

Review

How to identify “Material basis–Quality markers” more accurately in Chinese herbal medicines from modern chromatography–mass spectrometry data–sets: Opportunities and challenges of chemometric tools

Min He ^{*}, Yu Zhou

Department of Pharmaceutical Engineering, School of Chemical Engineering, Xiangtan University, Xiangtan 411105, China

ARTICLE INFO

Article history:

Received 21 January 2020

Revised 26 March 2020

Accepted 25 May 2020

Available online 6 August 2020

Keywords:

chemometric tools

chromatography–mass spectrometer

material basis

Q–markers

TCMs

ABSTRACT

Modern chromatography – mass spectrometer (MS) technology is an essential weapon in the exploration of traditional Chinese medicines (TCMs) which is based on the “effectiveness–material basis–quality markers (Q–markers)”. Nevertheless, the hardware bottleneck and irregular operation will limit the accuracy and comprehensiveness of test results. Chemometrics was thereby used to solve the existing problems: 1) The method of ‘design–modeling–optimization’ can be adopted to solve the multi–factor and multi–level problems in sample preparation/ parameter setting; 2) The approaches of signal processing can be used to calibrate the deviation from retention time (rt) dimension and mass–to–charge ratio (m/z) dimension in different types of instruments; 3) The methods of multivariate calibration and multivariate resolution can be utilized to analyze the co–eluting peaks in complex samples. When the researchers need to capture essential information on raw data sets extracting the higher level of information on essential features, 1) The significant components which affects the drug properties/efficacy can be found by the pattern recognition and variable selection; 2) Fingerprint–efficacy modeling is explored to clarify the material basis, or to screen out the Q–markers of biological significance; 3) Chemometric tools can apply to integrate chemical (metabolic) fingerprints with network pharmacology, bioinformatics, omics and others from a multi–level perspective. Under these programs, the qualitative/quantitative works will achieve in chemical (metabolic) fingerprint and metabolic trajectories, which leads to an accurate reflection of “material basis and Q–markers” in TCMs. Likewise, an in–depth hidden information can be disclosed, so that the components of drug properties/efficacy will be found. More importantly, multidimensional data can be integrated with fingerprints to acquire more hidden information.

© 2020 Tianjin Press of Chinese Herbal Medicines. Published by ELSEVIER B.V. This is an open access article under the CC BY–NC–ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	3
2. Effective preparation/separation of samples	4
3. Intelligent procession of chemical/metabolic fingerprints	4
3.1. Optimization of retention time (rt) dimension in fingerprints	4
3.1.1. Baseline correction	4
3.1.2. Peak deviation	5
3.1.3. Retention behaviors and their prediction	6
3.2. Optimization of m/z dimension in fingerprints	6
3.2.1. Optimization of ionic isotopes	6
3.2.2. Fragment rules and mass spectrum prediction	7
3.3. Mathematical separation of co–eluted peaks from fingerprints	8

* Corresponding author.

E–mail address: dahai8214813@gmail.com (M. He).

4. Fingerprint-data grouping methods	9
5. Challenges of fingerprint-efficacy modeling tools	10
6. Multidimensional data integration with fingerprints	11
7. Conclusion	12
Declaration of competing interest.	13
Acknowledgements	13
References.	13

1. Introduction

Traditional Chinese medicines (TCMs) have made a great contribution to the maintenance of people's health over the past several millennia. The chemical differences among those medicinal materials in different area or under processing techniques, however, it will inevitably lead to differences in clinical efficacy. In order to solve this strategic problem in herbal industry, the researches of quality standard needed urgently. Nevertheless, the current achievements are still failure to meet the requirements of quality control for TCMs. Especially the explanation for pharmacodynamic substance basis of TCMs is weakly, which has greatly limited the scientific/reasonable selection of quality indicators. The current quality standards are established by referring to those modes in the Western world. It has certain practical significance to evaluate TCMs by measuring the content of one or several components. However, this method cannot embody the traditional theory, such as “king, minister, assistant and guide”, even be caught into the trouble of “the clearer of the ingredients, the weaker of the efficacy”. Additionally, the same chemical evaluation is hard to reflect the characteristics of various medicinal materials. Therefore, the quality indicators should be gradually changed from chemical components to active components, from single component to multi-components. Moreover, this quality system ought to be guided by the practical clinical experience which can ensure the safety and effectiveness of TCMs.

“Markers” are a hot word, such as biomarkers, plant markers, etc. In 2016, academician Chang-xiao Liu has proposed a new concept of quality markers (Q-markers) (Liu et al., 2017; Li, Xie, He, Hou, Liao & Liu, 2019a) which based on the characteristics of the biological properties, manufacturing process and the compatibility theory in TCMs. As a core concept as well as an important basis, Q-marker was regarded as the industry supervision to TCMs. Furthermore, the concept of toxic Q-markers (Zhang, Li, et al., 2018) has also been put forward, which is of great significance to understand the toxic substance basis correctly, establishing a suitable range of dosage carefully, and using toxic TCMs reasonably. Q-markers are usually come from the material basis related with drug properties/efficacy, and represent their overall appearance. “Material basis and Q-markers” play as a team, in which they are harmony for a good prospect of TCMs. Nevertheless, how to explore the relationship between the material basis and drug properties/efficacy from the complex samples, and find out the valid and accurate Q-markers further? Modern analytical instruments and artificial intelligence can solve these problems effectively. At present, one-dimensional gas chromatography (GC), gas chromatography-mass spectrometry (GC-MS), liquid chromatography-ultraviolet detector (HPLC-UV), liquid chromatography-diode array detector-mass spectrometry (LC-DAD-MS) and two-dimensional chromatography (GC × GC or LC × LC) are widely used in the researches of chemical fingerprints, metabolic fingerprints, pharmacokinetics and biosynthesis pathway, etc (Yang et al., 2017). As far as the instrument itself is concerned, the two levels may be divided into “chemical separation” and “detection signal”. For the latter, m/z signals are parallel to other spectral information in each sampling point. Moreover, non-target and target m/z detection can both constitute

a hebal fingerprint, such as the chromatograms under single ion monitoring (SIM), selective reaction monitoring (SRM), multi reaction monitoring (MRM) and full scanning mode. Through GC/LC-MS determination, the variety / quality of Chinese herbs can be identified effectively (Liang, Xie, & Chan, 2010); the trends of Q-markers in the transmission process can be reflected; and the phytochemicals or their metabolites *in vivo* can be detected; And the pharmacokinetic parameters can be determined as well. Herbal samples are very complex (Gu, Shi, Zhang, Huang, & Chen, 2018; Chen et al., 2019) which need not only analytical technologies with high sensitivity, high specificity and full automation, but also efficient methods of high-throughput data procession, modeling and pattern recognition (Liu, Liang, & Liu, 2016) to obtain high-quality data. First of all, the hardware bottleneck or unsatisfactory operation will directly affect the accuracy of chromatography - MS data, such as, the “true” signal is often interfered by noise or other signals. Secondly, the correlation calculation between fingerprints and pharmacodynamic values is a significant source to screen out biological signals in the researches. Thirdly, the integrated pharmacology is needed to develop new tools to mine more information. In 1971, Wold S. proposed chemometrics for the first time, and made some researches on the basic theories/methods (Wold & Christie, 1984). Later academician Ru-qin Yu introduced chemometrics into China in the 1980s, and tried to explore the high-throughput data in TCMs (Liang, Wu, & Yu, 2016). The chemometric tools are presently used in many researches, such as, chemical (metabolic) fingerprints, fingerprint-efficiency modeling (Wang, Xiong, et al., 2017), network pharmacology (Liao et al., 2018) and chinmedomics (Sun et al., 2019). Only by data mining, can it identify the material basis and screen out Q-markers more accurately from GC/LC- MS data-sets.

This review summarized the application of chemometrics in GC/LC- MS data- sets, especially for Chinese herbal samples. Additionally, some suggestion and future designs have been offered as well. As shown in Fig. 1, the main contents include the following aspects: effective preparation/separation of samples, intelligent analysis of chemical (metabolic) fingerprints, fingerprint-data

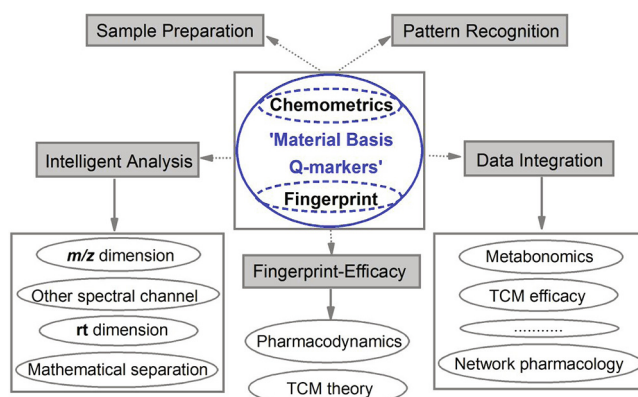


Fig. 1. Application of chemometrics in GC/LC- MS fingerprints.

grouping methods, challenges of fingerprint-efficacy modeling tools, and multidimensional data integration.

