32}O_{16}$	Isorhamnetin 3-*O*-rutinoside	74.95	Qualified (low score)[b]
8.905	263.1272	$C_{15}H_{20}O_4$	(+/-)-Abscisic acid[a]	69.40	Qualified (low score)
9.955	271.0613	$C_{15}H_{12}O_5$	Naringenin Pinobanksin	88.05 83.12	Qualified (multiple IDs)
10.158	285.0404	$C_{15}H_{10}O_6$	Luteolin	80.17	Qualified
10.496	637.1868	$C_{28}H_{32}O_{14}$	Linarin[a]	62.41	Qualified (low score)
10.564	315.0516	$C_{16}H_{12}O_7$	6-Methoxyluteolin (Nepetin)	72.33	Qualified (low score)
11.546	269.0456	$C_{15}H_{10}O_5$	Apigenin Genistein	95.86 84.05	Qualified (multiple IDs)
11.884	299.0559	$C_{16}H_{12}O_6$	Hispidulin[a]	95.31	Qualified
12.629	329.0672	$C_{17}H_{14}O_7$	Quercetin 3,4'-dimethyl ether	87.01	Qualified
13.746	343.0820	$C_{18}H_{16}O_7$	Eupatorin[a]	91.94	Qualified
14.423	253.0507	$C_{15}H_{10}O_4$	7-Hydroxyflavonol 3,6-Dihydroxyflavone Chrysin	78.68 67.76 67.22	Qualified (multiple IDs)
15.032	283.0610	$C_{16}H_{12}O_5$	Acacetin[a] Genkwanin[a]	88.14 81.75	Qualified (multiple IDs)
17.063	401.1608	$C_{22}H_{26}O_7$	Arzanol	93.69	Qualified
18.688	319.2279	$C_{20}H_{32}O_3$	Grindelic acid	95.65	Qualified
19.331	455.3533	$C_{30}H_{48}O_3$	Bacosine[a] Oleanolic acid	87.23 85.21	Qualified (multiple IDs)
20.279	283.2642	$C_{18}H_{36}O_2$	Stearic acid[a]	93.35	Qualified

[a] Calibration curve not available. [b] Peak not well resolved or shifted. [c] Structural isomer.

Table 7 – Qualified compounds of Grintuss® pediatric syrup in ESI+.

RT (min)	*m/z*	Formula	Compound name	Score	Flags (Tgt)
0.745	104.1070	$C_5H_{14}NO$	Choline chloride	99.51	Qualified (no H-adduct)
0.779	118.0867	$C_5H_{12}NO_2$	Betaine[a]	98.59	Qualified (no H-adduct)
1.050	136.0617	$C_5H_5N_5$	Adenine[a]	74.02	Qualified (low score)
2.336	385.1060	$C_{15}H_{22}O_{10}$	Catalpol	57.33	Qualified (low score)
2.844	369.1156	$C_{15}H_{22}O_9$	Aucubin	97.89	Qualified
3.656	355.1028	$C_{16}H_{18}O_9$	Neochlorogenic acid[a] Chlorogenic acid Cryptochlorogenic acid	93.44 91.25 88.49	Qualified (multiple IDs)
3.690	437.1058	$C_{18}H_{22}O_{11}$	Asperuloside[a]	91.20	Qualified
3.758	395.1330	$C_{17}H_{24}O_9$	Eleutheroside B[a]	77.12	Qualified
6.229	633.1454	$C_{27}H_{30}O_{16}$	Luteolin 3',7-di-*O*-glucoside	70.47	Qualified (low score)[b]
6.331	479.1550	$C_{23}H_{26}O_{11}$	Calceolarioside B	73.87	Qualified (low score)[b]
6.364	539.1156	$C_{25}H_{24}O_{12}$	3,4-Dicaffeoylquinic acid 3,5-Dicaffeoylquinic acid	91.76 88.08	Qualified (multiple IDs)
6.364	463.0872	$C_{21}H_{18}O_{12}$	Luteolin 7-*O*-glucuronide	81.21	Qualified
6.906	793.2535	$C_{35}H_{46}O_{19}$	Poliumoside	62.37	Qualified (low score)
7.143	163.0387	$C_9H_6O_3$	3-Hydroxycoumarin	79.10	Qualified[b]
7.143	479.1546	$C_{23}H_{26}O_{11}$	Calceolarioside B	83.79	Qualified
7.515	447.0928	$C_{21}H_{18}O_{11}$	Apigenin 7-*O*-glucuronide	96.29	Qualified
7.786	478.1062	$C_{22}H_{22}O_{12}$	Nepetin 7-glucoside Isorhamnetin 3-*O*-glucoside[a]	79.57 66.03	Qualified (multiple IDs, no H-adduct)
8.903	287.1257	$C_{15}H_{20}O_4$	(+/-)-Abscisic acid[a]	74.09	Qualified (low score)
9.513	303.0508	$C_{15}H_{10}O_7$	Herbacetin[a] Tricetin Quercetin[a]	86.17 82.74 71.24	Qualified (multiple IDs)
10.156	287.0548	$C_{15}H_{10}O_6$	Luteolin	75.18	Qualified
10.257	229.0474	$C_{11}H_{10}O_4$	Citropten	76.76	Qualified
10.528	593.1880	$C_{28}H_{32}O_{14}$	Linarin[a]	62.06	Qualified (low score)
10.562	317.0667	$C_{16}H_{12}O_7$	6-Methoxyluteolin (Nepetin)	85.52	Qualified
11.577	271.0604	$C_{15}H_{10}O_5$	Apigenin Genistein	82.07 68.39	Qualified (multiple IDs)
11.882	301.0713	$C_{16}H_{12}O_6$	Hispidulin[a]	90.18	Qualified

12.627	331.0804	$C_{17}H_{14}O_7$	Quercetin 3,4'-dimethyl ether	78.06	Qualified
13.744	345.0974	$C_{18}H_{16}O_7$	Eupatorin[a]	87.25	Qualified
13.812	257.0811	$C_{15}H_{12}O_4$	Pinocembrin[a]	63.10	Qualified (low score)
14.590	343.1177	$C_{19}H_{18}O_6$	Scutellarein tetramethyl ether 3',4',7,8-Tetramethoxyflavone Luteolin tetramethylether	90.68 85.96 77.58	Qualified (multiple IDs)
14.658	425.1222	$C_{21}H_{22}O_8$	3',4',5,5',6,7-Hexamethoxyflavone Nobiletin	86.77 80.36	Qualified (multiple IDs)
15.403	849.4237	$C_{42}H_{66}O_{16}$	Bayogenin 3-*O*-beta-*D*-glucuronopyranoside-28-*O*-beta-*D*-glucopyranosyl ester[a]	75.20	Qualified
15.470	373.1285	$C_{20}H_{20}O_7$	Tangeretin	84.01	Qualified
17.129	425.1574	$C_{22}H_{26}O_7$	Arzanol	95.45	Qualified
18.585	351.1571	$C_{20}H_{24}O_4$	5-Geranyloxy-7-methoxycoumarin	77.58	Qualified

[a] Calibration curve not available. [b] Peak not well resolved or shifted.

