Drug Discovery Today

DRUG DISCOVERY TODAY

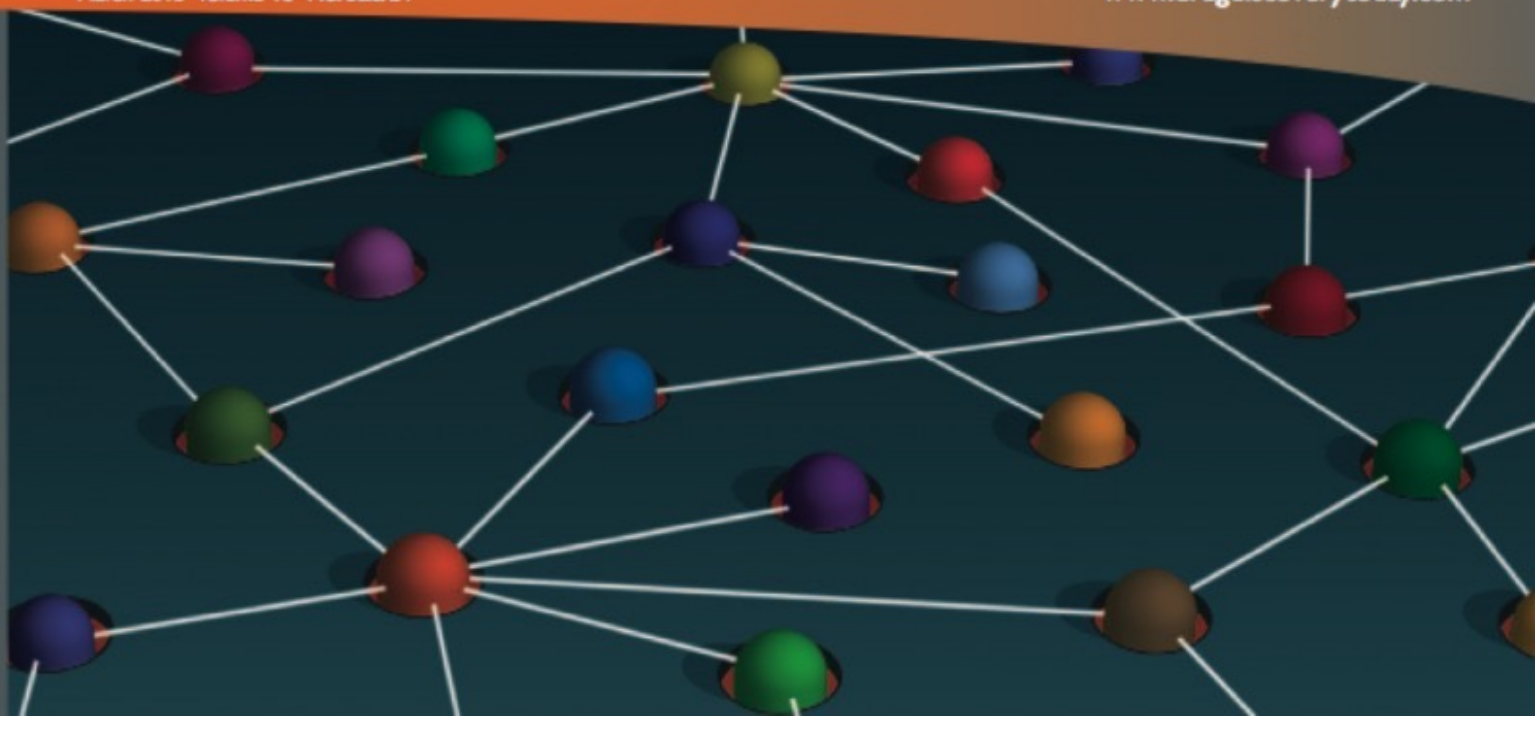
Authoritative reviews and expert opinion
Industry focused, peer reviewed

March 2013 Volume 18 No. 5&6/24 ISSN 1359-6446

- Social networks, web-based tools and diseases
- 3D cell culture: the missing link in drug discovery
- Biological network inference for drug discovery
- Expanding medicinal chemistry space

March 2013 Volume 18 No. 5&6/24

www.drugdiscoverytoday.com



Selected articles from this journal and other medical research on **Novel Coronavirus (2019-nCoV)** and related viruses are now available for free on ScienceDirect – start exploring directly or visit the Elsevier Novel Coronavirus Information Center

Drug Discovery Today

Supports *open access*

7.08
CiteScore

6.88
Impact Factor

Editor: Stephen Carney
View editorial board

View aims and scope

Explore journal content

- Latest issue
- Articles in press
- Article collections
- All issues

Sign in to set up alerts

RSS | *open access* RSS

Latest issues

Volume 25, Issue 2
pp. 259–466 (February 2020)

Volume 25, Issue 1
pp. 1–258 (January 2020)

Volume 24, Issue 12
pp. 2225–2350 (December 2019)

Volume 24, Issue 11
pp. 2105–2224 (November 2019)

View all issues

Find out more

- Submit your article
- Guide for authors
- About the journal

Search in this journal

Looking for an author or a specific volume/issue? Use advanced search

Articles

Articles in press | Latest published | Top cited | Most popular

Review article ○ Abstract only

Recent progress in antitumor functions of the intracellular antibodies

Yunfeng Lin, ... Lei Zhang

In Press, Journal Pre-proof, Available online 26 February 2020

Purchase PDF

Review article ○ Abstract only

Solute carrier proteins and c-Myc: a strong connection in cancer progression

Suman Panda, ... Subhrangsu Chatterjee

In Press, Uncorrected Proof, Available online 24 February 2020

Purchase PDF

Review article ○ Abstract only

Structure-based view of the druggable genome

Jiayan Wang, ... Matthieu Schapira

In Press, Corrected Proof, Available online 19 February 2020

Purchase PDF

Review article ○ Abstract only

Impact of the Protein Data Bank on antineoplastic approvals

John D. Westbrook, ... Stephen K. Burley

In Press, Corrected Proof, Available online 14 February 2020

Purchase PDF

Review article ● *Open access*

The quest for the holy grail: new antitubercular chemical entities, targets and strategies

Stanislav Huszár, ... Vinayak Singh

In Press, Corrected Proof, Available online 13 February 2020

Download PDF

Review article ○ Abstract only

Zeolites in drug delivery: progress, challenges and opportunities

Morteza Servatan, ... Masoud Mozafari

In Press, Uncorrected Proof, Available online 13 February 2020

Purchase PDF

Review article ○ Abstract only

Developing polyphenolic acetates as radiation countermeasure agents: current status and future perspectives

Kavya Venkateswaran, ... Bilikere S. Dwarakanath

In Press, Corrected Proof, Available online 13 February 2020

Purchase PDF

Mini review ○ Abstract only

Is cannabidiol a scheduled controlled substance? Origin makes the difference

Cinzia Citti, ... Giuseppe Cannazza

In Press, Corrected Proof, Available online 13 February 2020

Purchase PDF

More articles in press

Article collections

SI:Therapeutic Nanomaterials
Edited by Feng Xu
September 2017

SI:Tuberculosis
March 2017

Pathophysiology models for drug discovery
Edited by Dimitrios Zeugolis
September 2016

SI: Precision medicine
December 2015

View all article collections

View all issues

Journal info

Incorporating Drug Discovery Today: TARGETS, Drug Discovery Today: BIOSILICO and Pharmaceutical Science & Technology Today;

ISSN

1359-6446

More journal info

Help improve this page

Drug Discovery Today

Supports *open access*

Volume 25, Issue 1

Pages 1-258 (January 2020)

Download full issue

< Previous vol/issue
Next vol/issue >

Editorial

Editorial • Full text access

Learning and augmenting natural processes: potential means of combating antimicrobial resistance from a drug R&D perspective

Charles Oo, Sherwin K.B. Sy

Pages 1-3

Download PDF

Editorial • Full text access

Synthetic biology – reimagine drug discovery

Stephanie Brooking

Pages 4-6

Download PDF

Perspective

Feature

Mini review • Full text access

The importance of target binding kinetics for measuring target binding affinity in drug discovery: a case study from a CRF_1 receptor antagonist program

Sam R.J. Hoare, Beth A. Fleck, John P. Williams, Dimitri E. Grigoriadis

Pages 7-14

Download PDF Article preview

Mini review • Full text access

Dual-peptide ligand masks: a proposed treatment approach to stop prion disease dementias

Kevin Roe

Pages 15-21

Download PDF Article preview

Mini review • Full text access

Mind the gap! A survey of the challenges of biomarker commercialization

Susan Raths, Sven Parkel, Jesper Bredmose, Valérie Daussin

Pages 22-26

Download PDF Article preview

REVIEWS

Foundation

Review article • Full text access

Chemometrics: a complementary tool to guide the isolation of pharmacologically active natural products

Axhell A. Cornejo-Báez, Luis M. Peña-Rodríguez, Radamés Álvarez-Zapata, Maribel Vázquez-Hernández, Alberto Sánchez-Medina

Pages 27-37

Download PDF

Keynote

Review article • Full text access

Understanding the variability of the S1′ pocket to improve matrix metalloproteinase inhibitor selectivity profiles

Aleix Gimeno, Raúl Beltrán-Debón, Miquel Mulero, Gerard Pujadas, Santiago Garcia-Vallvé

Pages 38-57

Download PDF Article preview

Review article • Full text access

A new approach to the diagnosis and treatment of atherosclerosis: the era of the liposome

Nasim Kiaie, Armita Mahdavi Gorabi, Peter E. Penson, Gerald Watts, ... Amirhossein Sahebkar

Pages 58-72

Download PDF Article preview

Review article • Full text access

Design strategies for programmable oligonucleotide nanotherapeutics

Fitsum Feleke Sahle, Tao L. Lowe

Pages 73-88

Download PDF Article preview

Review article • Full text access

Neurovascular glucocorticoid receptors and glucocorticoids: implications in health, neurological disorders and drug therapy

Sherice Williams, Chaitali Ghosh

Pages 89-106

Download PDF Article preview

Review article • Full text access

Clinical applications of nanomedicine in cancer therapy

Mohammad Norouzi, Mehrnaz Amerian, Mahshid Amerian, Fatemeh Atyabi

Pages 107-125

Download PDF Article preview

Gene to Screen

Review article • Full text access

Validation of the epigenetic reader bromodomain-containing protein 4 (BRD4) as a therapeutic target for treatment of airway remodeling