2. Effective preparation/separation of samples

Herbal analysis is inseparable from sample preparation, in which a large number of items were handled at randomly. The sampling steps for Q-markers include: identification of medicinal materials; standardized treatment of processed pieces; optimization of sample extraction. The latter includes generally: material/liquid ratio, soaking time, extraction time, temperature, concentration, and so on. Then, the samples to be collected are determined by GC/LC related apparatus. Some parameters are necessary to be adjusted many times in this chromatographic determination, e.g. column, mobile phase and the elution gradient.

There are many multi-factor and multi-level problems in both sample preparation and GC/LC determination. Thus, many chemometric methods have been applied in industrial extraction (Sharif et al., 2014), laboratory extraction (Narendran, Meyyanathan, & ReddyKarri, 2019; Mousavi, Tamiji, & Khoshayand, 2018) and chromatographic analysis (Hibbert, 2012), e.g. full factorial design (FFD), partial factorial design (FRFD), pocket Burman (P-B) design, Taguchi design based on the orthogonal array, central composite design (CCD), Box-behnken design (BBD), Doehlert design, D-optimal design and other designs, etc. For example, full factorial design and response surface design were used to extract target components from medicinal materials (Miti et al., 2019; Mohammad Munawar, Surya Prakash, & Vangalapati, 2018). These programs can be obtained from Design-Expert (Stat-Ease Inc., www.statease.com/dx11.html); Fusion Pro (S-Matrix Corporation, www.smatrix.com/fusion_pro.html); Modde (Umetrics, <https://umetrics.com/product/modde-pro>); Unscrambler (Camo AS, www.camo.com/p2_tuf.htm); DOE wisdom (Launsby Consulting, www.launsby.com/BookIM.html). Meanwhile, statistical software can also be used in experimental designs, e.g. SPSS (IBM, www.ibm.com); Matlab (The Mathworks Inc., www.mathworks.com/); Origin (Microcal Software, www.originlab.com).

Many tools are used to analyze the complex relationships between response indicators and factors, including response surface methodology (RSM) (Carabajal, Teglia, Cerutti, Culzoni, & Goicoechea, 2019), Excel (Cabeza, Sobrón, García-Serna, & Cocero, 2016). RSM was employed to explain *Cyperus Rhizoma-Chuanxiong Rhizoma*, or *Cyperus Rhizoma-Angelicae Sinensis* interactions which based on LC-MS determination (Liu, Shang, Zhu, Qian, & Duan, 2018). During the analysis of triterpenic acids in TCMs, RSM was combined with BBD to optimize the main experimental parameters that will affect extraction efficiency and derivatization yield (Wu et al., 2015). Moreover, the researchers developed free and encapsulated *Arjuna* herb extract added vanilla chocolate dairy drink by using Central Composite Rotatable Design (CCRD) of RSM (Sawale, Patil, Hussain, Singh, & Singh, 2020). The ideal RSM results indicated that the experimental data can be applied to a mathematical equation, which is an effective statistical model (Liu et al., 2019). The object usually varies greatly in the practice of 'material basis and Q-markers' research, generating a complicated function in many problems. Nowadays, many new global optimization algorithms are emerging, such as gray wolf optimization (GWO) (Kulkarni & Kulkarni, 2018), particle swarm optimization (PSO) (Soepangkat Norcahyo, Effendi, & Pramujati, 2019), genetic algorithm (Mokhtari & Ghoreishi, 2019), ant colony optimization (Karri, Sahu, & Meikap, 2020), etc. For example, GWO algorithm was used to optimize the process parameters of essential oil extraction from *Cleome coluteoides* Boiss (Sodeifian, Ardestani, Sajadian, & Ghorbandoost, 2016). Nevertheless, GWO does have

numerous disadvantages, such as, an overreliance on the initial population, premature convergence, prone to local optimum and the unstable convergence process. Therefore, GWO combined with support vector machine (SVM) was used to predict the solubility of aromatic substances in super critical carbon dioxide (Bian, Zhang, Zhang, & Chen, 2017). Also, the scholars used an improved chaotic-GWO algorithm to optimize the experimental parameters in supercritical CO₂ extraction from Chaihu Shugan San (He, Hong, Yang, Yang, & Zeng, 2018).

3. Intelligent procession of chemical/metabolic fingerprints

Ancient books are a great treasure house for the development of TCMs. In the outline of strategic planning for the development of TCMs (2016–2030), the Chinese state council has clearly proposed to cultivate a number of famous prescriptions with international competitiveness. The weakness of 'substance basis' researches of TCMs in the past 20 years that limits the scientific/reasonable selection of quality indicators greatly. Chinese herbal preparations are thereby not widely accepted by the international community, and their effectiveness (safety) is always questioned. Nowadays, accompanying the great progress of science and technology, many technologies (fingerprint-efficiency modeling, metabonomics, serum pharmacology, bio-chromatography and affinity ultrafiltration) are committed to the identification of the material basis of pharmacodynamics. Nevertheless, the traditional concepts of drug properties/efficacy, such as "Yinjing Baoshi" (guiding action), "Xiangxu" (work in coordination) and "Guijing" (channel tropism) in mandarin, still remain unclear. Modern chromatography-MS technology has become the primary weapon of 'material basis and Q-markers' exploration in TCMs. Followings are its applications: 1) Applying to find the changes of active components during the process of collection, processing, preparation for herbal materials; 2) Utilizing to determine the "component- component" correlation in herbs; 3) Developing to explore the mechanism of drug absorption, distribution, metabolism and excretion (ADME). Unfortunately, its hardware bottleneck and improper operation will lead to the deviation or wrong conclusion in chemical (metabolic) fingerprints. Chemometrics is needed to remove the "mask" in the process of signal discrimination urgently, which came from the non-ideal data in TCMs / biological samples. Finally, the 'true' values can make the 'material basis and Q-markers' exposed to the maximum extent.

3.1. Optimization of retention time (rt) dimension in fingerprints

3.1.1. Baseline correction

In the chemical/metabolic fingerprint analysis, the baseline drift under non-ideal operation or instrumental fluctuation will affect statistical discrimination and qualitative/quantitative analysis. This will inevitably lead to an error recognition of 'material basis and Q-markers', also cannot qualify the herbal products. It is necessary for researchers to deal with these raw data by various methods, e.g. the improved iterative polynomial fitting with automatic threshold (Gan, Rian, & Mo, 2006). In Liang group, an adaptive iterative re-weighted penalty least squares algorithm (Zhang, Chen, & Liang, 2010) (airPLS, <http://code.google.com/p/airpls>) have been successfully used for the data sets from Chinese herbal samples. Besides, statistical entropy was used as an indicator to distinguish the real metabolite signals from (system) noises, so as to subtract their backgrounds (Krishnan et al., 2012). An automatic two side empirical baseline correction algorithm (ATEB) has been proposed as well, which is based on bilateral exponential smoothing algorithm and iterative fitting strategy (Liu et al., 2014). In recent years, some new baseline correction schemes (He et al., 2016; Qian, Wu,

& Hao, 2017; Lin, Zong, Tian, & Lin, 2018; Sawall et al., 2018) have been proposed constantly and applied to near fault ground motion data, Raman and NMR spectra, etc. It is worth noting that these methods can also be applicable to one-dimensional (1D) chromatographic data of Chinese herbal samples.