The outcome of the algorithm is represented by a compound list table described by different columns: the retention time assigned to the compound, the mass-to-charge ratio of the precursor or largest evidence ion for this compound, the empirical formula, the name of the compound, the overall score of the identification results and the Flags (Tgt). These flags explain the reasons why there is not a match for this target or the reasons for which the match may be incorrect (**Figure 20**).

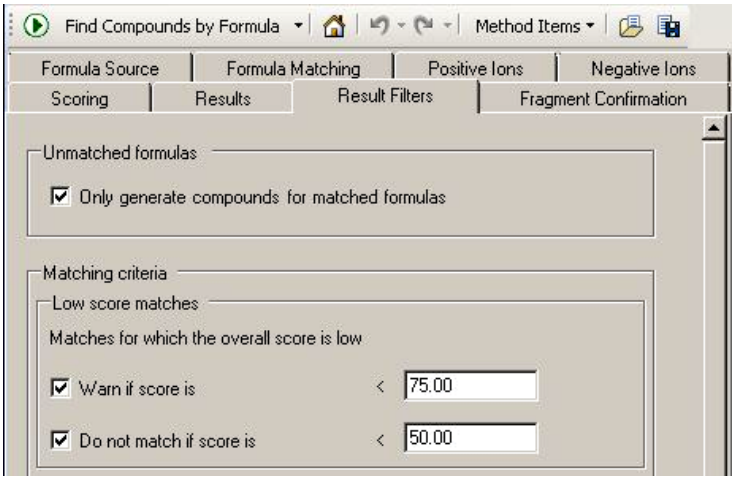


Figure 20 – Screenshot of the Result Filters tab of the Find-by-Formula algorithm of MassHunter qualitative analysis software.

In particular, compounds shown in **Tables 6** and **7** represent only the qualified entities with at least one fragment ion of the *in-house* MS/MS library. Moreover, the flag *Multiple IDs* indicates that a precursor formula found matched more than a compound of the *in-house* database.

Within the qualified compounds, 21 metabolites have been quantified so far and the results, expressed as % w/w (mean ± standard deviation) for the four batches of Grintuss® pediatric syrup prepared in triplicate, are shown in **Figure 21** and **Table 8**, while the relative EICs of the quantified compounds in negative and positive mode are reported in **Figure 22**.

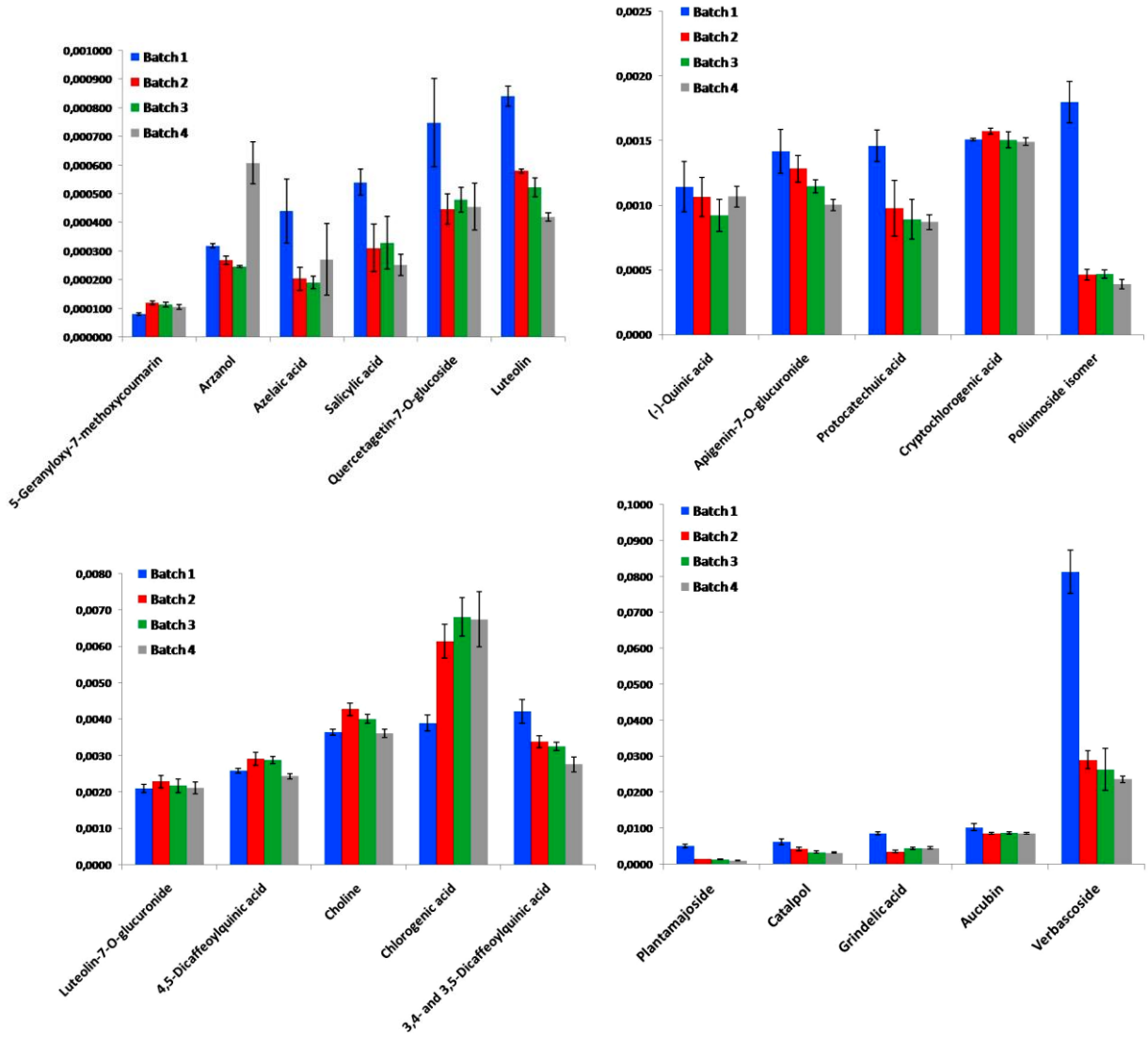


Figure 21 – Quantified metabolites in four batches of Grintuss® pediatric syrup.

Table 8 – Quantified compounds of Grintuss® pediatric syrup.