Allan R. Brasier, Jia Zhou

Pages 126-132

Download PDF Article preview

Review article • Full text access

Activity-based proteomic profiling: application of releasable linker in photoaffinity probes

Jin Wang, Qinhua Chen, Yuanyuan Shan, Xiaoyan Pan, Jie Zhang

Pages 133-140

Download PDF Article preview

Post Screen

Review article • Full text access

Iron oxide nanoparticles for therapeutic applications

Edouard Alphandéry

Pages 141-149

Download PDF Article preview

Review article • Full text access

Physicochemical properties affecting the fate of nanoparticles in pulmonary drug delivery

Qiaoyu Liu, Jian Guan, Lu Qin, Xin Zhang, Shirui Mao

Pages 150-159

Download PDF Article preview

Review article • Full text access

Heterogeneous surface-modified nanoplatforms for the targeted therapy of haematological malignancies

Sanjay Kulkarni, Abhijeet Pandey, Srinivas Mutalik

Pages 160-167

Download PDF Article preview

Review article • Full text access

Site-specific bioconjugation and self-assembly technologies for multi-functional biologics: on the road to the clinic

Sung In Lim

Pages 168-176

Download PDF Article preview

Review article • Full text access

Emergence of allosteric drug-resistance mutations: new challenges for allosteric drug discovery

Shaoyong Lu, Yuran Qiu, Duan Ni, Xinheng He, ... Jian Zhang

Pages 177-184

Download PDF Article preview

Review article • Full text access

Nose-to-brain delivery of antiglioblastoma drugs embedded into lipid nanocarrier systems: status quo and outlook

Fakhara Sabir, Ruba Ismail, Ildiko Csoka

Pages 185-194

Download PDF Article preview

Review article • Full text access

Decoy oligodeoxynucleotide technology: an emerging paradigm for breast cancer treatment

Maryam Mahjoubin Tehran, Samaneh Rezaei, Amin Jalili, Seyed Hamid Aghaee-Bakhtiari, Amirhossein Sahebkar

Pages 195-200

Download PDF Article preview

Review article • *Open access*

Drug-induced liver injury severity and toxicity (DILIst): binary classification of 1279 drugs by human hepatotoxicity

Shraddha Thakkar, Ting Li, Zhichao Liu, Leihong Wu, ... Weida Tong

Pages 201-208

Download PDF Article preview

Review article • Full text access

Enhancing the potential preclinical and clinical benefits of quercetin through novel drug delivery systems

Rubiya Khursheed, Sachin Kumar Singh, Sheetu Wadhwa, Monica Gulati, Ankit Awasthi

Pages 209-222

Download PDF Article preview

Review article • Full text access

Immune checkpoint inhibitors: a promising anticancer therapy

Sima Singh, Daniel Hassan, Hibah M. Aldawsari, Nagashekhara Molugulu, ... Prashant Kesharwani

Pages 223-229

Download PDF Article preview

Review article • Full text access

Bioactive modulators targeting STING adaptor in cGAS-STING pathway

Xi Feng, Dongyu Liu, Zhiyu Li, Jinlei Bian

Pages 230-237

Download PDF Article preview

Review article • Full text access

Antimicrobial peptides as novel therapeutics for non-small cell lung cancer

Nitesh K. Kunda

Pages 238-247

Download PDF Article preview

Review article • Full text access

A drug-likeness toolbox facilitates ADMET study in drug discovery

Chen-Yang Jia, Jing-Yi Li, Ge-Fei Hao, Guang-Fu Yang

Pages 248-258

Download PDF Article preview

Previous vol/issue

Next vol/issue

ISSN: 1359-6446

Teaser Chemometrics offers an important complementary tool to enhance the searching and isolation of bioactive natural products from natural sources.

Chemometrics: a complementary tool to guide the isolation of pharmacologically active natural products

Axhell A. Cornejo-Báez[1,4], Luis M. Peña-Rodríguez[2], Radamés Álvarez-Zapata[3], Maribel Vázquez-Hernández[1] and Alberto Sánchez-Medina[1]

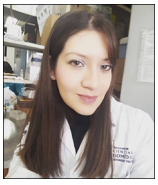
[1] Unidad de Servicios de Apoyo en Resolución Analítica, Universidad Veracruzana. Av. Luis Castelazo Ayala s/n, Col. Industrial Animas, C.P. 91190, Xalapa-Enríquez, Veracruz, México
[2] Centro de Investigación Científica de Yucatán, A.C. Calle 43 No. 130 Chuburná de Hidalgo, C.P. 97205, Mérida, Yucatán, México
[3] Centro de Estudios de Desarrollo Sustentable y Aprovechamiento de la Vida Silvestre (CEDESU) de la Universidad Autónoma de Campeche. Avenida Héroe de Nacozari #480, C.P. 24079, San Francisco de Campeche, Campeche, México
[4] Centro de Investigaciones Biomédicas, Universidad Veracruzana. Av. Luis Castelazo Ayala s/n, Col. Industrial Animas, C.P. 91190, Xalapa Enríquez, Veracruz, México

Natural products (NPs) are specialized metabolites from natural sources, such as plants, animals, fungi, and bacteria, that have had an important role in the discovery and development of new drugs. For many years, bioassay-guided isolation has been the most used strategy for the isolation of bioactive NPs. Chemometrics is the science of relating the measurements made in a system or chemical process to the application of mathematical and statistical methods. Recently, chemometric techniques have been widely used in NP studies, including the detection of bioactive NPs from medicinal plants, chemotaxonomy, determination of geographical origin of medicinal plants and herbal products, and for quality control purposes. In this review, we propose a chemometric-enhanced strategy as a new approach for speeding up the isolation of bioactive NPs while reducing the quantities of biological material and organic solvents used.

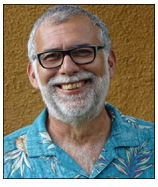
Introduction

The development of new drugs is a complex, slow, and expensive process. The time from the discovery of a potential drug until it reaches the market can be ~12 years or more, with an investment of ~ US$2870 million [1]. Discovery of a new lead compound involves the identification of new chemical entities (NCEs) that have the characteristics required for a drug. These entities can be obtained by chemical synthesis or by an isolation process of NPs. NPs are

Corresponding author: Sánchez-Medina, A. (albsanchez@uv.mx)



Axhell A. Cornejo-BÃez is currently a final-year PhD student in biomedical sciences. She also holds a MPhil in bioorganic chemistry and a BSc in clinical chemistry from the University of Veracruz. Her PhD focuses on the use of chemometrics to search for natural products from medicinal plants with antibacterial and antimycobacterial activity.



Luis M. Peþa-RodrÚguez is a senior scientist at the Plant Biotechnology Unit of Centro de Investigacœn CientÚfica de YucatÃn. He has a BSc degree in industrial chemistry from the University of Tamaulipas (Mexico) and a PhD in chemistry from the University of Alberta (Canada). He has carried out postdoctoral and sabbatical research stays at North Carolina State University (USA), University of New Brunswick (Canada), National University in Mexico City (MÕxico), and Technical University in Munich (Germany).

Radames Alvarez-Zapata received his BSc in biochemical engineering from the Technological Institute of MÕrida and his MSc in biotechnology from the Center for Scientific Research in YucatÃn. His current research centers on the characterization of biomolecules using chemometrics for the detection of molecules of technological interest from plants and microorganisms.

Maribel VÃzquez-HernÃndez is a senior researcher at the University of Veracruz. She was awarded a BSc in chemistry from the University of Veracruz and a PhD in chemical sciences from the National Polytechnic Institute in Mexico. She is an IUPAC Fellow. Her current research includes the synthesis and structural analysis of organic compounds with biological activity.

https://doi.org/10.1016/j.drudis.2019.09.016

specialized metabolites from natural sources, such as plants, animals, fungi, and bacteria, that have had an important role in the discovery and development of new drugs [2,3]. It is estimated that more than half of pharmaceutical products in clinical use today are derived from NPs [4,5]. The number of NCEs derived from NPs and approved for selling from 1981 to 2014 was 1328, the majority of which were in the anticancer, antibacterial, and antiviral areas. In addition, in 2014, ten of the 44 approved small-molecule drugs were from NPs or directly derived from them, representing 25% of the NCEs approved that year [6].

The WHO defines traditional medicine as the sum total of the knowledge, skill, and practices based on the theories, beliefs, and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health as well as in the prevention, diagnosis, improvement, or treatment of physical and mental illness [7]. Given the biological activity of medicinal plants, they have been used in traditional medicine for many years, and the active ingredients are currently used to develop medicines, dyes, preservatives, and cosmetics, among others [8]. Unlike synthetic drugs, medicinal plants contain a complex mixture of substances. The therapeutic mechanisms of medicinal plant are not well defined; therefore, it is important to identify as many chemical components as possible in medicinal plants to understand and explain their bioactivity [8,9]. Different strategies have been developed to isolate and identify the metabolites responsible for biological activity; one of the most used strategy is bioassay-guided isolation. In this review, we discuss the applications of chemometrics techniques and how they can be used as a complementary tool to enhance the isolation of bioactive NPs and reduce the amount of biological material and organic solvents required, while speeding up the process [10,11].

Alberto Sanchez Medina is a research professor at Universidad Veracruzana. He was awarded a PhD in pharmacognosy/pharmacology by the University of Greenwich and holds a BSc in food chemistry. Previously, he was a senior lecturer in pharmacology/pharmacognosy at the University of East London. His research experience includes the use of chromatographic and spectroscopic-based techniques for the analysis, isolation, and structural elucidation of bioactive natural products combined with the use of chemometrics and metabolomics.

Bioassay-guided isolation as a classical tool for NP studies

For many years, the most used strategy for the isolation of bioactive natural products has been bioassay-guided isolation (Fig. 1) [12]. This strategy involves repetitive fractionation at a preparative scale and the evaluation of biological activity until the isolation of the active compounds is achieved [13].