3.1.2. Peak deviation

Under the same experimental conditions, the same retention time (rt) should be displayed in the same phytochemicals from different batches of herbal samples. Due to the mobile phase, stationary phase, temperature, pressure, delayed injection and others, peak deviations were observed in different samples usually. These deviations will affect the following statistical discrimination, and also influence the second-order calibration for peak cluster from complex samples. Similarly, an error recognition of “material basis and Q-markers” will happen in this uncorrected dataset. At present, a number of peak alignment algorithms have been put forward, such as dynamic time warping (DTW) (Kassidas, Macgregor, & Taylor, 1998), correlation optimized warping (COW) (Vest Nielsen, Carstensen, & Smedsgaard, 1998), fuzzy warping (FW) (Walczak & Wu, 2005), chromalign (Sadygov, Maroto, & Hühmer, 2006), msalignment 2 (Palmlblad, Mills, Bindschedler, & Cramer, 2007), parametric time warping (PTW) (Bloemberg et al., 2010), peak alignment using reduced set mapping (PARS) (Torgrip, Åberg, Karlberg, & Jacobsson, 2010), multi-scale multiscale peak alignment (MSPA) (Zhang et al., 2012), etc. Among them, DTW cannot handle huge data, and COW will cost time to optimize parameters that may change the peak shape; MSPA does not have the shortcomings of the two algorithms above, but it is not adequate enough as to overlapped or embedded peaks. Recently, some peak alignment algorithms have come out, such as chromatogram alignment via mass spectrometry CAMS (Zheng et al., 2013), metabolite compound feature extraction and annotation (MET-COFEA) (Zhang et al., 2014), an automatic non targeted metabolic profiling analysis (anTMPA) (Fu et al., 2017). These algorithms can achieve more precise alignment in that they make full use of the mass spectral information in each sampling point. However, these alignments are easily affected by singular values in overlapped peaks or low signal-noise ratio (SNR) peaks. Unfortunately, such peaks could be found in the chromatograms under SIM, SRM, MRM and full scanning for Chinese herbal samples fre-

quently. To solve this problem, some researchers proposed an alignment algorithm of subwindow factor analysis based on mass spectral information (SFA-MS) (Yang et al., 2018). As shown in Fig. 2, the eigenvalues among different spectra from the chromatographic sub-windows can be calculated in this algorithm. As the eigenvalue approaches to 1, the signals in two samples are deemed as high similarity. SFA-MS can accurately align the co-eluted peaks without changing their shape, which has been applied in the GC-MS data set of *Bupleurum chinense* (He, Hong & Zhou, 2019a). At this stage, SFA-MS and CAMS are inappropriate for the raw data from hyphenated apparatus with high-resolution (HR) MS. Recently, deep learning method was used to align GC-MS data-set from both triple quadrupole MS and HR MS apparatus (Li & Wang, 2019b). This model is needless to input reference data, but it is unfit for the co-eluted peaks. In addition, the modified MSPA algorithm (mMSPA) (He, Yan, Yang, Zhang, et al., 2018) was applied in the second-order calibration, on account of it can align the multi-channel (m/z , wavelength) curves in two-dimensional data-set from herbal sample simultaneously.

Nowadays, comprehension two-dimensional gas chromatography (GC \times GC) and on line two-dimensional liquid chromatography (LC \times LC) have been widely used in the analysis of herbal/biological samples. These multidimensional technologies could better excavate ‘material basis and Q-markers’ to be discovered in the present and even into the future. Although their peak capacity was enhanced, the peak deviations could be observed in the GC \times GC/LC \times LC data array still. Peak detection is a vital step in GC \times GC peak alignment, and some methods have been summarized (van Stee & Brinkman, 2016). Many algorithms were used to align the peaks in GC \times GC-MS data-sets, e.g., a cylindrical mapping method (Weusten, Derks, & Mommers, 2012). Tauler's group (Parastar, Jalali-Heravi, & Tauler, 2012) proposed a bilinear method based on multivariate curve resolution, which can be used for GC \times GC peak alignment. In DISCO2 algorithm (Wang et al., 2011; Wang, 2013), multiple peak entries of the same metadata are firstly merged into one peak entry; peaks in all samples were then marked based on both rt and mass spectral similarity by Pearson's correlation coefficient; A local linear fitting method is finally used to align rt shift in GC \times GC / TOF-MS dataset. Besides, the semi-parametric approach was used to establish model between a two dimensional “warp function” and shifts, thus aligning the

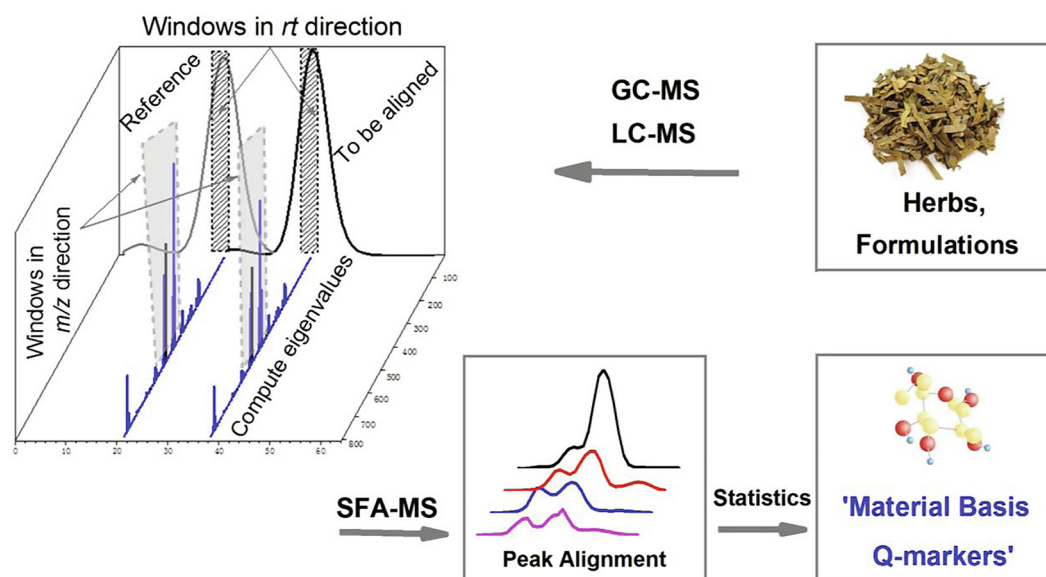


Fig. 2. Application of SFA-MS algorithm in herbal fingerprints. The data source is updated from the literature: Journal of Chromatography A, 1563, 162–170.

GC \times GC data from diesel oil (de Boer & Lankelma, 2014). A pixel based approach was also utilized to eliminate background interference and to perform peak alignment (Furbo, Hansen, Skov, & Christensen, 2014). In GC²MS platform (<http://gc2ms.web.cmdm.tw>), peak detection, baseline correction and peak alignment can be severed for GC \times GC data (Tian et al., 2016). All in all, the advanced instrument design and perfect signal procession (Prebihalo et al., 2018) can provide support for the precise analysis in the multidimensional chromatography.

3.1.3. Retention behaviors and their prediction

The identification of unknown signals in chemical (metabolic) fingerprints has become a hurdle that needs to be overcome, which restricts the recognition of 'material basis and Q- markers' in TCMs. For LC (GC) -MS data, the similarity search (in commercial or *in-house* databases) and fragmentation deduction for compound identification is fundamental. Nevertheless, the appearance of the similar mass spectra or fragmentation in many compounds will bring great trouble in the identification of unknown signals. To solve this problem, retention indices (RIs) in GC-MS data were used for the auxiliary identification of chromatographic peaks (Babushok, 2015). Regrettably, the RIs of desirable is not always available in the commercial/ *in-house* retention-data collections. Therefore, *in silico* RIs of analytes should be developed in quantitative structure-retention relationship (QSRR) calculations for chromatography (Amos, Haddad, Szucs, Dolan, & Christopher, 2018). Some software packages have been written to predict the retention of different chemicals in different columns, e.g. ACD/ChromGenius (http://www.acdlabs.com/products/com_iden/meth_dev/chrom-gen/). Many methods have been used for QSRR methodologies, e.g. SVM (Luan et al., 2005), random forests (RF) (Goudarzi, Shahsavani, Emadi-Gandaghi, & Arab Chamjangali, 2014), monte Carlo method (Veselinović et al., 2017), genetic algorithm (Zhang, Zheng, Xia et al., 2017), deep learning (Matyushin, Sholokhova, & Buryak, 2019). It's worth mentioning that the data sources and model size will affect the accuracy in the modeling processes. Under the appropriate parameter settings, these approaches adapted to the QSRR modeling of herbal ingredients. For example, a large number of RIs from the known terpenoids were used as the training set in a RF model, through which the predicted values of unknown terpenoids were proved to be close to the real values (He et al., 2013). All in all, RIs prediction can enhance the confidence level of compound identification when it combined with multivariate analysis, accurate mass determination and EI-MS spectral databases (Dossin et al., 2016; Matsuo, Tsugawa, Miyagawa, & Fukusaki, 2017).