Compound	MS mode	Batch 1	Batch 2	Batch 3	Batch 4
		Mean (% w/w) ± SD	Mean (% w/w) ± SD	Mean (% w/w) ± SD	Mean (% w/w) ± SD
5-Geranyloxy-7-methoxycoumarin	ESI+	0.000080±0.000004	0.000119±0.000005	0.00011±0.00001	0.000105±0.000008
Arzanol	ESI+	0.00032±0.00001	0.00027±0.00001	0.000246±0.000002	0.00061±0.00007
Azelaic acid	ESI-	0.0004±0.0001	0.00020±0.00004	0.00019±0.00002	0.0003±0.0001
Salicylic acid	ESI-	0.00054±0.00005	0.00031±0.00008	0.00033±0.00009	0.00025±0.00004
Quercetagetin 7-*O*-glucoside	ESI-	0.0007±0.0002	0.00045±0.00005	0.00048±0.00004	0.00046±0.00008
Luteolin	ESI-	0.00084±0.00004	0.00058±0.00001	0.00052±0.00003	0.00042±0.00001
(-)-Quinic acid	ESI-	0.0011±0.0002	0.0011±0.0002	0.0009±0.0001	0.00107±0.00008
Apigenin 7-*O*-glucuronide	ESI-	0.0014±0.0002	0.0013±0.0001	0.00115±0.00005	0.00100±0.00004
Protocatechuic acid	ESI-	0.0015±0.0001	0.0010±0.0002	0.0009±0.0002	0.00087±0.00006
Cryptochlorogenic acid	ESI-	0.00151±0.00001	0.00157±0.00002	0.00151±0.00006	0.00149±0.00003
Poliumoside isomer	ESI-	0.0018±0.0002	0.00047±0.00004	0.00047±0.00003	0.00039±0.00004
Luteolin 7-*O*-glucuronide	ESI+	0.0021±0.0001	0.0023±0.0002	0.0022±0.0002	0.0021±0.0002
4,5-Dicaffeoylquinic acid	ESI-	0.00257±0.00007	0.0029±0.0002	0.00288±0.00009	0.00243±0.00007
Choline chloride	ESI+	0.00364±0.00008	0.0043±0.0002	0.0040±0.0001	0.0036±0.0001
Chlorogenic acid	ESI-	0.0039±0.0002	0.0061±0.0005	0.0068±0.0005	0.0067±0.0008
3,4- and 3,5-dicaffeoylquinic acid	ESI-	0.0042±0.0003	0.0034±0.0002	0.0033±0.0001	0.0028±0.0002
Plantamajoside	ESI-	0.0050±0.0004	0.00140±0.00002	0.00128±0.00008	0.00106±0.00007
Catalpol	ESI-	0.0062±0.0008	0.0042±0.0005	0.0034±0.0003	0.0032±0.0002
Grindelic acid	ESI-	0.0085±0.0004	0.0035±0.0003	0.0044±0.0004	0.0045±0.0004
Aucubin	ESI-	0.010±0.001	0.0086±0.0002	0.0086±0.0003	0.0086±0.0003
Verbascoside	ESI-	0.081±0.006	0.029±0.002	0.026±0.006	0.0236±0.0009

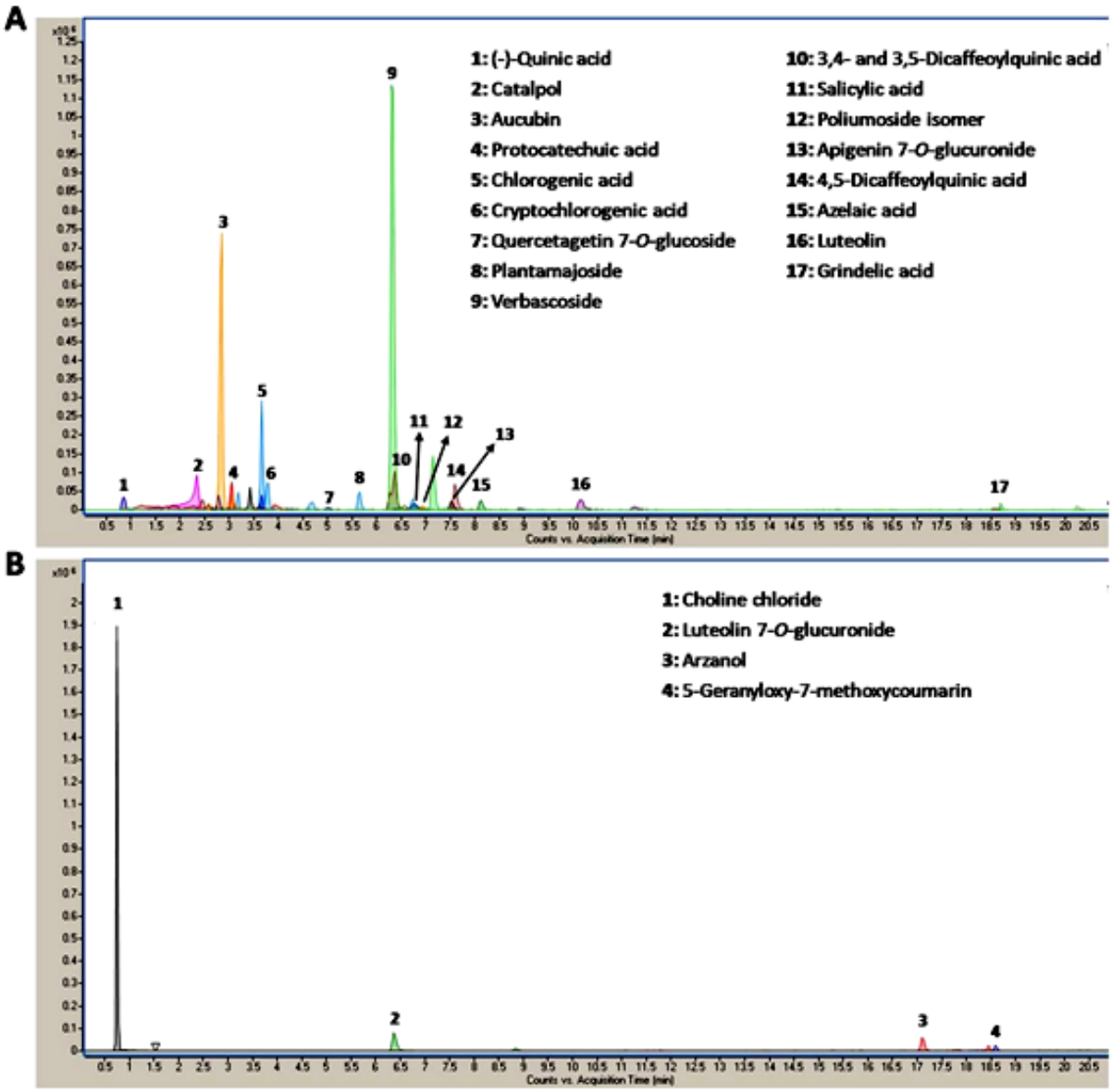


Figure 22 – Extracted ion chromatograms (EICs) in ESI- (**A**) and ESI+ (**B**) of the quantified compounds in Grintuss® pediatric syrup.