Using this strategy, numerous bioactive chemical compounds have been isolated from plants, fungi, bacteria, and marine organisms. Recent compounds studied using this strategy include: antibacterial compounds, such as lupeol-3β-*O*-cinnamate, lupeol-3β-*O*-dihydrocinnamate, and phyllacanthon (MIC = 0.5, 0.25 mg ml^{-1}, and 5.0 mg ml^{-1}, respectively); and antioxidants, such as α, γ, and δ-tocopherols and the phytosterol, β-sitosterol, isolated from extracts of *Cnidoscolus quercifolius* Pohl [14,15]. It also includes compounds with antituberculous activity, such as decarin and *N*-isobutyl-(2E,4E)-2,4-tetradecadienamide (MIC = 1.6 μg/ml) isolated from *Zanthoxylum capense* [16], and the chloroform fraction against *Mycobacterium tuberculosis*, *Mycobacterium smegmatis, and Mycobacterium avium* (MIC ranging from 156.25 to 625 μg/ml) obtained from *Chenopodium ambrosioides* [17]; compounds with antiplasmodial activity, such as lamalbide 6,7,8-triacetate (IC_{50} = 16.39 ± 0.43 μg/ml) and its aglycone lamiridosin 6,7,8-triacetate (IC_{50} = 0.44.56 ± 1.12 μg/ml) [18]. Furthermore, compounds with cytotoxic activity, such as 5,7,4′-trihydroxy-3′-methoxyflavone-4′-*O*-β-D-glucoside, 5,7,4′-trihydroxy-3′-methoxy flavone-4′-*O*-(6″-acetyl)-β-D-glucoside, and 5,7,4′-trihydroxy-3′-methoxy flavone from the ethanolic extract of *Coronopus didymus* (IC_{50} = 43.50, 0.63, and 3.67 μM, respectively) [19], and an essential oil from *Chenopodium ambrosioides* also showing cytotoxic activity [20]. Two iridoids with antiproliferative activity were obtained from the dichloromethane fraction of *Euphrasia pectinata*: euphrasin and campsinol (IC_{50} = 0.2 and 1.9 μM, respectively) [21]. This strategy has proved to be successful for guiding the isolation of bioactive compounds from plant extracts. However, this can be enhanced when combined with chemometrics.

Analysis of chemical data using multivariate analysis

For many years, multivariate analysis has been used in many areas, such as informatics, engineering, pharmacy, chemistry, among others, for the reduction of a set of data of interest. Multivariate analysis has also been used to study the taxonomic relationships and allocations of animal species [22]; in pharmacology, it has contributed to the development of new drugs, their pharmaceutical forms, and the manufacture thereof [23,24]. Another area in which multivariate analysis is applied is in chemistry, such as by Dos Santos *et al.*, who evaluated the effect of storage temperature on methane carbon stable isotope analysis [25]. Multivariate analysis when performed using chemical data is called 'chemometrics'.

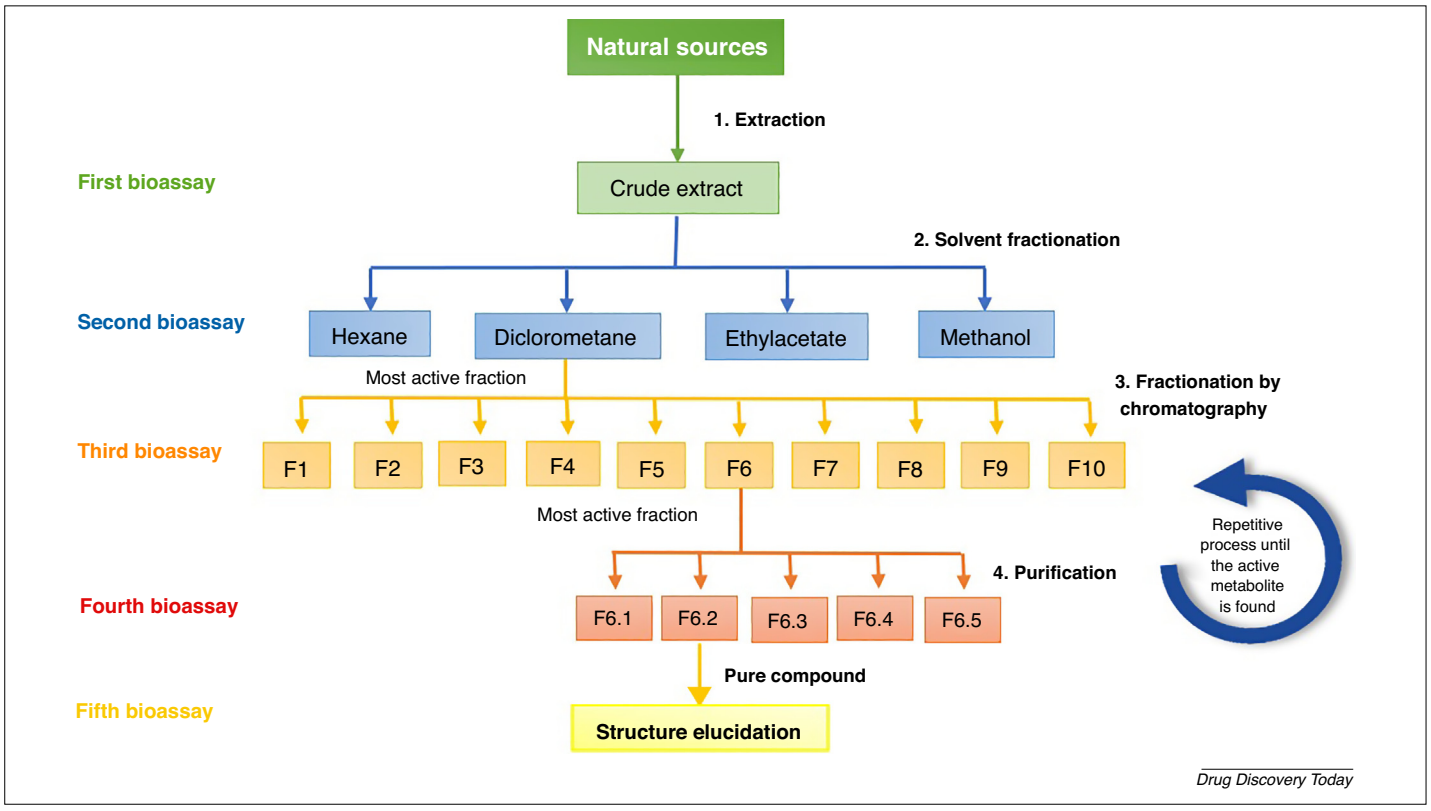


FIGURE 1

Steps involved in a typical bioassay-guided fractionation process. Adapted from Ref. [12].

The term 'chemometrics' was introduced by Svante Wold in 1971. Later, in 1974, together with Bruce Kowalski, he launched the International Chemometrics Society [26], which defined chemometrics as the science of relating the measurements made in a system or chemical process to the state of the system through the application of mathematical and statistical methods [27]. The first paper using the term 'chemometrics' was published by Wold in 1972, and during the 1980s, the first dedicated journals, *Chemometrics and Intelligent Laboratory Systems* and *Journal of Chemometrics*, and specialized software (Arthur, SIMCA, and Unscrambler) appeared. Since then, chemometrics has rapidly developed and has been widely applied in the chemistry and chemical engineering fields [26,28]. Currently, several commercial software programs for multivariate analysis and chemometrics are available, such as AM IX-TOOLS (Bruker BioSpin, GmbH, MA, USA), Matlab with statistics tool x (MathWorks, Natick, MA, USA), Minitab (Minitab, State College, PA, USA), Pirouette (Infometrix, Bothell, WA, USA), SIMCA-P (Umetrics, Umeå, Sweden), SPSS (SPSS, Chicago, IL, USA) or Unscrambler (CAMO software, Woodbridge, NJ, USA) [29]. Additionally, there are many freeware programs for multivariate analysis, such as Chemostat, Metaboanalyst, R (cran) among others, although the latter requires prior programming knowledge [30–32].

In recent articles, the term 'biochemometrics' has been used, which incorporates in a statistical model the chemical data and the biological or pharmacological activity data [33–35]. However, here, we use 'chemometrics', given that, in chemometric tools such as OPLS and other discriminant analysis, a variable Y is required and this can also correspond to any given biological activity. Hence, biochemometrics is encompassed within chemometrics.

Chemometrics provides us with methods to reduce a large amount of data produced by automated instruments, such as high-performance liquid chromatography (HPLC), Gas Chromatography (CG), nuclear magnetic resonance (NMR), and mass spectrometry (MS), among others, and allows us to extract information and identify certain patterns and characteristics of interest. Currently, this discipline has gained importance because of the need to analyze large amounts of information obtained through instrumental equipment and by the increasing capacity of computers [36]. Therefore, to handle the large data set obtained during the chemical analysis of samples and each chromatogram or spectrum obtained comprising hundreds of signals, suitable chemometric methods are required. All the information provided by statistical software should be interpreted with a critical eye, which makes human participation indispensable in the multivariate optimization process [28,37].

There are several techniques that are commonly used for chemometric studies. Emphasis is placed here on principal component analysis (PCA), partial least squares (PLS), and orthogonal projections to latent structures (O-PLS) because these are tools that have shown us the best results in the determination of active compounds. However, we also describe other commonly used chemometrics techniques.

Principal component analysis

PCA is a statistical technique of nonsupervised multivariate analysis that allows the information contained in a set of p loadings of

interest to be sectioned into new independent loadings. The variables are also called 'loadings'. Each explains a specific part of the information and, through a linear combination of the original loadings, it is possible to summarize the information in a few components that reduce the dimension of the problem [38,39]. PCA is used for the investigation of chemical profile analysis in complex mixtures, such as plant extracts. The data analysis procedure requires the preprocessing of the raw data to generate a data matrix, the columns of which represent loadings, whereas the rows contain the samples that are included for analysis [40]. PCA aims to find and interpret complex and casually determined relationships (similarities and differences) between samples. The key to PCA is that it quantifies the importance of the loadings that explain the observed group and the patterns of the inherent properties of individual samples [41].

The spatial distribution of the samples can be observed in a score plot, where similar samples are grouped together and different samples fall further apart [42]. The data for the PCA must be collected in a two-way matrix or matrix, called X, in which the column vectors represent loadings (e.g., absorbances, wavelengths, retention times, chemical shifts, peaks, area under the peaks, toxicity values, or biological responses) and row vectors represent 'objects' for which these loadings are measured, often also called cases, samples, measurements, and so on [42,43].

PCA reduces the number of loadings and allows the visualization of the information included in the matrix. By linear combinations of the original loadings, PCA produces the so-called latent variables or principal components (PC) in a way that they describe the greatest possible variation in X. The score plot gives information regarding the (dis) similarity of the objects, for example, on their tendency to group, whereas the loading plot provides information on the contribution of the original loadings [44].