RIs can also improve the component identification in GC \times GC data-sets (Jiang, Kulsing, Nolvachai, & Marriott, 2015) from herbal samples. Some methods are used to calculate ²D RIs in GC \times GC data-sets, such as, regression algorithm (Mazur, Zenkevich, Artaev, Polyakova, & Lebedev, 2018), facile approach (Jiang, 2019). The systematic error in GC \times GC data-set is related to the flow rate and the heating rate (Jaramillo & Dorman, 2018), which can be reduced by a model calculation. For the *in silico* RIs of unknown compounds, a lot of modeling of GC \times GC separations have been built (Jaramillo & Dorman, 2019). These prediction values of GC \times GC were applied in the identification of biological components later, e.g. steroids (Randazzo, Bileck, Danani, Vogt, & Groessl, 2019). Equally, *in silico* RIs of GC \times GC can make out the secondary metabolites accurately with the combination of the mathematical separation, fragmental rules, and so on (He, Yan, Yang, Ye, et al., 2018).

Most flavonoids, saponins, alkaloids and other phytochemicals with high-boiling point characteristics are inappropriate for GC separation, which are the major source of 'material basis and Q-markers'. The LC separation without any standard retention-data

collection, the main complementary way of identification for herbal components relies on the reference. The trouble is that too many herbal ingredients need to be identified by LC-related technologies, which is dissatisfied by a few reference materials. Some researchers tried to make the *rt* prediction (Lochmuller, 1995; Taraji et al., 2018) that based on a set of known values under the same experimental conditions. Scientists developed PredRetplatform (<http://predret.org>) that makes community sharing of *rt* information across laboratories possibly (Stanstrup, Neumann, & Vrhovšek, 2015). Thus, the researchers can distinguish those isomers with similar mass spectra but different retention behaviors based on the predicted values effectively. For example, the combination of high-resolution MS analysis and predicted LC *rt* filtering (Chervin et al., 2017) were used to identify compounds in *Streptomyces* extracts. The phthalide isomers were also distinguished in *Ligusticum chuanxiong* by using qualitative analysis and *rt* prediction (Zhang, Huo, Zhang, Qiao, & Gao, 2018). It is notable that the appropriate machine learning method is suitable for different circumstances (Bouwmeester, Martens & Degroev, 2019), especially for the application to diverse types of compounds in each Chinese herbal medicine.

3.2. Optimization of *m/z* dimension in fingerprints

3.2.1. Optimization of ionic isotopes

Unknown signals in chemical (metabolic) fingerprints are manifested in active components in TCMs frequently. This has become a knotty problem to identify 'material basis and Q-markers'. Usually the accurate mass and isotopic distribution can be transformed into the molecular formula of target compounds, which is the first step of identification. Seven golden rules (<https://fiehnlab.ucdavis.edu/projects/seven-golden-rules>) or an online molecular formula tool (available through chemcalc software) (Doucette & Chisholm, 2019) can deal with it well. The difficulty lies in the accuracy of measurement in fingerprints.

Nowadays, triple quadrupole MS, TOF MS, IT-TOF MS, qTOF MS instruments are predominant in most laboratories. In daily works, non-ideal operation will inevitably give rise to large deviations between the measured values and the true values. And, the different types of instruments present various isotopic structures due to their own operation principles. Under TOF-electrospray ionization (ESI) condition, accurate mass axes and unstable isotopes are exhibited (Mihaleva et al., 2008). In LC-qTOF apparatus, the drifts of mass axes and isotope abundances are affected by the signal intensity greatly (He, Nie, Wu, & Liang, 2015). Nevertheless, low mass accuracy is often displayed in weak molecular ion under electron ionization (EI) mode (Lau et al., 2019). Similarly, these subtle variations appeared in MSⁿ fragment species, e.g. ionic mass of LC-MSⁿ, LC-qTOF-MS/MS, GC-MS determination. In general, the variations of these errors are relevant to signal intensities and *m/z* values in high resolution instruments (Vergeynst, Van Langenhove, Joos, & Demeestere, 2013). By report, GC-MS tuning frequency is a key variable in relative abundance variation (Kelly, Brooks, & Bell, 2019). Presently, modern fourier transform-ion cyclotron resonance (FT-ICR, Bruker) (Guan et al., 2018) and orbitrap analyzer (Thermo Fisher Scientific) (Xu et al., 2019) takes on both ultra-high mass accuracy and ultra-high resolving power has been used in the study of herbal constituents. Their systematic errors are also non-linearly existence in the measurements, which is correlated with the signal intensities or others (Cox & Mann, 2009). Therefore, these instruments should be improved to meet higher requirements in different samples as well.

Internal calibrations based on different algorithms have been developed for the distorted isotopes in the LC/GC-MS data-sets (Cappellin et al., 2010; Doherty, Grégoire, & Bertrand, 2008). Other approaches have also been implemented to reduce systematic

errors in hyphenated chromatographic apparatus, e.g. an EXCEL[®] application GIMiCK (Stoll-Werian et al., 2019). For the saturated mass spectra, an open source computational method was created to re-calculate the precursor m/z values and intensities (Bilbao et al., 2018). As to the imaging MS, a novel automated work-flow for spectral alignment and mass calibration was constructed to obtain mass errors as low as 5 ppm using a TOF instrument (Ràfols, del Castillo, Yanes, Brezmes, & Correig, 2018). However, random errors will cause the measurement values fluctuate around the real value, and their range could be affected by signal strength or others. A comprehensive strategy for conducting measurements of isotope ratios and mass values was thereby made, accommodating the observed run-to-run variations (Graczyk, McLain, Tsai, Chamberlain, & Steeb, 2019). In order to reduce systematic/ random errors, the concept of spectral accuracy for MS was subsequently proposed (Wang & Gu, 2010). On this theoretical basis, the exploitation of MassWorks software is developed to correct the mass axes and isotope abundances in triple quadrupole MS (Jiang & Erve, 2012). Inspired by this theory, an external calibration was used to improve the ionic mass accuracy after an isotopic shape correction, and terpenoids from *Ephedra* is illustrated as an example (He et al., 2013). However, the subjective factor is easily introduced in the modeling procedure, e.g. peak shape function. Therefore, an algorithm to correct ionic isotopes was implemented, calling it automatic averaging of target ions in an interesting domain for internal correction (AAID-IC) (Hong, Li, He, Zhao, & Li, 2020). This procedure (Fig. 3) can be applied in herbal data-sets from various types of MS apparatus, which also needs continuous upgrades and improvements in the environment of modern data.

3.2.2. Fragment rules and mass spectrum prediction

After the molecular formula calculation, the m/z fragments become an important evidence for the identification of drug prop-

erties/efficacy related components. This qualitative result provides a guarantee for exploring the ‘material basis and Q-markers’ in TCMs, which based on the thought of “property- effect - substance”. How to complete the peak attribution by using the m/z fragments in GC (LC)-MS data-set? The standard libraries become the main accesses in which the similarity between measure and reference spectra are scored. Among them, the commercial libraries include Wiley library, NIST library (<https://webbook.nist.gov/chemistry/>), Massbank (<http://www.massbank.jp/>), saditer library, etc; the special libraries include standard pesticide library, Pflieger drug library, essential oil library, etc; *in-house* libraries are also a significant source which can be built by the users. In TCMID 2.0 (<http://www.megabionet.org/tcmid/>) (Huang et al., 2018), 3895 MS spectra of 729 ingredients was collected.

There are no reference EI/ESI spectra record in standard databases (NIST, Mass bank etc.) for many herbal components. Even worse, those m/z fragments from several types of LC-ESI-MSⁿ are significantly different. Even for the same ESI-MSⁿ instrument under various bombardment energies, the fragments with different abundances can also be observed. Therefore, fragmentation rules can give a reference to identify herbal components. At present, a large number of studies have been made on the fragmentation behaviors of phytochemicals in MS apparatus (Steinmann & Ganzera, 2011; Ganzera & Sturm, 2018; Zhang, Sun, Yan, & Wang, 2017), including flavonoids, terpenes, alkaloids, etc. Therefore, the fragmentation rules can be summarized from these researches. After accurate mass determination (formula calculation), the researchers can obtain the candidate compounds from an *in-house* TCMs library. Next, the experimental MSⁿ deduction is used for their structure confirmation under the guidance of these fragmentation patterns.