In **Tables 9** and **10** are reported the parameters for the calibration curve obtained for each quantified compound and the quantifier/qualifier ions used for the quantification, respectively.

Table 9 – Parameters for the calibration curve obtained for each quantified compound.

Compound	Slope	Intercept	R^2	Accuracy (%)
5-Geranyloxy-7-methoxycoumarin	0.006369	-0.026075	0.993	80.5
Arzanol	0.006726	-0.019887	0.997	70.2
Azelaic acid	0.016553	-0.005007	0.999	50.0
Salicylic acid	0.014720	-0.099543	0.995	122.7
Quercetagetin 7-*O*-glucoside	0.003952	-0.091069	0.985	37.5
Luteolin	0.016199	-0.262530	0.989	98.2
(-)-Quinic acid	0.006094	-0.206513	0.975	250.3
Apigenin 7-*O*-glucuronide	0.004105	-0.114936	0.987	152.7
Protocatechuic acid	0.009157	-0.279345	0.997	170.9
Cryptochlorogenic acid	0.008126	-0.348388	0.978	135.0
Poliumoside isomer[a]	0.004289	-0.024018	0.996	54.3
Luteolin 7-*O*-glucuronide	0.000948	+0.046902	0.997	81.9
4,5-Dicaffeoylquinic acid	0.002673	-0.020450	0.994	137.3
Choline chloride	0.004900	+0.172359	0.987	141.3
Chlorogenic acid	0.002489	-0.009334	0.997	174.2
3,4- and 3,5-Dicaffeoylquinic acid[b]	0.006219	-0.155355	0.994	53.0
Plantamajoside	0.005780	-0.143145	0.982	49.3
Catalpol	0.005843	-0.061174	0.985	67.7
Grindelic acid	0.000596	-0.015451	0.986	120.2
Aucubin	0.009797	-0.343048	0.991	73.7
Verbascoside	0.002859	-0.002965	0.995	151.4

[a] The reported parameters refer to the calibration curve obtained for poliumoside. [b] The reported parameters refer to the calibration curve obtained for 3,4-dicaffeoylquinic acid.

Table 10 – Quantifier and qualifier ions used for the quantification.

Compound	Quantifier Ion Species	Quantifier (*m/z*)	Qualifier (*m/z*)
5-Geranyloxy-7-methoxycoumarin	$[M+H]^+$	329.1747	193.0480
Arzanol	$[M+H]^+$	403.1751	181.0502; 155.0705
Azelaic acid	$[M-H]^-$	187.0976	125.0971
Salicylic acid	$[M-H]^-$	137.0244	93.0346; 65.0397
Quercetagetin 7-*O*-glucoside	$[M-H]^-$	479.0831	317.0293
Luteolin	$[M-H]^-$	285.0405	133.0292
(-)-Quinic acid	$[M-H]^-$	191.0561	85.0295; 59.0138
Apigenin 7-*O*-glucuronide	$[M-H]^-$	445.0776	269.0447
Protocatechuic acid	$[M-H]^-$	153.0193	109.0293
Cryptochlorogenic acid	$[M-H]^-$	353.0878	173.0455; 191.0558
Poliumoside isomer[a]	$[M-H]^-$	769.2561	161.0238
Luteolin 7-*O*-glucuronide	$[M+H]^+$	463.0871	287.0549
4,5-Dicaffeoylquinic acid	$[M-H]^-$	515.1195	173.0454; 353.0875
Choline chloride	$M+$	104.1070	60.0807; 58.0652
Chlorogenic acid	$[M-H]^-$	353.0878	191.0559; 85.0293
3,4- and 3,5-Dicaffeoylquinic acid	$[M-H]^-$	515.1195	173.0451; 179.0342
Plantamajoside	$[M-H]^-$	639.1931	477.1611
Catalpol	$[M+COOH]^-$	407.1195	169.0504; 199.0605
Grindelic acid	$[M-H]^-$	319.2279	205.1594
Aucubin	$[M+COOH]^-$	391.1246	165.0541; 183.0653
Verbascoside	$[M-H]^-$	623.1981	461.1650

[a] The reported information refers to poliumoside.

In conclusion, this study allowed the identification of almost hundred metabolites contained in Grintuss® pediatric syrup. Within the qualified compounds, 21 metabolites have been quantified so far allowing to reach 0.07-0.14% (w/w) of the overall composition of Grintuss® pediatric syrup. In particular, the quantified metabolites belong to the following classes: flavonoids (quercetagetin 7-*O*-glucoside, apigenin 7-*O*-glucuronide, luteolin, luteolin 7-*O*-

glucuronide); phenylpropanoids (poliumoside, chlorogenic acid, cryptochlorogenic acid, 3,4- and 3,5-dicaffeoylquinic acid, 4,5-dicaffeoylquinic acid, 5-geranyloxy-7-methoxycoumarin); phenethanoids (verbascoside, plantamajoside); salicylates (salicylic acid); simple phenols (arzanol); acid phenols (protocatechuic acid); iridoides (aucubin, catalpol); diterpenes (grindelic acid); carboxylic acids ((-)-quinic acid, azelaic acid) and hydro-soluble vitamins (choline).

As previously discussed, this work is still currently ongoing with a continuous implementation of the *in-house* library of standard reference compounds as well as by building of additional calibration curves required for the quantification of the remaining qualified metabolites. The quali-quantitative profiling obtained by applying this targeted metabolomics LC-MS-based approach was then combined with the results deriving from others analytical facilities for the analysis of additional compounds as polysaccharides and minerals allowing to reach so far up to 99.9% of the overall composition of Grintuss® pediatric syrup (**Table 11**).

Table 11 – Overall composition of Grintuss® pediatric syrup determined so far.

Class of compounds	Mean % (w/w)	Analytical technique
Phenols	0.0002047	GC-MS
Monoterpenes	0.0520554	GC-MS
Sesquiterpenes	0.0000500	GC-MS
Apocarotenids	0.0000340	GC-MS
Aromatic and aliphatic alcohols	0.0033180	GC-MS
Aromatic aldehydes, acids, esters and ketones	0.0026370	GC-MS
Esters	0.0000630	GC-MS
Mono- and di-saccharides	71.050000	GC-FID
Sterols	0.0022290	GC-FID
Threecarboxylic acids	0.1230000	HPLC-DAD
Polysaccharides	0.5487500	HPLC-RID
Fructoligosaccharides	0.2000000	Gravimetric
Minerals	0.0854530	ICP-MS
Proteins and aminoacids	0.0263730	HPLC-Fluorescent
Water	27.785000	Karl-Fisher
This study (mean)	**0.0866237**	**UHPLC-MS**
TOTAL	**99.9657908%**	

Although this work is primarily focused on the quali-quantitative profiling of Grintuss® syrup, the general approach and workflow described possess a more widely scope, thus opening the way for the metabolomics investigation of others formulated natural complex products and crude plant extracts according to the requirements recently approved by the European Parliament in the field of medical devices.