Projections of partial least squares and orthogonal projections to latent structures

Partial least squares (PLS) projections are a widely used technique with great predictive power, but the interpretation of the regression coefficients of the model can be complicated for complex biological samples because of the modeling of small orthogonal variations in the data matrix X [45]. PLS is a regression technique to model the relationship between a descriptor (X) and a response (Y) (X and Y) that maximizes the square covariance X-Y. PLS is useful not only for the interpretation model, but also for predicting Y of X [46].

Orthogonal projections to latent structures (O-PLS) is an extension of PLS, the objective is to split the systematic variation in the X block into two parts, one that models the correlation between X and Y (predictive) and another that shows the systematic X variation not related (orthogonal) to Y [47,48]. In O-PLS, these orthogonal variations are removed from the data matrix before the construction of a PLS model of a component. This improves the interpretability of the model while preserving its predictive power [45,48]. Both methods include the construction of a regression model maximizing the covariance between the descriptor loadings and the response. The loads of the descriptor loadings in those components specify the origin of this noncorrelation, also called orthogonal variation. The variation correlated with the response can be interpreted by the loads in the predictive component of the O-PLS model. The R^2X values of the prediction and orthogonal components are measures of the structured fraction of the original data variation that describes the response and the fraction uncorrelated with the response. The quality of an O-PLS model is described by the value R^2Y, that is, the correlation between the observed and predicted values for the studied response, and the Q^2 value, that is, the correlation between the observed and the cross-validation of the predicted answer. The higher the value of R^2Y and Q^2, the better the response that can be described and predicted as a function of the descriptor loadings, respectively [49–51]. In addition, O-PLS provides a superior model interpretation by separating the systematic variation contained in the X data matrix into two parts, a predictive part that correlates to Y and an orthogonal part that is not correlated to Y [46].

Other commonly used chemometric tools

Among other commonly used chemometric tools are the linear unsupervised classification methods, such as cluster analysis (CA), an unsupervised pattern detection method that partitions all cases into smaller groups or clusters of relatively similar cases that are dissimilar to other groups. The analysis is based on the correct evaluation of the number of samples in each cluster, cluster location, ranking the variables that best describe the clusters, similarities and dissimilarities among the clusters, and the most significant variables of the clusters. Each cluster is mathematically represented by a component distribution, such as a Gaussian (continuous) or a Poissonian (discrete) distribution, and the entire data set is modeled by a mixture of these distributions [52,53]. Another technique commonly used is hierarchical cluster analysis (HCA), also known as hierarchical clustering. This is a popular method for cluster analysis in big data research and data mining aiming to establish a hierarchy of clusters. HCA uses an algorithm to produce a dendrogram that assembles variables or objects into a single tree. It is usually used as a clustering tool to evaluate intra- and intergroup similarities and differences and, thus, is similar to PCA. There are two types of strategy used in HCA: the agglomerative and the divisive strategies. In the agglomerative strategy, clustering goes from 'the leaves' to 'the roots' of a cluster tree, and this is called a 'bottom-up' approach. Divisive clustering is considered a 'top-down' approach, going from 'the roots' to 'the leaves'. All observations are initially considered as one cluster, and then splits are performed recursively as one moves down the hierarchy [54–56].

One of the most widely used statistical techniques is the multiple linear regression analysis (MLRA), which models the best combination of two or more independent variables to predict or estimate the dependent variable (Y) by fitting a linear equation. A MLRA model is built with descriptive variables using the least squares methods to minimize the residuals, and it shows the contribution of each independent variable to the dependent variable [57–59]. Discriminant data analysis (DA) is a method for analyzing dependence that is a special case of correlation; one of its objectives is to discriminate between two or more groups in terms of the means of the discriminating variables. DA is performed on original data without affecting the results and comparability with other chemometrics methods and constructs a discriminant function for each group [53]. In addition, least discriminant analysis (LDA) is a method of pattern recognition.

It produces a linear classification structure that is able to separate series of the sample within different classes on the basis of their spectral features [60]. Likewise, the discriminant analysis (DA) model can be aligned with PLS and O-PLS; PLS-DA provides an avenue for predicting group membership based on a set of high-dimensional measurements. It tends to construct overly complex models when variation exists in the measurements that do not correlate with membership to an experimental group. In studies of complex mixtures, such as chemometrics, spectral signals having high variability that does not relate to group membership are almost unavoidable. For this reason, O-PLS-DA is often used in lieu of PLS-DA to disentangle group-predictive and group-unrelated variation in the measured data. O-PLS-DA is the preferable model because it has better interpretability compared with PLS-DA. [61,62]. Finally, *artificial neural networks (ANNs)* are a tool for nonparametric data-based modeling that learn nonlinear relations and make a connection between input and output variables. ANN is constructed from neural layers and imitates the learning processes of the human brain. Neurons in neural networks receive inputs and generate outputs through nonlinear operations. By training a network to match a certain set of input data with the desired set of corresponding output data, the process that generated data is modeled [51,63].

Chemometrics applied to NP studies

WHO established fingerprints analysis using chemometrics as a methodology for assessing the chemical variation of commercial medicinal plants. A fingerprint can be defined as a characteristic profile that reflects the complex chemical composition of any analyzed sample, which can be obtained by spectroscopic, chromatographic, or electrophoretic techniques [64]. Several techniques have been proposed for this purpose, including high-performance liquid chromatography (HPLC) combined with mass spectrometry (MS) and/or ultraviolet (UV)-Vis detection spectrophotometry, gas chromatography (GC), nuclear magnetic resonance spectroscopic (NMR), spectrophotometry, near-infrared spectroscopy (NIRS), ultra-HPLC (UHPLC), ion mobility (MS), capillary electrophoresis (CE), paper-based colorimetric assays, and infrared and fluorescence spectroscopy. The HPLC analytical technique is the most popular for the analysis of herbal products and extracts from natural sources. It is an easy-to-use, fully automated technique with high resolution, selectivity, and sensitivity. One of the main advantages of HPLC is the ability to couple the technique to different detectors, the most used of which in HPLC are the UV detector and diode array detection (DAD) for UV absorbent compounds [9,65,66]. By contrast, NMR allows the simultaneous detection of different groups of primary and specialized metabolites. In an NMR spectrum, the signals are proportional to the molar concentration, allowing the comparison of the concentrations of all the compounds in the sample, without needing calibration curves for each individual compound [67,68].

Chemometric techniques have also been applied in the standardization and quality control of herbal medicines [69–71]. Diverse analytical methods are available for the quality standardization and identification of adulterants in herbal and food materials [72]. Given that herbal extracts are complex and often include hundreds or

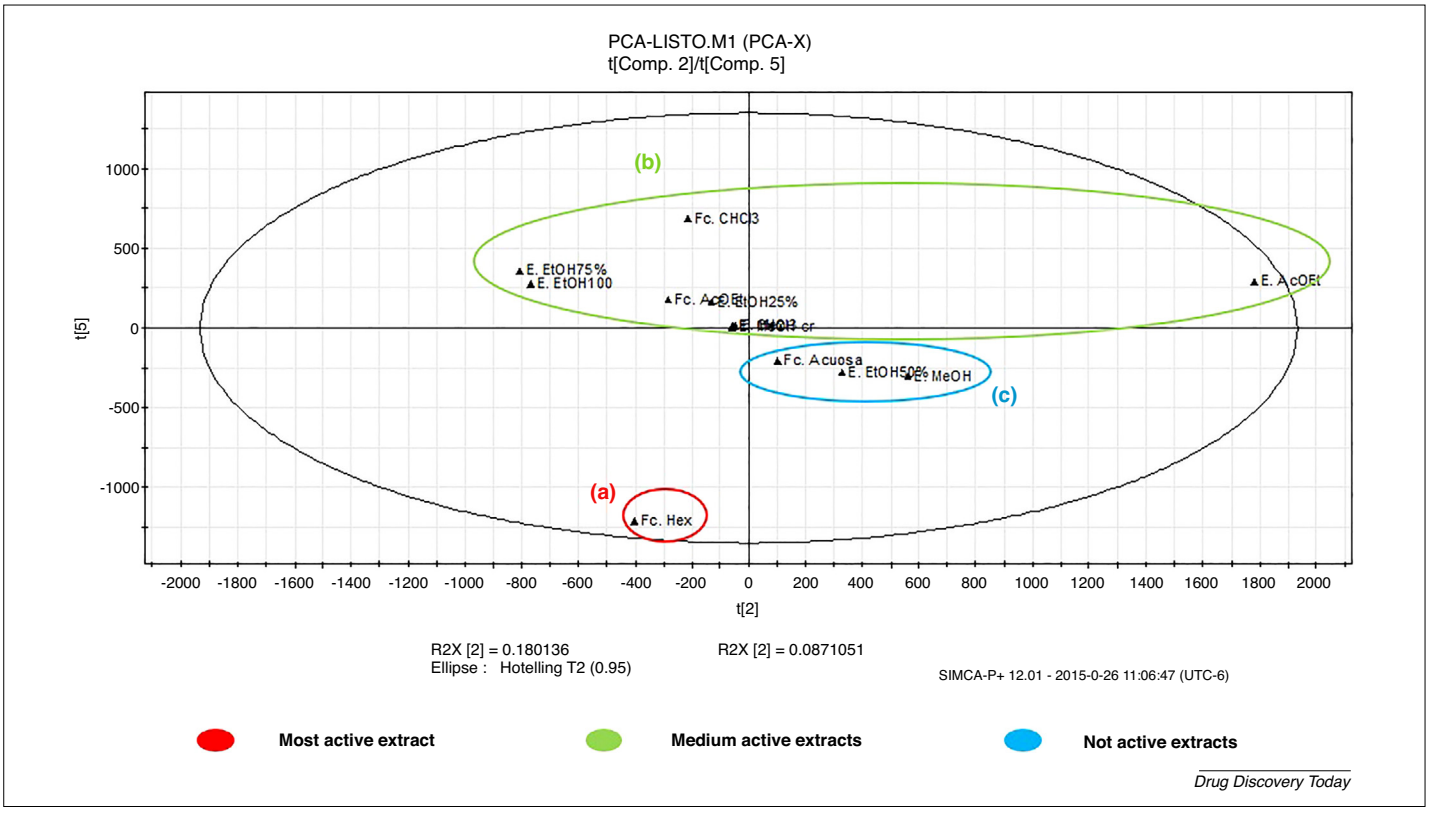


FIGURE 2

Principal component analysis (PCA) of extracts and fractions from a plant. The formation of two clusters can be observed, grouped according to the polarity. This also reflects their similar biological activities [115].

thousands of compounds, chemometrics provides a means for understanding their chemical composition, which is essential for safety risk assessments [73]. Chemometric techniques are also used to evaluate the authenticity of plant materials in herbal remedies and other extracts [74]. They have also been shown to be useful for the production and standardization of plant extracts when a certain concentration of a metabolite of interest is required. Chemometrics is widely used in the taxonomic classification of plants and their different varieties and in their classification according to their geographical origin [27,75,76].