For unknown peaks, the alternative strategy is resort to *in silico* MS/MS library or software (Ma, Kind, Vaniya, & Fahrman,

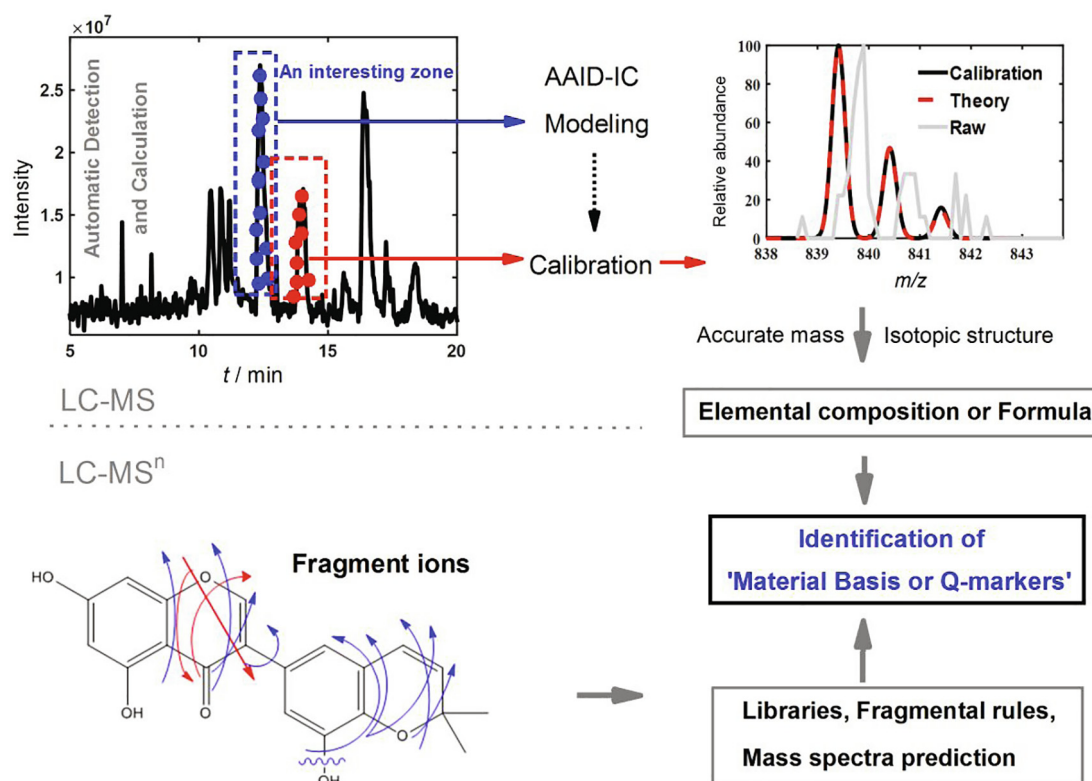


Fig. 3. Chemometric researches of accurate mass, isotopic profile, MS/MS fragments in herbal analysis. The data source is updated from the literature: Journal of Chromatography A, 1613, 460668.

2015; Allard et al., 2016). On the basis of operation mechanisms, the software can be divided into three types of groups: (1) *in silico* fragmentation methods: MetFrag (<http://c-ruttkies.github.io/Met-Frag/>) (Ruttkies, Schymanski, Wolf, Hollender, & Neumann, 2016), CFM-ID3.0 (<http://cfmid3.wishartlab.com.>) (Djoumbou-Feunang et al., 2019), MAGMa+ (<https://github.com/savantas/MAGMA-plus>) (Verdegem, Lambrechts, Carmeliet, & Ghesquière, 2016), MIDAS (<http://midas.omicsbio.org>) (Wang, Kora, Bowen, & Pan, 2014), and MS-Finder (http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/) (Tsugawa et al., 2016); (2) fingerprint-based methods: CSI-FingerID (www.csi-fingerid.org) (Dührkop, Shen, Meusel, Rousu, & Böcker, 2015); (3) MS/MS spectra prediction based on the structural similarity, that is “known-to unknown” methods. New tools in recent years: SF-Matching (http://www.bork.embl.de/Docu/sf_matching) (Li, Kuhn, Gavin & Bork, 2019c) based on RF model; SIRIUS 4 (<https://bio.informatik.uni-jena.de/sirius/>) (Dührkop et al., 2019); DeepMASS (<https://github.com/hcji/DeepMASS.>) (Ji, Xu, Lu, & Zhang, 2019) based on structure similarity approach.

Using the approaches above, the mass spectra can be mapped to the corresponding chemicals as much as possible. For flavonoids in Licorice, the researchers combined fragmentation rules with MS-FINDER prediction to explain their MSⁿ data (He, Wu, et al., 2017). In addition, TCM molecular networking based on *in silico* MS² spectra with integration of virtual screening and affinity MS screening were used to discover functional ligands from natural herbs (Wang, Kim, et al., 2019). Through these methods of qualitative analysis, the drug properties/efficacy- related components (peaks) can be accurately identified in herbs (data-set). Likewise, it has laid the foundation for the selection of Q-markers in relevant researches.

3.3. Mathematical separation of co-eluted peaks from fingerprints

On account of the herbal complexity, the appearance of co-eluted peaks is existed in the 1D or even 2D chromatographic separations. From the quantitative point of view, SIM, SRM and MRM, all of them can solve this problem. Whereas in full scan mode, the distorted signals will seriously affect the qualitative accuracy of the target compounds, which has become a huge barrier to identify ‘material basis and Q-markers’. In order to solve this problem, the usage of multivariate calibration and multivariate resolution could deal with these complex data-sets, and form a new green analytical chemistry idea of “mathematical separation” (Fig. 4) gradually. The commonly used methods includes: direct trilinear decomposition method (DTLD) (Sanchez & Kowalski, 1990), heuristic evolving latent projections (HELP) (Liang & Kvalheim, 1992), parallel factor analysis (PARAFAC) (Bro, 1997; Mitchell & Burdick, 2010), alternating trilinear decomposition (ATLD) (Wu, Shibukawa, & Oguma, 1998), bilinear least squares/ residual bilinearization (BLS/RBL) (Linder & Sundberg, 1998), generalized rank annihilation method (GRAM) (Faber, 2001), alternative moving window factor analysis (AMWFA) (Zeng et al., 2006), selective ion analysis (SIA) (Tan, Liang, & Yi, 2010), PARAFAC2 (Kiers, ten Berge, & Bro, 2010; Bro, Andersson, & Kiers 2010), multivariate curve resolution - alternating least squares (MCR-ALS) (Kumar & Mishra, 2015), etc. These classical algorithms have been widely used in the analysis of two-dimensional (2D) matrix (Liang, Xie, & Chan, 2004) and three-dimensional (3D) array (Zhao et al., 2012) from TCMs, and searching the qualitative/ quantitative information of the target compounds will be more precisely.

With the development of multidimensional chromatography technologies, the increased peak capacity has become a sharp