Materials and methods

All solvents were of high purity analytical grade and were used without further purification. ULC/MS grade absolute methanol was purchased from Biosolve (*Dieuze, France*). Ultrahigh purified water was prepared in a PURELAB® Ultra water purification system (*ELGA, UK*). Formic acid 98%-100% for LC-MS LiChropur® and dimethyl sulfoxide ≥99% were purchased from Sigma-Aldrich (*St. Louis, MO*). The high-purity reference standards, used both for the construction of the *in-house* database and for the construction of the calibration curves, were purchased from Extrasynthese (*Genay, France*), Sigma-Aldrich (*St. Louis, MO*), PhytoLab GmbH & Co. KG (*Vestenbergsgreuth, Germany*) and ChromaDex (*Irvine, CA*). High-purity reference standards stock solutions were prepared in methanol, water/methanol (80:20, v/v) or methanol/dimethyl sulfoxide (80:20, v/v) at 500 ppm. The working solutions were prepared by diluting appropriate volumes of the stock solutions with water/methanol (50:50, v/v). Internal standard sulfadimethoxine-d^6 was purchased from Sigma-Aldrich (*St. Louis, MO*). Internal standard stock solution was prepared in methanol at 5 ppm. All the stock solutions were stored in glass vials at -80 °C. Grintuss® pediatric syrup was produced by Aboca S.p.a. (*Sansepolcro, Italy*). Samples preparation consisted in the solubilisation of 0.5 g of Grintuss® pediatric syrup in a 50 mL glass volumetric flask with water/methanol (50:50, v/v) in ultrasonic bath (Branson 5800 Ultrasonic Cleaner) for 10 minutes setting temperature at 35 °C, then make up to the mark at room temperature. The solution obtained was then homogenized and, after filtration on 0.2 µm Millipore cellulose acetate syringe filter, was transferred in a vial and analyzed at different concentrations. The instrumental platform used consisted of an UHPLC series 1290 coupled to a hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer series 6545 (*Agilent Technologies, Santa Clara, CA*). The UHPLC was endowed with a binary pump, a multisampler, a multicolumn thermostat, an isocratic pump and a solvent cabinet. The multisampler was maintained at 15 °C. The analytical column was a Cortecs® C18 (100 x 2.1 mm, 1.6 µm) protected by a Cortecs® UPLC® C18 VanGuard™ pre-column (5 x 2.1 mm, 1.6 µm) both supplied by Waters (*Milford, MA*)

and thermostated at 40 °C. The analyses were performed in elution gradient using aqueous HCO_2H (0.1%, v/v) as mobile phase A and HCO_2H (0.1%, v/v in MeOH) as mobile phase B according to the following gradient table (**Table 12**), with a flow rate of 0.3 mL min^{-1} and an injector volume of 3 µL.

Table 12 – Elution gradient.

Time (min)	A%	B%
0.0	99	1
1.0	99	1
2.0	75	25
10.0	50	50
11.0	50	50
15.0	25	75
17.0	15	85
19.0	1	99
19.5	1	99
21.0	99	1
24.0	99	1

The UHPLC system was coupled to a Q-TOF mass spectrometer endowed with a Dual AJS ESI source operating in negative and positive ionization modes with a scan range from 50 *m/z* to 1,700 *m/z*. The optimized instrument parameters are reported in **Table 13**.

Table 13 – MS parameters.

Instrumental parameters	ESI-	ESI+
Gas temperature (°C)	325	325
Gas flow (L min^{-1})	11	11
Nebulizer (psig)	35	35
Sheat gas temperature (°C)	350	350
Sheat gas flow (L min^{-1})	12	12
VCap	3,500	3,500
Nozzle voltage (V)	0	1,000
Fragmentor	100	100
Skimmer	65	65
Octopole RF Peak	750	750

The construction of the *in-house* database was performed by acquiring each analytical standard reference compound in data dependent mode (Auto MS/MS, Agilent Technologies) using three different collision energy values (20, 30 and 40 eV) by N_2. While the acquisition of the calibration curves and the acquisition of the sample under investigation was performed in All Ions mode using simultaneously four acquisition collision energy values (0, 20, 30 and 40 eV). In both the acquisition methods, a reference mass solution containing purine and hexakis(1H, 1H, 3H-tetrafluoropropoxy)phosphazine was direct injected to the ESI source by the isocratic pump and ionized together with the sample solution for mass correction allowing to get accurate mass time-of-flight data. MassHunter software version B.07 (*Agilent Technologies, Santa Clara, CA*) was used for data acquisition and processing.

3. Concluding Remarks and Future Perspectives

Crude plant extracts and natural products deriving from them have been used since ancient in folk medicine for treating a broad range of human diseases and represent still today a priceless source of new therapeutically remedies. However, the high complexity, diversity and variability of formulated natural products in terms of composition makes a challenging task the guarantee of their quality, efficacy and safety requirements. In this regard, the current year has represented a breaking point in the regulation of medical devices, included those that derive from medicinal plant extracts and natural products, due to the recent approval of the EU regulation 2017/745 which will force the manufactures to clearly investigate the overall quali-qualitative composition of the device. In view of this stricter standardization of plant-based medical devices, an intense research towards advanced technologies and robust analytical protocols is required both for evaluating the composition and for reaching the required quality standards.

Based on these considerations and as a continuation of the phytochemical research activity in Aboca S.p.a., this PhD Thesis focuses on the application of targeted and untargeted mass spectrometry-based analytical platforms for the metabolomics investigation of formulated natural complex products. In pursuing this aim, Grintuss® syrup, a medical device used for treating cough and based on honey and functional vegetable substances from *Grindelia robusta*, *Plantago lanceolata* and *Helychrisum italicum*, was selected as a model of complex formulated natural products to set-up two complementary metabolomics approaches.

The first approach was based on an untargeted fingerprinting for monitoring the batch-quality of Grintuss® adult syrup by electrospray ionization mass spectrometry flow injection analysis (FIA-ESI-MS). The workflow used has involved the preliminary sample preparation followed its analysis by FIA-ESI-MS. Thus, after having assessed the validity of the model by using multivariate statistical process control (MSPC), the quality verification of a new batch

under investigation was achieved by comparing its profile with the fingerprinting profile of batches prepared according to the optimized manufacturing process and validated by traditional continuous quality verification throughout the production chain and the profile obtained from batches deriving from pre-established production errors. Remarkably, this approach is currently under investigation at the Quality Control Department of Aboca S.p.a. as a new, alternative and/or complementary method to the traditional quality control process, for monitoring the quality of formulated natural complex products.