TABLE 1

Frequently chemometric and analytical techniques used in NP studies

Natural product	Chemometric techniques used	Outcome	Analytical approach	Refs
Romanian medicinal plant extracts	HCA, PCA, LDA	Classified medicinal herbs according to their therapeutic effects and usage (asthma, nervous system, digestive system, skin problems and digestive system)	HPTLC	[81]
Extracts from 42 medicinal plants produced by Dacia Plant Manufacturer (Fares Orăștie, Hunedoara, România)	HCA, PCA-LDA	Classification of medicinal plants according to phylum (Pteridophyte, Magnoliophyte, and Spermatophyte)	Spectrophotometer (UV–Vis, 200 and 430 nm)	[82]
Coffee	PCA, PLS-DA	Discrimination of espresso coffees with different sensory characteristics	ATR-FTIR	[83]
Cultivars of grapes	PCA, DA	Phenolic content and antioxidant properties.	UHPLC	[84]
Material plants and herbal supplements	PCA, PLS	Detect adulteration of herbal suplements	IR, HPLC-DAD, UHPLC-MS	[85]
Material plants and herbal supplements	PCA, PLS-DA	Detection of regulated or toxic plants in plant food supplements.	IR (FT-Mid-IR and near-IR)	[86]
Sarcandra glabra	HCA, PCA	Evaluation of similarity and differences of *S. glabra* from different regions	FT-IR, HPSEC, PCD-HPLC	[87]
Natural aged white tea and fresh white tea	HCA, PCA	Effects of rapid aging technology on aroma compounds and characteristic of white tea investigated compared with natural aged white tea and fresh white tea	GC–MS	[88]
White tea	PCA	Characterization of white tea metabolome and comparison with green and black tea	UHPLC-QTOF/MS	[89]
Camellia oil (extracted from seeds of *Camellia oleifera* Abel)	LDA, PLS-DA	Identification and quantification of camellia oil adulteration with other cheaper vegetable oils	Excitation-emission matrix fluorescence spectroscopy	[90]
Myrtus communis L.	PCA	Variability in essential oil composition and antibacterial activity of a population of samples of *M. communis*	GC and GC/MS	[91]
Corydalis yanhusuo W.T. Wang	MLRA	Cardioprotection efficiency of *C. yanhusuo* extracts evaluated	UPLC-UV, Q-TOF MS	[92]
Alchemilla vulgaris	OPLS	Collagenase inhibitory activity	^{1}H NMR	[93]
Eugenia uniflora	HCA, PCA	Evaluation of composition, antioxidant capacity, and cytotoxic activity	GC, GC–MS	[94]
Extracts from 28 Thai plants	PCA, PLS	Antioxidant activity and composition of organic solvent extracts	Spectrophotometer (UV–Vis, 765 nm).	[95]
Extracts from 20 indigenous plants	PCA, OPLS	Antiplasmodial activity	^{1}H NMR	[96]
Andrographis paniculate (Burm. f.) Wall.ex Nees	PCA	Ovicidal and larvicidal activity of extracts from leaves of *A. paniculate* agains tfield isolates of human hookworm	HPLC-MS	[97]
Magnolia biondii Pamp	PCA, PLS-DA	Anti-inflammatory mechanisms of volatile oil in *M. biondii* Pamp on rats	GS-MS	[98]
Glycyrrhetinic acid obtained from *Glycyrrhizae Radix et Rhizoma*	PCA, PLS-DA	Hepatoprotective effect of glycyrrhetinic acid on realgar-induced subchronic hepatotoxicity in mice	^{1}H NMR	[99]
Agrimonia eupatoria L., *Arctium minus* (Hill) Bernh. and *Potentilla reptans* L.	PCA	Evaluation of antimicrobial activity	HPLC	[100]
Annona squamosa Linn	PCA, PLS-DA, OPLS-DA	Evaluation of toxicity	UHPLC-MS	[101]
Gardenia jasminoides fruit	HCA	Evaluation of antioxidant and anti-influenza properties	UFLC-MS	[102]
Bidens pilosa	PCA, OPLS	Modulation of renin–angiotensin system	HPLC-MS	[103]

[a]Abbreviations: ATR-FTIR, attenuated total reflectance Fourier transform spectroscopy; HPSEC, high-performance size-exclusion chromatography; HPTLC, high-performance thin-layer chromatography; ICP-MS, inductively coupled plasma MS; ICP OES, inductively coupled plasma optical emission spectrometry; LC-ESI/LTQOrbitrap/MS, HPLC coupled with an hybrid mass spectrometer, which combines the linear trap quadrupole and OrbiTrap mass analyzer; PCD-HPLC, postcolumn derivatization HPLC; UFLC, ultra-flow liquid chromatography; UHPLC-QTOF/MS, UHPLC coupled with quadrupole time-of-flight MS.

In the analysis of plant extracts, PCA reveals clusters based on the chemical characteristics of each sample, so that chemically different samples appear separated and those chemically similar are grouped together. Given that their biological activities are influenced by the chemical compounds present in them, it is expected that those grouped together will have similar activities (Fig. 2).

Chemometrics has been used to evaluate the variability of the chemical composition of propolis and corn from different regions to determine the main classes of chemical compound and the variability of these in different species of the same genus [41,77,78]. It has also been used in the comparison and classification of essential oils from different plants based on their chemical composition and biological activities [79] and for the authentication of plants [80]. Several studies using a diversity of analytical instruments and chemometrics tools have been reported (Table 1).

Chemometrics as a complementary tool guide the isolation of NPs

Chemometric techniques can be used as a complementary strategy to bioassay-guided isolation and can help reduce the amount of biological material and solvents used for extraction, because they require grams of biological material to carry out a whole study [100]. This can be helpful in contributing to the knowledge and documentation of some plants species with limited geographical distribution or that are endangered. At the same time, they reduce working time and associated costs.

In 2009, Nguyen *et al*. analyzed the fingerprints of several *Mallotus* species by means of HPLC and evaluated their antioxidant capacity using 1,1-diphenyl-2-picrylhydrazyl radical (DPPH) scavenger activity. Using chemometric techniques, such as PCA and PLS, the observed antioxidant activity was correlated with the peaks given in the HPLC fingerprints, revealing the peaks responsible for the antioxidant activity [104]. Tistaert and collaborators then applied and evaluated several multivariate calibration techniques to *Mallotus* fingerprints to search for the best technique that indicates the peaks responsible for antioxidant activity. A data set containing HPLC fingerprints of 39 samples from *Mallotus* was developed and its antioxidant activity evaluated. The authors concluded that the O-PLS was the best-performing technique to indicate the antioxidant compounds in *Mallotus* extracts because of its simplicity and reproducibility [44,105].

By contrast, Gao *et al*. evaluated the antiproliferative activity of extracts from roots of *Scutellaria baicalensis* on the human lung cell line SK-MES-1. From the HPLC and 1H-NMR chemical profiles of the extracts combined with the use of PCA and PLS, the authors identified baicalin, baicalein, and wogonin as the compounds responsible for the cell growth inhibition activity of the extracts [106].

An important concept to consider is the metabolomics, which is used to identify and quantify low-molecular-weight metabolites and to produce profiles of metabolites that reflect the state of a biological system of interest [107,108]. This approach measures changes in the relative concentrations of metabolites as a result of internal or external factors and, thus, can identify complex interactions between biological systems, drugs, and diseases [109,110]. However, metabolomics is based on chemometric techniques, because it makes use of these to understand the variability in biological systems. Overall, the role of 'omics tools is now fully recognized as a powerful strategy for the prioritization of bioactive NPs [111].

An example of this methodology was developed by Yuliana *et al*., who studied the correlation of signals obtained by NMR of

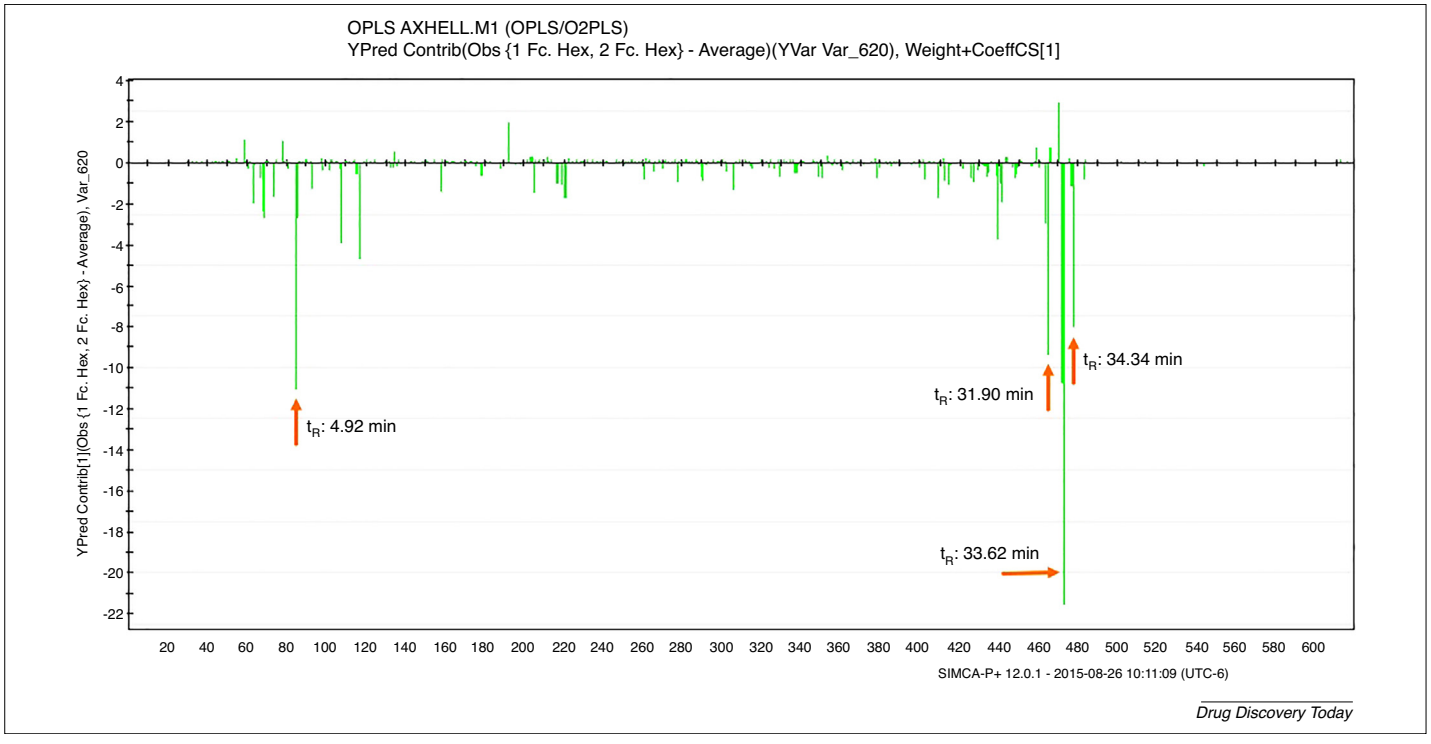


FIGURE 3

Predictive vector plot of the orthogonal projections to latent structures (O-PLS) obtained by high-performance liquid chromatography-ultraviolet (HPLC-UV) analysis of active fractions. The plot indicates the potential bioactive nanoparticles based on their retention time [115].