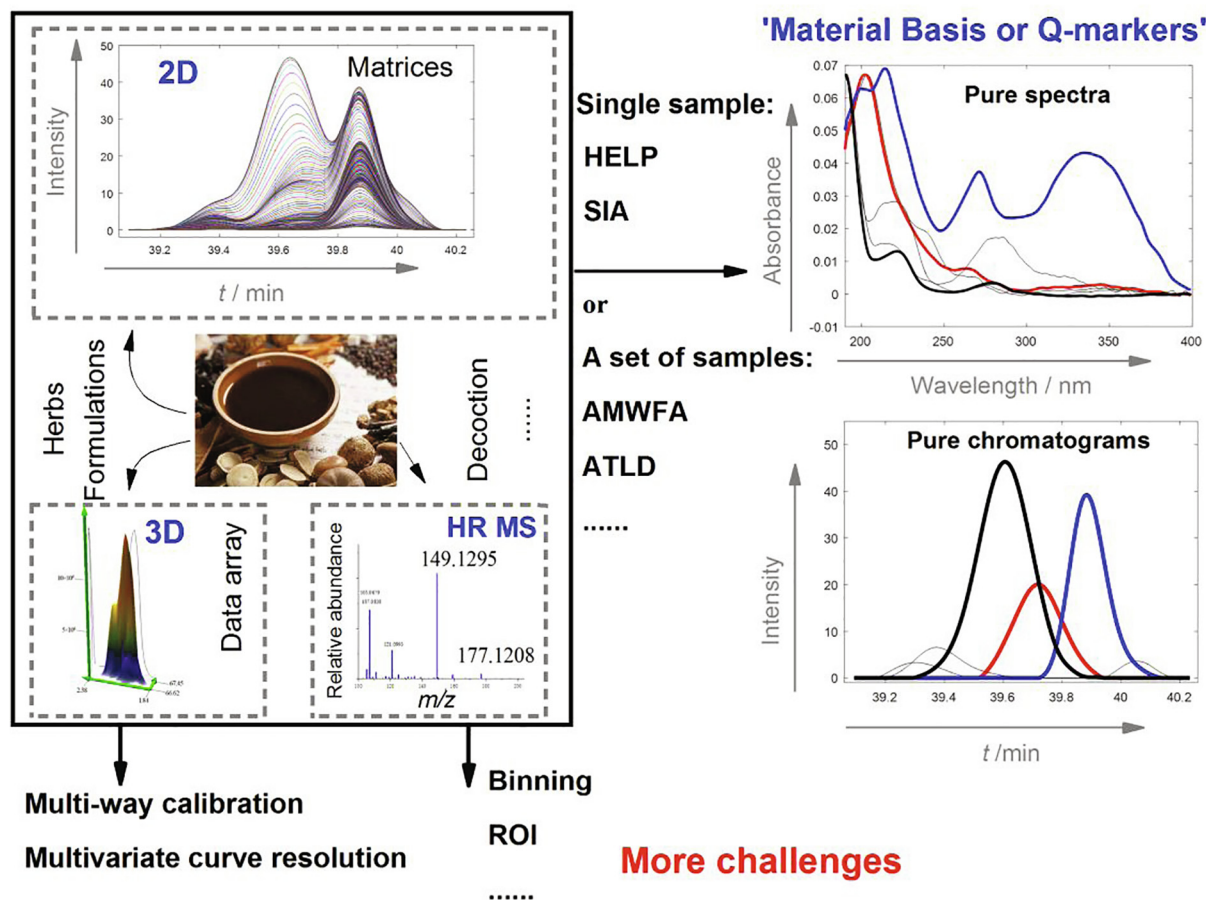


Fig. 4. Mathematical separation for co-eluted peaks from herbal fingerprints.

weapon to unlock TCMS box. Under the MS detection, such instruments have been used in the studies of herbal or biological samples diffusely. Thus, it could provide the accurate identification of material basis in herbs, and give some clues for the disclose of Q-markers. For example, an off-line LC \times LC/ultra-high performance supercritical fluid chromatography tandem qTOF MS system was used to separate and identify 229 bufadienolides in *Bufo venenum* (Wei et al., 2019). LC \times LC-quadrupole-Orbitrap MS was utilized to separate 270 components in Erzhi Pill, and 146 compounds were identified (Fu et al., 2019). Moreover, GC \times GC technology has been applied in the pharmaceutical and biomedical fields, showing more advantages compared with other methods (Aspromonte, Wolfs, & Adams, 2019). However, the hardware is still unable to meet the needs of all compounds in a complex herbal sample, and the co-eluted peaks remains exist in the 3D array. So the researchers used the second-order calibration methods such as MCR-ALS trilinear and PARAFAC2 to analyze the co-eluted peaks from LC \times LC-MS data set (Navarro-Reig, Jaumot, van Beek, Vivó-Truyols, & Tauler, 2016). Tauler' group also used wavelets to compress GC \times GC TOFMS data-sets, and established appropriate column-wise data matrix augmentation arrangement, then used MCR-ALS modeling to analyze eighty *D. magna* metabolites (Izadmanesh et al., 2017). The scholars further proposed two approaches (He, Yang, et al., 2017): 1) transforming a raw 3D array into "2D row-wise slice" set, expressed as $\mathbf{X}_{aug} (\mathbf{I} \times \mathbf{K}, \mathbf{J})$, then use non-iterative multivariate curve resolution methods (HELP and SIA) to analyze the sub matrix $\mathbf{X}_i (1, 2 \dots k)$; 2) using the second-order/three way algorithm to resolve a 3D sub-array, such as ATLD (Vignaduzzo, Maggio, & Olivieri, 2020); certainly, the third-order/four way algorithms can also be used for a set of samples. In order to identify the compounds more exactly, a multiple-strategy analysis taking *Cyperus* as an example was proposed (He, Yan, Yang, Ye, et al., 2018): the combination of physical separation and mathematical separation; the union of similarity searches and MS/MS *in silico*; and the alliance of RIs calculation and QSRR approaches.

HR-MS has been repeatedly used in the herbal analysis, such as qTOF-MS, IT-TOF MS and Orbit MS. Compared with triple quadrupole MS, such HR-MS can obtain more accurate mass axis, wider mass range and lower matrix interference. These apparatuses are widely used in the non-target/target detection of drug properties/efficacy related ingredients (Gröger, Käfer, & Zimmermann, 2020; Alvarez-Rivera, Ballesteros-Vivas, Parada-Alfonso, Ibañez, & Cifuentes, 2019) provide 'material basis and Q-markers' exploration sharper weapons. These data-sets are influenced by noise interference, incomplete separation and other problems, which needs to be analyzed by chemometric tools. However, a more complex internal structure exists in the non-target data-sets, which is unable transformed into a huge matrix of equal distance to be analyzed by chemometric tools. Some methods, such as binning and region of interest (ROI) compression, are in sore need of HR-MS data. ROI-MCR was thereby written to resolve the overlapped or embedded peaks in high-resolution MS data-sets (Dalmau, Bedia, & Tauler, 2018; Navarro-Reig, Jaumot, & Tauler, 2018). Furthermore, ROI-SIA was put forward to obtain the clean mass spectra of various compounds in *Ligusticum chuanxiong* (He, Peng, Xie, Hong & Gao, 2019b), which is a parameter-free method.

4. Fingerprint-data grouping methods

The external factor of climate conditions, soil factors, harvest time, genetic variations, processing methods and preparation technologies have great influence on the variety and content of the secondary metabolites, which seriously affect the process of TCMS'

modernization. And, chemical pattern recognition is an indispensable scientific tool which can be applied in the researches of 'material basis and Q-markers' in TCMS, e.g, material base of five flavors, variety identification, origin, growth periods, storage years, decoction process, and so on. It is also an important tool in the study of fingerprint- efficacy modeling, metabonomics, serum pharmacology, and so on. Chemical pattern analysis and chemical pattern recognition have always been a pivotal part of chemometrics. After chemical measurements, it can reveal the hidden information from the complex samples, providing the valuable clues to analytical chemists. Therefore, pattern recognition can be used to extract the overall information of various components in TCMS, and to screen out possible Q-markers through variable selection. The main steps include: a series of multivariate data was applied to build the training set; features extraction and data preprocessing; training and classification by machine learning methods; verify the model availability; and analyze/distinguish different samples. At present, pattern recognition methods are mainly divided into supervised pattern recognition, unsupervised pattern recognition – cluster analysis, pattern recognition based on projection, classification and regression trees.

The chemical fingerprints can reflect the chemical characteristics of secondary metabolites in TCMS. Their "integrity" and "fuzziness" can be utilized to evaluate the quality of medicinal materials, processed products and herbal preparations. Because of a lot of multivariable data reflecting chemical information in TCMS, the data mining tools seem to be extremely crucial. Many unsupervised pattern recognition algorithms are used for herbal analysis, for instance, principal component analysis (PCA) (Russo et al., 2019), nonlinear mapping (NM) (Sammon, 1969) and cluster analysis (CA) (Rajadurai & Sankaranarayanan, 2012). In order to predict the unknown samples, the supervised pattern recognition methods are proposed in which a large number of known samples are used as training set. 1) The linear methods mainly include: PLS-discrimination analysis (PLS-DA) (Ballabio & Consonni, 2013), linear discriminant analysis (LDA) (Ni et al., 2012), orthogonal projections to latent structure discriminant analysis (OPLS-DA) (Kang et al., 2008), etc; 2) The nonlinear methods include: RF (Svetnik et al., 2003), SVM (Yan, Zhan, & Zhu, 2009), etc.