The second part of this PhD Thesis describes the development and the application of a targeted metabolomics LC-MS-based approach for the quali-quantitative profiling of Grintuss® pediatric syrup. In pursuing this aim, an *in-house* database of high-purity standard reference compounds was built by collecting compounds belonging to different chemical classes. Thus, after the acquisition of MS and MS/MS spectra of standard compounds of interest, the matching between the detected metabolites from samples of Grintuss® syrup and standard reference compounds allowed to identify almost hundred metabolites, while 21 compounds have been quantified so far which account for 0.07-0.14% (w/w) of the overall composition of Grintuss® pediatric syrup. Worth of note, more than 99.9% of the overall composition of Grintuss® syrup has been reached so far by combining different analytical approaches, including GC-MS, GC-FID, HPLC-DAD, HPLC-RID, HPLC-fluorescent, ICP-MS, gravimetric analysis, Karl-Fischer tritation.

Overall, the two complementary metabolomics investigations developed so far represent a preliminary attempt to face the fascinating challenge to guarantee the quality, efficacy and safety of medical devices based on natural components according to the recently approved EU regulation 2017/745. In particular, while the untargeted approach can be considered as a routine method to establish the batch quality compliance of formulated natural complex products, the targeted method would be useful to gain a deeper understanding of the quali-quantitative profile of the specific metabolites of batch defined non compliant.

In conclusion, although this *Philosophiae Doctor* Thesis is mainly focussed on the application of LC-MS techniques, it is a part of a larger and multidisciplinary project aimed at achieving the comprehensive knowledge of the whole quali-quantitative profile of formulated natural complex products by combining different analytical methods.

4. References

1. a) Dias, D. A., Urban, S., Roeener, U. A historical overview of natural products in drug discovery. *Metabolites*, **2012**, *2*, 303-336. b) Cragg, G. M., Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta*, **2013**, *1830,* 3670-3695.
2. a) Mishra, B. B., Tiwari, V. K. Natural products: an evolving role in future drug discovery. *Eur. J. Med. Chem.* **2011**, *46*, 4769-4807. b) Carter, G. T. Natural products and Pharma 2011: strategic changes spur new opportunities. *Nat. Prod. Rep.* **2011**, *28*, 1783-1789.
3. a) Koehn, F. E., Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* **2005**, *4*, 206-220. b) Schmidt, B. M., Ribnicky, D. M., Lipsky, P. E., Raskin I. Revisiting the ancient concept of botanical therapeutics. *Nat. Chem. Biol.* **2007**, *3*, 360-366. c) David B., Wolfender, J., Dias, D. A. The pharmaceutical industry and natural products: historical status and new trends. *Phytochem. Rev.* **2015**, *14*, 299-315.
4. a) Nelson, A., Warriner S. Parallels with nature. *Nat. Chem.* **2014**, *6*, 845-846. b) Bucar, F., Wube, A., Schmid, M. Natural product isolation – how to get from biological material to pure compounds. *Nat. Prod. Rep.*, **2013**, *30*, 525-545. c) Rochfort, S. Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *J. Nat. Prod.* **2005**, *68*, 1813-1820.
5. Oksman-Caldentey, K., Inze, D. Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. *Trends Plant Sci.* **2004**, *9*, 433-440.
6. Kabera, J. N., Semana, E., Mussa, A. R., He, X. Plant secondary metabolites: biosynthesis, classification, function and pharmacological properties. *J. Pharm. Pharmacol.* **2014**, *2*, 377-392.
7. a) Borchardt, J. K. The beginnings of drug therapy: ancient mesopotamian medicine. *Drug News Perspect.* **2002**, *15*, 187-192. b) Attinger, P. La médicine mésopotamienne. *Le journal des Médecines Cunéiformes*, **2008**, *11*, 1-96.
8. Huang, K. C. *The pharmacology of chinese herbs.* CRC Press, Boca Raton, FL, **1999**.

9. Dev, S. Ancient-modern concordance in Ayurvedic plants: some examples. *Environ. Health Persp.* **1999**, *107*, 783-789.
10. Sneader, W. *Drug discovery: a history.* Wiley, **2005**.
11. Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E., Linder, T., Wawrosch, C., Uhrin, P., Temml V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., Stuppner, H. Discovery and resupply of pharmacologically active plant-derived natural products: a review. *Biotechnol. Adv.* **2015**, *33*, 1582-1614.
12. Ngo, L. T., Okogun, J. I., Folk, W. R. 21st Century natural product research and drug development and traditional medicines. *Nat. Prod. Rep.* **2013**, *30*, 584-592.
13. Farnsworth, N. R., Akerele, O., Bingel, A. S., Soejarto, D. D., Guo, Z. Medicinal plants in therapy. *Bull. World Health Org.* **1985**, *63*, 965-981.
14. a) Paterson, I., Anderson, E. A. Chemistry. The renaissance of natural products as drug candidates. *Science*, **2005**, *310*, 451-453. b) Baker, D. D., Chu, M., Oza, U., Rajgarhia, V. The value of natural products to future pharmaceutical discovery. *Nat. Prod. Rep.* **2007**, *24*, 1225-1244. c) Butler, M. S. Natural products to drugs: natural product-derived compounds in clinical trials. *Nat. Prod. Rep.* **2008**, *25*, 475-516. d) Newman, D. J., Cragg, G. M. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **2007**, *70*, 461-477.
15. a) Fabricant, D. S., Farnsworth, N. R. The value of plants used in traditional medicine for drug discovery. *Environ. Health Perspect.* **2001**, *109*, 69-75. b) Raskin, I., Ribnicky, D. M., Komarnytsky, S., Ilic, N., Poulev, A., Borisjuk, N., Brinker, A., Moreno, D. A., Ripoll, C., Yakoby, N., O'Neal, J. M., Cornwell, T., Pastor, I., Fridlander, B. Plants and human health in the twenty-first century. *Trends Biotech.* **2002**, *20*, 522-531.
16. Cragg, G. M., Newman, D. J., Snader, K. M. Natural products in drug discovery and development. *J. Nat. Prod.* **1997**, *60*, 52-60.
17. Patridge, E., Gareiss, P., Kinch, M.S., Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today*, **2016**, *21*, 204-207.
18. https://www.pharmacompass.com/data-compilation/product-sales-data-from-annual-reports-of-major-pharmaceutical-companies-2016 (*Last access: 11/03/2018*).
19. Schmidt, B., Ribnicky, D. M., Poulev, A., Logendra, S., Cefalu, W. T., Raskin, I. A natural history of botanical therapeutics. *Metab. Clin. Exp.* **2008**, *57*, S3-S9.
20. a) Hoffman, F. A., Kishter, S. R. Botanical new drug applications – the final frontier. *Herbalgram*, **2013**, *10*, 66-69. b) Chen, S. T., Dou, J., Temple, R., Agarwal, R., Wu, K., Walker, S. New therapies from old medicines. *Nat. Biotechnol.* **2008**, *26*, 1077-1083.