metabolites present in the *Boesenbergia rotunda* extracts with the aim of quickly identifying compounds that correlated with the activity of A1 receptor binding. It was predicted by O-PLS that two previously reported compounds in this plant, pinocembrin and hydroxypanduratin, were responsible for the A1 adenosine binding activity. In addition, the authors stated that the identification of active compounds from plant extracts using chemometric techniques from ^{1}H-NMR was more efficient than using fractionation guided by bioassays [112].

In 2012, Sharma *et al*. compared the ^{1}H-NMR metabolic profiles of different extracts from *Galphimia glauca*. Following a chemometrics analysis using PLS-DA and data from the evaluation of the anxiolytic, sedative and anti-inflammatory activities of the extracts, they were able to correlate galphimines with the sedative activity of the plant. The results supported the efficacy of ^{1}H-NMR and chemometrics for the determination of bioactive NPs [113]. In addition, Bapela and collaborators evaluated the antiplasmodial activity of extracts from 20 plant species, followed by PCA analysis of the ^{1}H-NMR spectra. The PCA did not show a clustering pattern of the samples according to their biological activity. However, the application of O-PLS to the ^{1}H-NMR profiles resulted in a pattern of discrimination that correlated with biological activity. From these results, two indole alkaloids, tabernaemontanin and dregamine, with antiplasmodial activity (IC_{50} = 12.0 ± 0.8 μM and 62.0 ± 2.4 μM respectively) were identified from *Tabernaemontana elegans*. Furthermore, friedelin (IC_{50} = 7.20 ± 0.5 μM) and morindolide (IC_{50} = 107.1 ± 0.6 μM) were isolated from *Vangueria infausta* subsp. *infausta*. This showed that chemometrics can be used to globally identify the classes of specialized metabolites responsible for the biological activity of plant species [114].

In addition to the use of chemometric techniques for the detection of NPs responsible for the biological activity of extracts from plants, and the construction of predictive models for their biological activity, another interesting, but rarely used application of chemometrics is guiding the isolation of NPs from plants for which no phytochemicals reports are available.

For example, the leishmanicidal activity of the extracts and fractions of *Colubrina greggii* were evaluated, and their chemical fingerprints were acquired using HPLC-UV. The data were analyzed using PCA and O-PLS and the results showed the loadings and retention times (t_R) of the chemical compounds responsible for th leishmanicidal activity. These data were used to guide the final purification by semipreparative HPLC of two components (t_R = 48.7 and 49.5 min, respectively). Further *in vitro* evaluation of these confirmed that those were the compounds responsible for the biological activity. Additionally, these peaks were minor

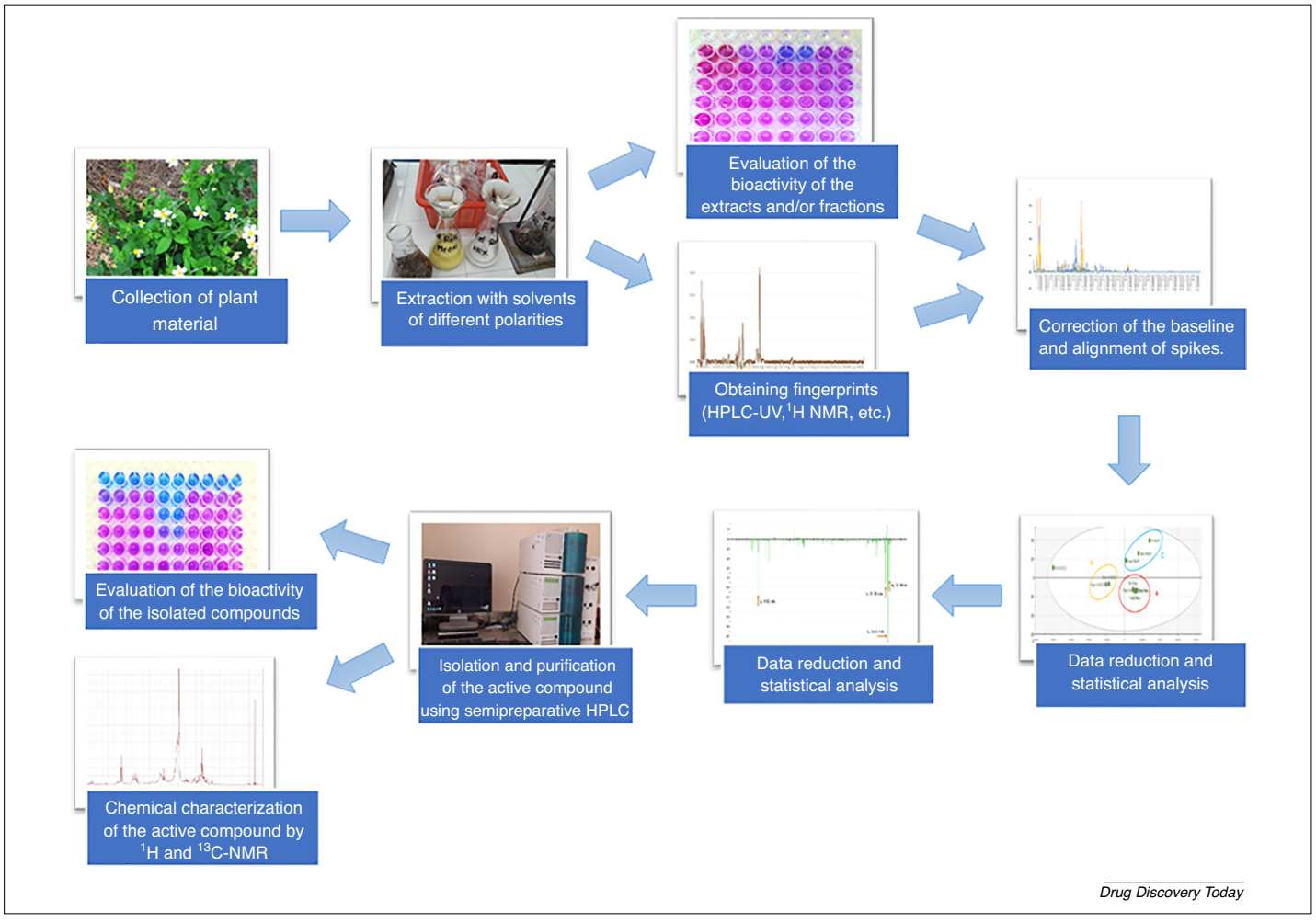


FIGURE 4

Proposed steps for the isolation of active nanoparticles with biological activity based on chemometric techniques.

components present in the plant extracts. Hence, the authors showed evidence of the applicability of chemometrics for enhancing and guiding the isolation of bioactive NPs, even when they are minor components [115] (Fig. 3). Based on this evidence, we propose an experimental design for the isolation of NPs using a chemometric-enhanced strategy. In Fig. 4, we show the steps to follow in a chemometrics-guided isolation of NPs. Additionally, in our experience, we have found that a minimum of 8–10 samples is required to obtain a reliable model. We also recommend to performing the analytical analysis of each sample in duplicate or triplicate. Finally, this experimental design can also be adapted to include fractions from an early chromatographic column of a crude extract.

Concluding remarks

With the ever-increasing impact of humans on the natural environment, it is more important than ever before to collect information and document in detail the chemical constituents of medicinal plants, especially those that are endangered or of limited geographical distribution. These can result in lead molecules for the development of new drugs with different types of biological activity. Chemometric techniques provide the necessary tools to enhance the isolation of bioactive NPs from plants or any other source of NPs. In addition, the isolation of chemometrically enhanced bioactive NPs reduces the quantity of plant material and solvents required, saving both time and resources. When correctly applied, chemometrics can also help to detect bioactive NPs even if they are found in small quantities. Thus, the use of chemometric techniques can become a powerful tool to enhance the detection and isolation of bioactive NPs.

Acknowledgment

A.A.C.B. would like to thank Consejo Nacional de Ciencia y Tecnología (CONACYT) for funding a PhD scholarship with number 560976/297059.