Among the tools above, PCA is commonly used in herbal quality evaluation through the feature reduction, image clustering and image classification, e.g, planting, procession, preparation and storage of TCMS. For example, 30 *Bupleurum* samples were collected from five regions in China, namely Luliang City in Shanxi Province, Aba and Ganzi Prefecture in Sichuan Province, Longnan City and Dingxi City in Gansu Province. After GC-MS determination and SFA-MS alignment, PCA was used to classify (distinguish) these *Bupleurum* samples in Fig. 5 (He, Hong & Zhou, 2019a). In addition, PLS-DA models were utilized to find the species-specific markers from *Fritillaria* species, after UPLC-qTOF MS analysis (Liu et al., 2020). In this procedure, the most promising ions responsible for class separation were selected by VIP plot; the intensities of the selected ions were visualized to further identify the potential species-specific markers; the sorted specific markers were rechecked in the raw data that set to ensure the peak quality and specificity among assessed species. Several tools are usually combined to distinguish different groups and find out the important constituents, as the examples, PCA, hierarchical cluster analysis (HCA) and heatmap in the research of aromatic components from fresh, natural aged and the accelerated aged white tea (Qi et al., 2018); PCA, HCA and similarity analysis in the research of Q-markers from *Ligusticum chuanxiong*- *Cyperus Rhizome* (Guo, Gong, Wu, Qiu, & Ma, 2020). Also, PCA, factor analysis (FA) and HCA were utilized to evaluate the differences of 14 indicators in different *Poria cocos* decoction pieces from different levels (Zhu et al., 2019). In addition, PLS-DA, SVM and RF are also widely used

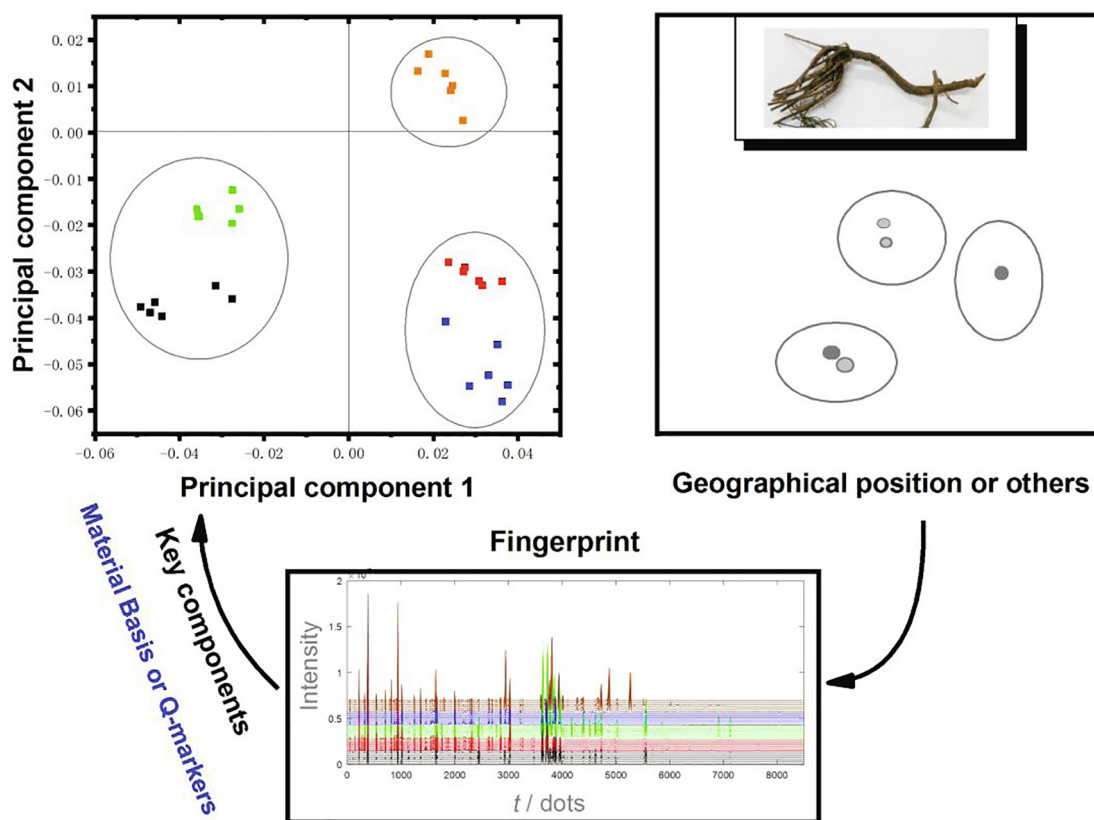


Fig. 5. Classification for *Bupleurum* samples by PCA. The data source is updated from the literature: Journal of Separation Science, 42(11), 2003–2012.

in the study of the geographical location, administration and processing technology of Chinese herbal medicines (Wu, Zuo, Zhang, & Wang, 2018).

5. Challenges of fingerprint-efficacy modeling tools

Since the 1930s, scientists have isolated and purified monomers to determine their biological activities, and to explain their mechanisms. Alternately, many pharmacological models have been used for the tracking extraction, separation and identification of active compounds. The efficacy as an important evaluation index against the different batches of herbal preparations, these approaches are time/labor-consuming, assessment indicators-unclear, or integrity-deficiency. Therefore, the current achievements are still unable to meet the requirements of quality control for TCMs. As a result, single or several components were used as the evaluation index in TCMs, which has been constantly questioned in domestic and overseas. The fingerprint was regarded as a powerful tool for assessing the batch-to-batch chemical consistency of botanical drugs. This method has been accepted by the world health organization (WHO), food and drug administration (FDA), and European medicines agency (EMA). In the process of development, newborn fingerprint strategy will encounter various problems. Quite a few of studies have proved that samples with high fingerprint-similarity values (> 0.95) do not always exhibit the expected equivalent efficacy. In other words, those high-content components in fingerprint do not maintain the predominance in the treatment. The significance of solo-fingerprint approach in evaluation of efficacy consistency is weakened. It is necessary to extend the fingerprint-activity relationship to discover the bio-active components in TCMs, especially for those herbs with phylogenetic relationships. Many scholars have proposed some approaches in

modeling between fingerprint and herbal efficacy, by which the active compounds in different extraction sites from herbs or Chinese patent medicines were clarified (Fig. 6). Moreover, the addition, subtraction/ removal of herbs or their dosage in prescriptions was utilized to modeling, through which the best combination with efficiency enhancement/ toxicity reduction was shown.

Nowadays, fingerprint- efficacy modeling is not only used to study the substance basis related to drug properties/efficacy, but also to screen the Q-markers in TCMs. For example, the potential Q-markers (antimicrobial, cytotoxic, anti-inflammatory and analgesic) of *N. sativa* oils were obtained from fingerprint- efficacy modeling (Shawky et al., 2018). Currently, the approaches (Zhang, Zheng, Ni, Li, & Li, 2018) includes: 1) methods of elaborate the relevance between components and efficacy, e.g. grey relational analysis, correlation analysis, Cluster analysis; 2) manners to measure the contribution of components to efficacy, e.g. multivariate linear regression (MLR) analysis, partial least-squares regression (PLSR) analysis and PCA; 3) Methods to find the main active ingredients by simplifying data structure, e.g. canonical correlation analysis, ANN, SVM etc. The feasibility of these approaches can be verified by the following examples, such as, PLSR and back propagation - artistic network modeling (BP-ANN) were used to screen out the Q-markers in *Sophora* flower-bud / *Sophora* flower (Wang, Xiong, et al., 2017). The similarity analysis (SA), HCA and PCA were used to divide *Emilia prenanthoidea* DC samples into two categories; and gray correlation analysis (GRA), PLSR and ANN were used to correlate the fingerprints with the anti-inflammatory activities of different samples, deeming three compounds as Q-markers (Jiang et al., 2018).