21. Nicoletti, M. Nutraceuticals and botanicals: overview and perspectives. *Int. J. Food Sci. Nutr.* **2012**, *63*, 2-6.
22. Pan, S., Zhou, S., Gao, S., Yu, Z., Zhang, S., Tang, M., Sun, J., Ma, D., Han, Y., Fong, W., Ko, K. New perspectives on how to discover drugs from herbal medicines: CAM's outstanding contribution to modern therapeutics. *Evid. Based Complement. Alternat. Med.* **2013**, 1-25.
23. Wagner, H., Ulrich-Merzenich, G. Synergy research: approaching a new generation of phytopharmaceuticals. *Phytomedicine*, **2009**, *16*, 97-110.
24. Yang, M., Poon, J., Wang, S., Jiao, L., Poon, S., Cui, L., Chen, P., Man-Yuen Sze, D., Xu, L. Application of genetic algorithm for discovery of core effective formulae in TCM clinical data. *Comput. Math. Methods Med.* **2013**, 1-16.
25. Wagner, H. Synergy research: approaching a new generation of phytopharmaceuticals. *Fitoterapia*, **2011**, *82*, 34-37.
26. Naoghare, P. K., Song, J. M. Chip-based high-throughput screening of herbal medicines. *Comb. Chem. High Throughput Screen.* **2010**, *13*, 923-931.
27. U.S. Food and Drug Administration. News and Events, FDA News Release (2012) FDA approves first anti-diarrheal drug for HIV/AIDS patients. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm599657.htm (*Last access: 11/03/2018*).
28. Mattoli, L., Burico, M., Fodaroni, G., Tamimi, S., Bedont, S., Traldi, P., Stocchero, M. New frontiers in pharmaceutical analysis: a metabolomic approach to check batch compliance of complex products based on natural substances. *J. Pharm. Biomed. Anal.* **2016**, *126*, 156-162.
29. Vlietinck, A., Pieters, L., Apers, S. Legal requirements for the quality of herbal substances and herbal preparations for the manufacturing of herbal medicinal products in the european union. *Planta Med.* **2009**, *75*, 683-688.
30. Racchi, M., Govoni, S., Lucchelli, A., Capone, L., Giovagnoni, E. Insights into the definition of terms in European medical device regulation. *Expert Rev. Med. Devices*, **2016**, *13*, 907-917.
31. https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-1/dir_2004_24/dir_2004_24_en.pdf (*Last access: 11/03/2018*).
32. Calapai, G. European legislation on herbal medicines: a look into the future. *Drug Safety*, **2008**, *31*, 428-431.
33. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003644.pdf (*Last access: 11/03/2018*).
34. http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2014/11/news_detail_002211.jsp&mid=WC0b01ac058004d5c1 (*Last access: 13/12/2017*).
35. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:247:0021:0055:en:PDF (*Last access: 11/03/2018*).

36. https://www.cromsource.com/wp-content/uploads/2012/12/EU-Recast-of-the-Medical-Device-directives.pdf (*Last access: 11/03/2018*).
37. a) http://eur-lex.europa.eu/legal-content/SK/ALL/?uri=CELEX:32017R0745 (*Last access: 11/03/2018*). b) http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0746 (*Last access: 11/03/2018*).
38. http://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework_en#new_regulations (*Last access: 11/03/2018*).
39. Villas-Bôas, S. G., Roessner, U., Hansen, M. A. E., Smedsgaard, J., Nielsen, J. *Metabolome analysis: an introduction*, John Wiley & Sons, Inc., **2007**.
40. Oliver, S., Winson, M. K., Kell, D. B., Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **1998**, *16*, 373-378.
41. Dettmer, K., Aronov, P. A., Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.
42. a) Sumner, L. W., Mendes, P., Dixon, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **2003**, *62*, 817-836. b) Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N., Lothar, W. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157-1161. c) Fiehn, O. Metabolomics – the link between genotypes and phenotypes. *Plant. Mol. Biol.* **2002**, *48*, 155-171. d) Fernie, A. R. The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. *Phytochemistry*, **2007**, *68*, 2861-2880. e) Saito, K., Matsuda, F. Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* **2010**, *61*, 463-489.
43. a) Patel, K. N., Patel, J. K., Patel, M. P., Rajput, G. C., Patel, H. A. Introduction to hyphenated techniques and their applications in pharmacy. *Pharm. Methods*, **2010**, *1*, 2-13. b) Wolfender, J. L., Marti, G., Thomas, A., Bertrand, S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromatogr. A.* **2015**, *1382*, 136-164. c) Sarker, S. D., Nahar, L. Hyphenated techniques and their applications in natural products analysis. *Methods Mol. Biol.* **2012**, *864*, 301-340. d) Zhang, A., Sun, H., Wang, P., Han, Y., Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst*, **2012**, *137*, 293-300. e) Sarker, S. D., Nahar, L. Hyphenated techniques, in *Methods in Biotechnology*, Vol. 20, Natural Products Isolation, ed. S. D. Sarker, Z. Latif, and A. I. Gray, 2nd ed., Humana Press, Ottowa, New Jersey, **2005**, pp. 233-267.
44. Dunn, W. B., Ellis, D. I. Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem.* **2005**, *24*, 285-294.
45. a) Kim, H. K., Choi, Y. H, Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549. b) Kim, H. K., Choi, Y. H., Verpoorte, R. NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol.* **2011**, *29*, 267-275.

46. Allwood, J. W., Ellis, D. I., Goodacre, R. Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiol. Plant.* **2008**, *132*, 117-135.
47. Jorge, T. F., Rodrigues, J. A., Caldana, C., Schmidt, R., van Dongen, J. T., Thomas-Oates, J., António, C. Mass spectrometry-based plant metabolomics: metabolite responses to abiotic stress. *Mass Spectrom. Rev.* **2016**, *35*, 620-649.
48. De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., Hall, R. D. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2007**, *2*, 778-791.
49. Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **2006***, 1*, 387-396.
50. Ramautar, R., Somsen, G. W., de Jong, G. J. CE-MS for metabolomics: developments and applications in the period 2012-2014. *Electrophoresis*, **2015**, *36*, 212-224.
51. Junot, C., Fenaille, F., Colsch, B., Bécher, F. High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom. Rev.* **2014**, *33*, 471-500.
52. Dunn, W. B., Wilson, I. D., Nicholls, A. W. Broadhurst, D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, **2012**, *4*, 2249-2264.
53. Kim, H. K., Verpoorte, R. Sample preparation for plant metabolomics. *Phytochem. Anal.* **2010**, *21*, 4-13.
54. Bijttebier, S., Van der Auwera, A., Foubert, K., Voorspoels, S., Pieters, L., Apers, S. Bridging the gap between comprehensive extraction protocols in plant metabolomics studies and method validation. *Anal. Chim. Acta*, **2016**, *935*, 136-150.
55. Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2012**, *403*, 1523-1548.
56. Salem, M. A., Jüppner, J., Bajdzienko, K., Giavalisco, P. Protocol: a fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample. *Plant Methods,* **2016**, *12*, 1-15.
57. Hyötyläinen, T. Sample collection, storage and preparation, in *Chromatographic Methods in Metabolomics*, T. Hyötyläinen and S. Wiedmer, eds. RCS, **2013**, pp. 11-42.
58. Bucar, F., Wube, A., Schmid, M. Natural product isolation – how to get from biological material to pure compounds. *Nat. Prod. Rep.* **2013**, *30*, 525-545.
59. Koulman, A., Tapper, B. A., Fraser K., Cao, M., Lane, G. A., Rasmussen, S. High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics. *Rapid. Commun. Mass Spectrom.* **2007**, *21*, 421-428.