References

1 DiMasi, J.A. *et al.* (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* 47, 20–33
2 Guo, Z. (2017) The modification of natural products for medical use. *Acta Pharm. Sin. B* 7, 119–136
3 Patridge, E. *et al.* (2016) An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today* 21, 204–207
4 Awan, A.R. *et al.* (2016) Biosynthesis of therapeutic natural products using synthetic biology. *Adv. Drug Deliv. Rev.* 105, 96–106
5 Rao, T. *et al.* (2019) The pharmacogenetics of natural products: a pharmacokinetic and pharmacodynamic perspective. *Pharmacol. Res.* 146, 104283
6 Newman, D.J. and Cragg, G.M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661
7 WHO (2013) *WHO Herbal Sale Rates*. WHO
8 Garza-Juárez, A. *et al.* (2011) Correlation between chromatographic fingerprint and antioxidant activity of *Turnera diffusa* (Damiana). *Planta Med.* 77, 958–963
9 Mazina, J. *et al.* (2015) Fluorescence, electrophoretic and chromatographic fingerprints of herbal medicines and their comparative chemometric analysis. *Talanta* 139, 233–246
10 Moreira, S.A. *et al.* (2019) Effect of emergent non-thermal extraction technologies on bioactive individual compounds profile from different plant materials. *Food Res. Int.* 115, 177–190
11 Kvalheim, O.M. *et al.* (2011) Chromatographic profiling and multivariate analysis for screening and quantifying the contributions from individual components to the bioactive signature in natural products. *Chemom. Intell. Lab. Syst.* 107, 98–105
12 Cunha, L.C.S. *et al.* (2017) Bioassay-guided fractionation and antimicrobial and cytotoxic activities of cassia bakeriana extracts. *Braz. J. Pharmacogn.* 27, 91–98
13 Brusotti, G. *et al.* (2014) Isolation and characterization of bioactive compounds from plant resources : the role of analysis in the ethnopharmacological approach. *J. Pharm. Biomed. Anal.* 87, 218–228
14 de Oliveira-Júnior, R.G. *et al.* (2018) Antibacterial activity of terpenoids isolated from *Cnidoscolus quercifolius* Pohl (Euphorbiaceae), a Brazilian medicinal plant from Caatinga biome. *Eur. J. Integr. Med.* 24, 30–34
15 Santos, K.A. *et al.* (2017) Chemical composition, antioxidant activity and thermal analysis of oil extracted from favela (*Cnidoscolus quercifolius*) seeds. *Ind. Crops Prod.* 97, 368–373
16 Luo, X. *et al.* (2013) *Zanthoxylum capense* constituents with antimycobacterial activity against *Mycobacterium tuberculosis in vitro* and ex vivo within human macrophages. *J. Ethnopharmacol.* 146, 417–422
17 Jesus, R.S. *et al.* (2018) *In vitro* antimicrobial and antimycobacterial activity and HPLC–DAD screening of phenolics from *Chenopodium ambrosioides* L. *Braz. J. Microbiol.* 49, 296–302
18 Tshisekedi Tshibangu, P. *et al.* (2017) Antiplasmodial activity of *Heinsia crinita* (Rubiaceae) and identification of new iridoids. *J. Ethnopharmacol.* 196, 261–266
19 Noreen, H. *et al.* (2016) Bioassay-guided isolation of cytotoxic flavonoids from aerial parts of *Coronopus didymus*. *J. Ethnopharmacol.* 194, 971–980
20 Degenhardt, R.T. *et al.* (2016) Characterization and evaluation of the cytotoxic potential of the essential oil of *Chenopodium ambrosioides*. *Braz. J. Pharmacogn.* 26, 56–61
21 Marcotullio, M.C. *et al.* (2015) Bioassay-guided fractionation of *Euphrasia pectinata* Ten. and isolation of iridoids with antiproliferative activity. *Phytochem. Lett.* 12, 252–256
22 Tuttle, K.L. (2018) PCA and DFA as multivariate tools to resolve ambiguous species relationships in museum collections. *Proc. Biol. Soc. Washingt.* 131, 132–143
23 Kubova, K. *et al.* (2017) The influence of thermal treatment and type of insoluble poly(meth)acrylates on dissolution behavior of very soluble drug from hypromellose matrix tablets evaluated by multivariate data analysis. *Pharm. Dev. Technol.* 22, 206–217
24 Walicová, V. *et al.* (2017) Design and evaluation of mucoadhesive oral films containing sodium hyaluronate using multivariate data analysis. *Pharm. Dev. Technol.* 22, 229–236
25 dos Santos, V.H.J.M. *et al.* (2018) Evaluation of different samplers and storage temperature effect on the methane carbon stable isotope analysis. *Environ. Monit. Assess.* 190, 573–586
26 Brereton, R.G. (2014) A short history of chemometrics: a personal view. *J. Chemom.* 28, 749–760
27 Bansal, A. *et al.* (2014) Chemometrics: a new scenario in herbal drug standardization. *J. Pharm. Biomed. Anal.* 4, 223–233
28 Li, C.Q. *et al.* (2019) Collaboration patterns and network in chemometrics. *Chemom. Intell. Lab. Syst.* 191, 21–29
29 Kumar, N. *et al.* (2014) Chemometrics tools used in analytical chemistry: an overview. *Talanta* 123, 186–199
30 Helfer, G.A. *et al.* (2015) Chemostat, um software gratuito para análise exploratória de dados multivariados. *Quim. Nova* 38, 575–579
31 Chong, J. *et al.* (2019) MetaboAnalystR 2.0: From raw spectra to biological insights. *Metabolites* 9, 57–66
32 Varmuza, K. and Filzmoser, P. (2009) *Introduction to Multivariate Statistical Analysis in Chemometrics*. pp. 1–336, CRC Press
33 Kellogg, J.J. *et al.* (2016) *Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds*. Publisher
34 Caesar, N.B. *et al.* (2018) Integration of biochemometrics and molecular networking to identify bioactive constituents of ashitaba (*Angelica keiskei* Koidzumi). *Planta Med.* 84, 721–728
35 Britton, N.B. *et al.* (2018) Biochemometrics to identify synergists and additives from botanical medicines: a case study with *Hydrastis canadensis* (Goldenseal). *J. Nat. Prod.* 81, 484–493
36 Hendriks, M.M.W.B. *et al.* (2005) Preprocessing and exploratory analysis of chromatographic profiles of plant extracts. *Anal. Chim. Acta* 545, 53–64
37 Ferreira, S. (2019) *Experimental Design* (3rd edn), Elsevier
38 Charles, A.L. and Alamsjah, M.A. (2019) Application of chemometric techniques: an innovative approach to discriminate two seaweed cultivars by physico-functional properties. *Food Chem.* 289, 269–277

39 Xia, J. *et al.* (2019) Fourier transform infrared spectroscopy and chemometrics for the discrimination of paper relic types. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 219, 8–14
40 Tanaka, K. *et al.* (2008) Phytochemistry quality evaluation of *Astragali Radix* using a multivariate statistical approach. *Phytochemistry* 69, 2081–2087
41 Arumugam, R. *et al.* (2012) Determination of element contents in herbal drugs: chemometric approach. *Food Chem.* 135, 2372–2377
42 Kasote, D. *et al.* (2014) Chemical profiling and chemometric analysis of South African propolis. *Biochem. Syst. Ecol.* 55, 156–163
43 Esbensen, K.H. and Geladi, P. (2009) Principal component analysis: concept, geometrical interpretation, mathematical background, algorithms, history, practice. *Compr. Chemom.* 2, 211–226
44 Tistaert, C. *et al.* (2009) Potential antioxidant compounds in *Mallotus* species fingerprints. Part I: Indication, using linear multivariate calibration techniques. *Anal. Chim. Acta* 652, 189–197
45 Tistaert, C. *et al.* (2011) Chromatographic separation techniques and data handling methods for herbal fingerprints: a review. *Anal. Chim. Acta* 690, 148–161
46 Galindo-Prieto, B. *et al.* (2015) Variable influence on projection (VIP) for OPLS models and its applicability in multivariate time series analysis. *Chemom. Intell. Lab. Syst.* 146, 297–304
47 Souihi, N. *et al.* (2015) OPLS in batch monitoring — opens up new opportunities. *Anal. Chim. Acta* 857, 28–38
48 Ortiz-Herrero, L. *et al.* (2019) OPLS multivariate regression of FTIR-ATR spectra of acrylic paints for age estimation in contemporary artworks. *Talanta* 205, 120114
49 Silva, A.F.T. *et al.* (2017) Statistical process control of cocrystallization processes: A comparison between OPLS and PLS. *Int. J. Pharm.* 520 (1–2), 29–38
50 Souihi, N. *et al.* (2013) A quality by design approach to investigate the effect of mannitol and dicalcium phosphate qualities on roll compaction. *Int. J. Pharm.* 447, 47–61
51 Trygg, J. and Wold, S. (2002) Orthogonal projections to latent structures (O-PLS). *J. Chemom.* 16, 119–128
52 Mutihac, L. and Mutihac, R. (2008) Mining in chemometrics. *Anal. Chim. Acta* 612, 1–18
53 Zhou, F. *et al.* (2007) Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong. *Mar. Pollut. Bull.* 54, 745–756
54 Tomsone, L. *et al.* (2012) The application of hierarchical cluster analysis to clasify genotypes of horseradish (*Armoracia rusticana* L.) roots. *Cheminé Technologija* 62, 52–56
55 Zhang, Z. *et al.* (2017) Hierarchical cluster analysis in clinical research with heterogeneous study population: Highlighting its visualization with R. *Ann. Transl. Med.* 5, 75–85
56 Caesar, L.K. *et al.* (2018) Hierarchical cluster analysis of technical replicates to identify interferents in untargeted mass spectrometry metabolomics. *Anal. Chim. Acta* 1021, 69–77
57 Lee, S.I. *et al.* (2019) Microstructure-based prediction of yield ratio and uniform elongation in high-strength bainitic steels using multiple linear regression analysis. *Mater. Sci. Eng. A* 758, 56–59
58 Kong, W.J. *et al.* (2017) Combination of chemical fingerprint and bioactivity evaluation to explore the antibacterial components of *Salvia miltiorrhizae*. *Sci. Rep.* 7, 1–12
59 Gadurić, S.B. *et al.* (2016) Multivariate chemometrics with regression and classification analyses in heroin profiling based on the chromatographic data. *Iran. J. Pharm. Res.* 15, 725–734
60 Yadav, P.K. and Sharma, R.M. (2019) Classification of illicit liquors based on their geographic origin using Attenuated total reflectance (ATR) – Fourier transform infrared (FT-IR) spectroscopy and chemometrics. *Forensic Sci. Int.* 295, e1–e5
61 Worley, B. and Powers, R. (2016) PCA as a predictor of OPLS-DA model reliability. *Curr. Metabolomics* 4, 97–103
62 Homayoun, S.B. *et al.* (2011) Compared application of the new OPLS-DA statistical model versus partial least squares regression to manage large numbers of variables in an injury case-control study. *Sci. Res. Essays* 6, 4369–4377
63 Yousefi, S. *et al.* (2019) A robust hybrid artificial neural network double frontier data envelopment analysis approach for assessing sustainability of power plants under uncertainty. *Expert Syst.* In press https://doi.org/10.1111/exsy.12435
64 WHO (2018) Guidelines on good manufacturing practices for the. *WHO Tech. Rep. Ser.* 10, 153–178
65 Bijauliya, R.K. *et al.* (2017) A comprehensive review on standardization of herbal drugs. *Int. J. Pharm. Sci. Res.* 8, 3663–3677
66 Masondo, N.A. and Makunga, N.P. (2019) Advancement of analytical techniques in some South African commercialized medicinal plants: current and future perspectives. *South Afr. J. Bot.* 126, 40–57
67 Kim, H.K. *et al.* (2010) NMR-based metabolomic analysis of plants. *Nat. Protoc.* 5, 536–549
68 Alaerts, G. *et al.* (2010) Recent developments in chromatographic fingerprints from herbal products: set-up and data analysis. *Comb. Chem. High Throughput Screen.* 13, 900–922
69 Mendiola, J.A. *et al.* (2010) Design of natural food antioxidant ingredients through a chemometric approach. *J. Agric. Food Chem.* 58, 787–792
70 Sereshti, H. *et al.* (2018) Quality control of saffron and evaluation of potential adulteration by means of thin layer chromatography-image analysis and chemometrics methods. *Food Control* 90, 48–57
71 Różańska, A. *et al.* (2018) Novel analytical method for detection of orange juice adulteration based on ultra-fast gas chromatography. *Monatshefte fur Chemie* 149, 1615–1621
72 Moon, B.C. *et al.* (2018) Establishment of a PCR assay for the detection and discrimination of authentic *Cordyceps* and adulterant species in food and herbal medicines. *Molecules* 23, 1–16
73 Mukherjee, P.K. (2019) Qualitative analysis for evaluation of herbal drugs. In *Quality Control and Evaluation of Herbal Drugs* (Mukherjee, P.K., ed.), pp. 79–149, Elsevier
74 Lucio-Gutiérrez, J.R. *et al.* (2019) Selective 1D-TOCSY and chemometrics to evaluate authenticity of *Turnera diffusa* and related botanical extracts. *Phytochem. Lett.* 30, 62–68
75 Schulze, A.E. *et al.* (2014) Chemometric analysis of chromatographic fingerprints shows potential of *Cyclopia maculata* (Andrews) Kies for production of standardized extracts with high xanthone content. *J. Agric. Food Chem.* 62 (3), 10542–10551
76 Hu, L. *et al.* (2019) Quality evaluation and traceability of Bletilla striata by fluorescence fingerprint coupled with multiway chemometrics analysis. *J. Sci. Food Agric.* 99, 1413–1424
77 Deconinck, E. *et al.* (2017) Chemometrics and chromatographic fingerprints to classify plant food supplements according to the content of regulated plants. *J. Pharm. Biomed. Anal.* 143, 48–55
78 Kuhnen, S. *et al.* (2010) Metabolic fingerprint of Brazilian maize landraces silk (stigma/styles) using NMR spectroscopy and chemometric methods. *J. Agric. Food Chem.* 58, 2194–2200
79 Semeniuc, C.A. *et al.* (2018) Chemometric comparison and classification of some essential oils extracted from plants belonging to Apiaceae and Lamiaceae families based on their chemical composition and biological activities. *Molecules* 23, 2261–2273
80 Hidalgo, M.J. *et al.* (2018) Classification of organic olives based on chemometric analysis of elemental data. *Microchem. J.* 142, 30–35
81 Simion, I.M. *et al.* (2019) Classification of Romanian medicinal plant extracts according to the therapeutic effects using thin layer chromatography and robust chemometrics. *J. Pharm. Biomed. Anal.* 163, 137–143
82 Simion, I.M. and Sârbu, C. (2019) The impact of the order of derivative spectra on the performance of pattern recognition methods. Classification of medicinal plants according to the phylum. *Spectrochim. Acta A* 219, 91–95
83 Belchior, V. *et al.* (2019) Attenuated total reflectance fourier transform spectroscopy (ATR-FTIR) and chemometrics for discrimination of espresso coffees with different sensory characteristics. *Food Chem.* 273, 178–185
84 Sridhar, K. and Charles, A.L. (2018) Application of multivariate statistical techniques to assess the phenolic compounds and the *in vitro* antioxidant activity of commercial grape cultivars. *J. Chemom.* 32, 1–13
85 Deconinck, E. *et al.* (2019) A strategy based on fingerprinting and chemometrics for the detection of regulated plants in plant food supplements from the Belgian market: two case studies. *J. Pharm. Biomed. Anal.* 166, 189–196
86 Deconinck, E. *et al.* (2017) Detection of regulated herbs and plants in plant food supplements and traditional medicines using infrared spectroscopy. *J. Pharm. Biomed. Anal.* 142, 210–217
87 Li, H. *et al.* (2019) Multiple fingerprint profiles and chemometrics analysis of polysaccharides from *Sarcandra glabra*. *Int. J. Biol. Macromol.* 123, 957–967
88 Qi, D. *et al.* (2018) Study on the effects of rapid aging technology on the aroma quality of white tea using GC–MS combined with chemometrics: in comparison with natural aged and fresh white tea. *Food Chem.* 265, 189–199
89 Dai, W. *et al.* (2017) Characterization of white tea metabolome: comparison against green and black tea by a nontargeted metabolomics approach. *Food Res. Int.* 96, 40–45
90 Wang, T. *et al.* (2019) Rapid identification and quantification of cheaper vegetable oil adulteration in camellia oil by using excitation-emission matrix fluorescence spectroscopy combined with chemometrics. *Food Chem.* 293, 348–357
91 Fadil, M. *et al.* (2017) Intrapopulation variability of *Myrtus communis* L. growing in Morocco: chemometric investigation and antibacterial activity. *J. Appl. Res. Med. Aromat. Plants* 7, 35–40
92 Li, Q. *et al.* (2017) Fingerprint-efficacy study of the quaternary alkaloids in *Corydalis yanhusuo*. *J. Ethnopharmacol.* 207, 108–117

93 Mandrone, M. *et al.* (2018) Identification of a collagenase-inhibiting flavonoid from *Alchemilla vulgaris* using NMR-based metabolomics. *Planta Med.* 84 (–13), 941–946
94 Figueiredo, P.L.B. *et al.* (2019) Composition, antioxidant capacity and cytotoxic activity of *Eugenia uniflora* L. chemotype-oils from the Amazon. *J. Ethnopharmacol.* 232, 30–38
95 Maisuthisakul, P. *et al.* (2008) Relationship between antioxidant properties and chemical composition of some Thai plants. *J. Food Compos. Anal.* 21, 229–240
96 Bapela, M.J. *et al.* (2015) Isolation of antiplasmodial compounds from traditionally used plants by means of NMR-based metabolomics and chromatography. *South Afr. J. Bot.* 98, 170–171
97 Banerjee, T. *et al.* (2019) Ovicidal and larvicidal effects of extracts from leaves of Andrographis paniculata (Burm. f.) Wall.ex Nees against field isolates of human hookworm (*Ancylostoma duodenale*). *J. Ethnopharmacol.* 235, 489–500
98 Qi, Y. *et al.* (2011) A plasma metabonomic investigation into the intervention of volatile oil of *Magnolia biondii* Pamp on rat model of acute inflammation. *J. Ethnopharmacol.* 137, 487–494
99 Huo, T. *et al.* (2017) Plasma metabolomics study of the hepatoprotective effect of glycyrrhetinic acid on realgar-induced sub-chronic hepatotoxicity in mice via 1H NMR analysis. *J. Ethnopharmacol.* 208, 36–43
100 Watkins, F. *et al.* (2012) Antimicrobial assays of three native British plants used in Anglo-Saxon medicine for wound healing formulations in 10th century England. *J. Ethnopharmacol.* 144, 408–415
101 Miao, Y.J. *et al.* (2016) Metabolomics study on the toxicity of *Annona squamosa* by ultraperformance liquid-chromatography high-definition mass spectrometry coupled with pattern recognition approach and metabolic pathways analysis. *J. Ethnopharmacol.* 184, 187–195
102 Shan, M.Q. *et al.* (2019) Comparative analysis of sixteen active compounds and antioxidant and anti-influenza properties of *Gardenia jasminoides* fruits at different times and application to the determination of the appropriate harvest period with hierarchical cluster analysis. *J. Ethnopharmacol.* 233, 169–178
103 Chen, Y. *et al.* (2016) Identification of active substances for dually modulating the renin-angiotensin system in *Bidens pilosa* by liquid chromatography-mass spectrometry-based chemometrics. *J. Funct. Foods* 21, 201–211
104 Nguyen Hoai, N. *et al.* (2009) Development of HPLC fingerprints for *Mallotus* species extracts and evaluation of the peaks responsible for their antioxidant activity. *J. Pharm. Biomed. Anal.* 50, 753–763
105 Tistaert, C. *et al.* (2012) Potential antioxidant compounds in *Mallotus* species fingerprints. Part II: fingerprint alignment, data analysis and peak identification. *Anal. Chim. Acta* 721, 35–43
106 Gao, J. *et al.* (2010) Secondary metabolite mapping identifies *Scutellaria* inhibitors of human lung cancer cells. *J. Pharm. Biomed. Anal.* 53, 723–728
107 Mishra, A. *et al.* (2019) Omics approaches for microalgal applications: prospects and challenges. *Bioresour. Technol.* 291, 121890
108 Patti, G.J. *et al.* (2012) Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269
109 Chen, L.L. *et al.* (2018) Effects of processing adjuvants on traditional Chinese herbs. *J. Food Drug Anal.* 26, S96–S114
110 Borgmann-Winter, K.E. *et al.* (2019) The proteome and its dynamics : a missing piece for integrative multi- omics in schizophrenia. *Schizophr. Res.* In press https://doi.org/10.1016/j.schres.2019.07.025
111 Wolfender, J.L. *et al.* (2019) Innovative omics-based approaches for prioritisation and targeted isolation of natural products-new strategies for drug discovery. *Nat. Prod. Rep.* 36, 855–868
112 Dewi, N. *et al.* (2013) NMR metabolomics for identi fi cation of adenosine A1 receptor binding compounds from *Boesenbergia rotunda* rhizomes extract. *J. Ethnopharmacol.* 150, 95–99
113 Sharma, A. *et al.* (2012) A comparison on the metabolic profiling of the Mexican anxiolytic and sedative plant *Galphimia glauca* four years later. *J. Ethnopharmacol.* 141, 964–974
114 Bapela, M.J. *et al.* (2019) H^1NMR-based metabolomics of antimalarial plant species traditionally used by Vha-Venda people in Limpopo Province, South Africa and isolation of antiplasmodial compounds. *J. Ethnopharmacol.* 228, 148–155
115 Alvarez-Zapata, R. *et al.* (2015) Chemometrics-enhanced high performance liquid chromatography-ultraviolet detection of bioactive metabolites from phytochemically unknown plants. *J. Chromatogr. A* 1422, 213–221