60. Kirwan, J. A., Weber, R. J. M., Broadhurst, D. I., Viant, M. R. Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Scientific Data*, **2014**, *1*, 1-13.
61. Lei, Z., Huhman, D. V., Sumner, L. W. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **2011**, *286*, 25435-25442.
62. Allwood, J. W., Goodacre, R. An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochem. Anal.* **2010**, *21*, 33-47.
63. Villas-Bôas, S. G., Mas, S., Åkesson, M., Smedsgaard, J., Nielsen, J. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.* **2005**, *24*, 613-646.
64. Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **2004**, *22*, 245-252.
65. Goodacre, R., Broadhurst, D., Smilde, A. K., Kristal, B. S., Baker, J. D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A., Ebbels, T., Kell, D. B., Manetti, C., Newton, J., Paternostro, G., Somorjai, R., Sjöström, M., Trygg, J., Wulfert, F. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, **2007**, *3*, 231-241.
66. The handbook of plant metabolomics in *Molecular Plant Biology Handbook*. Weckwerth, W., Kahl, G., eds. Wiley-Blackwell, **2013**.
67. Esbensen, K. H., Guyot, D., Westad, F., Houmoller, L. P. *Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design.* Camo Software, **2004**.
68. Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J., Viant, M. R. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, **2013**, *9*, 44-66.
69. Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., Thaden, J. J., Viant, M. R. Proposed minimum reporting standards for chemical analysis. *Metabolomics*, **2007**, *3*, 211-221.
70. a) http://www.genome.jp/kegg (*Last access: 11/03/2018*). b) https://pubchem.ncbi.nlm.nih.gov (*Last access: 11/03/2018*). c) http://www.chemspider.com (*Last access: 11/03/2018*).
71. Rasche, F., Svatoš, A., Maddula, R. K., Böttcher, C., Böcker, S. Computing fragmentation trees from tandem mass spectrometry data. *Anal. Chem.* **2011**, *83*, 1243-1251.
72. http://www.hmdb.ca/ (*Last access: 11/03/2018*).
73. https://metlin.scripps.edu (*Last access: 11/03/2018*).
74. http://www.massbank.jp (*Last access: 11/03/2018*).
75. Gaetano, A., Lucci, J., Mattoli, L., Mercati, V., Murgia, V., Ragazzi, E., Tisato, F., Traldi, P. *Complex systems in pharmacology: an evolutionary perspective*. V. Mercati and P. Traldi, Aboca, **2015**.

76. a) Mattoli, L., Cangi, F., Ghiara, C., Burico, M., Maidecchi, A., Bianchi, E., Ragazzi, E., Bellotto, L., Seraglia, R., Traldi, P. A metabolite fingerprinting for the characterization of commercial botanical dietary supplements. *Metabolomics*, **2011**, *7*, 437-445. b) Mattoli, L., Cangi, F., Maidecchi, A., Ghiara, C., Ragazzi, E., Tubaro, M., Stella, L., Tisato, F., Traldi, P. Metabolomic fingerprinting of plant extracts. *J. Mass Spectrom.* **2006**, *41*, 1534-1545.
77. a) Mercati, V., Capone, L., Giovagnoni, E. Composition for use in the treatment of persistent cough. *Patent* (Aboca S.p.a. Società Agricola) **2016**, *US20160331794A1*. b) Canciani, M., Murgia, V., Caimmi, D., Anapurapu, S., Licari, A., Marseglia, G. L. Efficacy of Grintuss® pediatric syrup in treating cough in children: a randomized, multicenter, double blind, placebo-controlled clinical trial. *Ital. J. Pediatr.* **2014**, *40*, 56.
78. Naz, S., Vallejo, M., García, A., Barbas, C. Method validation strategies involved in non-targeted metabolomics. *J. Chromatogr. A*, **2014**, *1353*, 99-105.
79. Sangster, T., Major, H., Plumb, R., Wilson, A. J., Wilson, I. D. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis**.** *Analyst*, **2006**, *131*, 1075-1078.
80. Pereira, H., Martin, J., Joly, C., Sébédio, J., Pujos-Guillot, E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*, **2010**, *6*, 207-218.
81. Trygg, J., Holmes, E., Lundstedt, T. Chemometrics in metabonomics. *J. Proteome Res.* **2007**, *6*, 469-479.
82. Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J. C., Sansone, S. The metabolomics standards initiative (MSI). *Metabolomics*, **2007**, *3*, 175-178.
83. Kirwan, J. A., Broadhurst, D. I., Davidson, R. L., Viant, M. R. Characterizing and correcting batch variation in automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal. Bioanal. Chem.* **2013**, *405*, 5147-5157.
84. Jackson, J. E. *A users guide to principal components*, John Wiley, New York, **1991**.
85. a) MacGregor, J. F., Kourti, T. Statistical process control of multivariate processes. *Control Eng. Pract.* **1995**, *3*, 403-414. b) Wikström, C., Albano, C., Eriksson, L., Fridén, H., Johansson, E., Nordahl, Å., Rännar, S., Sandberg, M., Kettaneh-Wold, N., Wold, S. Multivariate process and quality monitoring applied to an electrolysis process – part II. *Chemom. Intell. Lab. Syst.* **1998**, *42*, 233-240. c) Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., Wold, S. Multi- and megavariate data analysis, part II, MKS Umetrics AB, Umeå, **2006**.

86. Wu, H., Guo, J., Chen, S., Liu, X., Zhou, Y., Zhang, X., Xu, X. Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography-mass spectrometry. *J. Pharm. Biomed. Anal.* **2013**, *72*, 267-291.
87. Kind, T., Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2010**, *2*, 23-60.
88. Halket, J. M., Waterman, D., Przyborowska, A. M., Patel, R. K. P., Fraser, P. D., Bramley, P. M. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2005**, *56*, 219-243.
89. http://www.waters.com/webassets/cms/library/docs/720004675en_1.6.pdf (*Last access: 11/03/2018*).