Numerous challenges are still occurring in fingerprint-efficiency modeling. First of all, the herbal medicines require explaining their pharmacological effects from animals, organs, cells and molecule

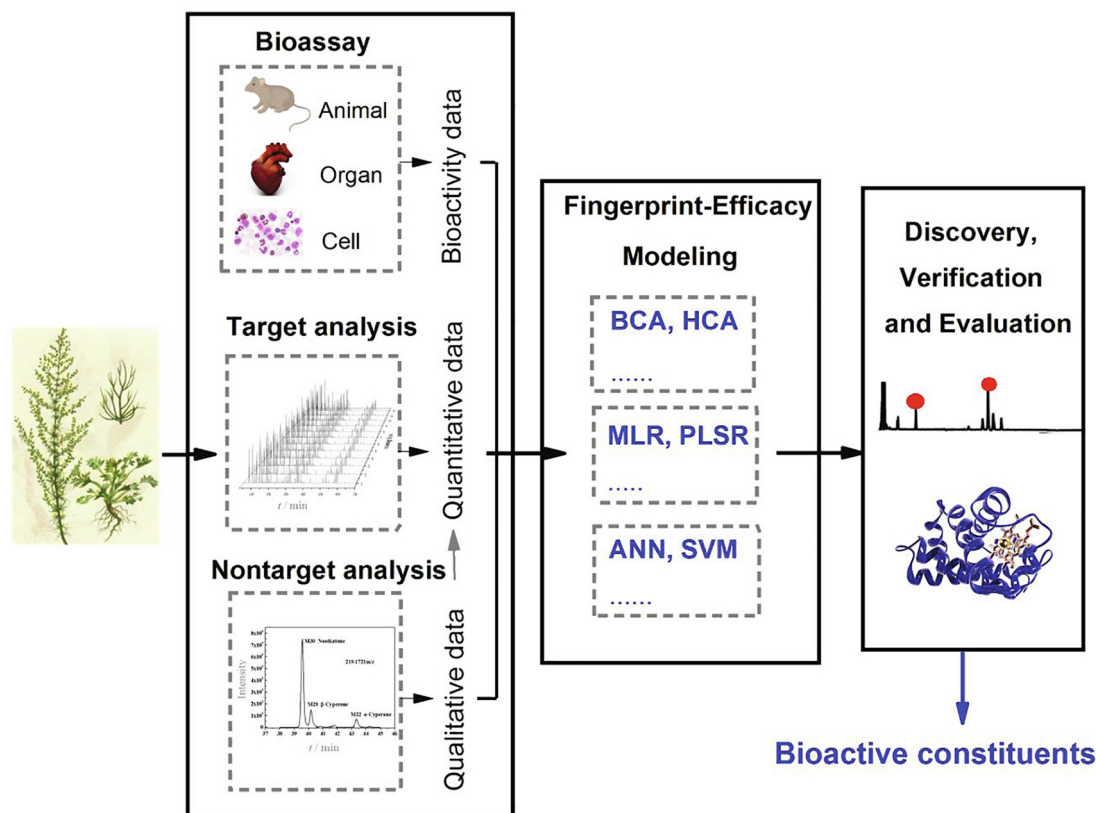


Fig. 6. Bio-active constituents discrimination through fingerprint- efficacy modeling.

levels, which is extremely challenging for the mathematical modeling. Secondly, there are no fixed mathematical-models between fingerprints and biological efficiency that can reflect their intermediate relationships. The chemometric methods with crossing application should be encouraged in constructing fingerprint- efficacy relationship, which can ensure accuracy to the extent mostly. Thirdly, the dialectic theory, such as “couplet medicines”, “monarch, minister, assistant and guide”, the mechanism of these still require researching. That is, the interactions probably exist between compounds and other active components or unknown ingredients in TCMs. How to find these inextricable connections from these complex samples? In the past, the fingerprints (*in vitro/in vivo*) were integrated into omics and network pharmacology. The new question is how to further establish their relationships from multi-level and multi-angle? Lastly, the fingerprint signals with biological activities needs to be verified still. Moreover, the relationships between drug properties and drug efficacy need to be elucidated by mathematical modeling, such as “five flavors and drug efficacy”, “effectiveness and properties are identical”, “similar effectiveness and different properties”. Gratifyingly, many new concepts have been put forward to these some problems, such as effect-constituent index (Xiong et al., 2018), molecular connectivity index (Liu, Zhang, et al., 2018), super Q markers (Li, Liu, et al., 2018).

6. Multidimensional data integration with fingerprints

Since the complexity of active ingredients is a significant feature in TCMs, only through the apparatus of chromatography-MS is unable to verify it. Therefore, plenty of approaches are thereby developed to disclose the mysterious material basis of active or

toxic effects. Moreover, different methods were utilized to explore the Q-markers from different angles or levels. It is necessary to develop multidimensional data integration approaches to identify ‘material basis and Q-markers’ more accurately for huge amount of data (Fig. 7). The scope of these data involves TCM theory, pharmacodynamics, pharmacokinetics, chemistry, mathematics, systems biology and so on. In addition, the conceptions of “whole and part”, “*in vivo* and *in vitro*”, “bioavailability and efficacy” were stressed in the data integration. In terms of research content, it comes down to chemical fingerprints, metabolic fingerprints, metabolic trajectory, network pharmacology, omics and pharmacodynamic data, etc. In order to identify ‘material basis and Q-markers’, some schemes have also been come up with to integrate data from multiple sources at this stage. It mainly includes three levels: 1) ‘material specificity, relevance and druggable’-based exploration; 2) ‘drug properties-efficacy’-based researches; 3) ‘chemical fingerprint - metabolic fingerprint - network targets’-based researches. All of these schemes are needed chemometrics tools, which can be seen in Fig. 7. As an example, GRA and least squares support vector machine (LS-SVM) techniques were used to integrate chemical and biosynthetic analysis, drug metabolism and network pharmacology, clarifying the Q-markers from Yuanhu Zhitong Tablets (Li, Li, et al., 2018). For other examples, chemical fingerprints, pharmacodynamics and network pharmacology were integrated to discover the Q-markers of *Alisma orientale* (Sam.) Juzep. (Liao et al., 2018); chemical fingerprints combined with network pharmacology were utilized to find the potential Q-markers which are related to arrhythmia from Shenxian Shengmai Oral Liquid (Xiang et al., 2018). Furthermore, a network of “toxicity - toxic chemical composition - toxic target - effect pathway” was constructed to predict the potential Q-markers, and the result could be verified by traceability and testability (Li et al., 2019d).

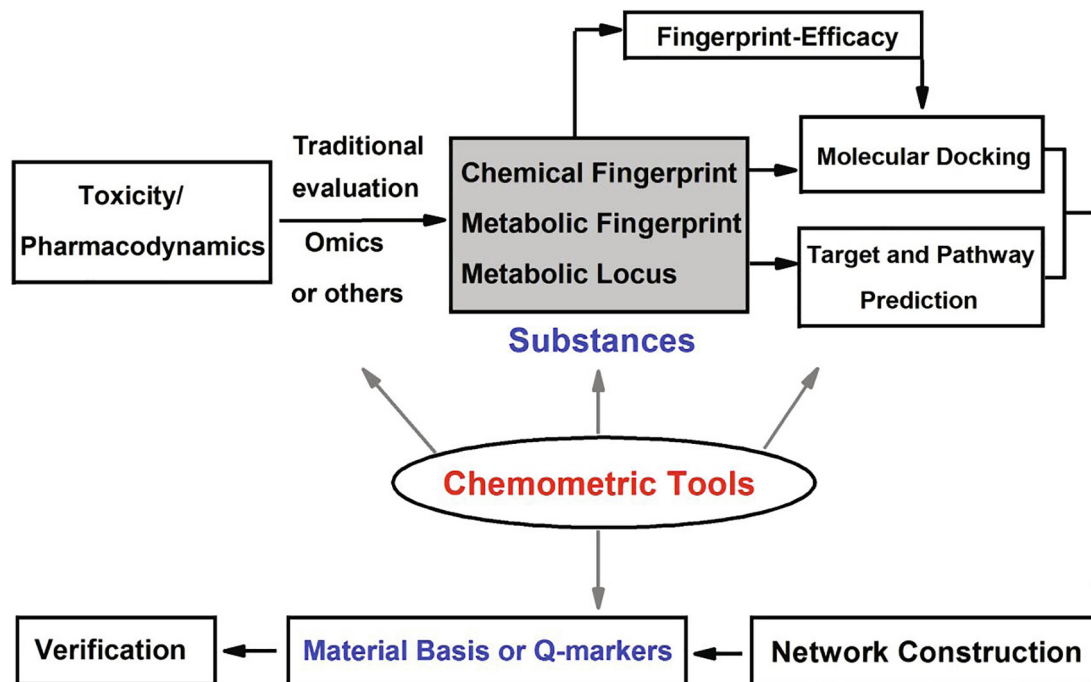


Fig. 7. Multidimensional data integration with herbal fingerprints.

System biology is permeated rapidly in pharmacological researches, such as, pharmacodynamic substance basis, pharmacological mechanism and Q-marker discovery in TCMs. Li Shao's team has established an online network-pharmacological platform, which has been committed to the mechanism research of TCMs (Zheng et al., 2018). Guo-an Luo proposed a strategy of new compound drugs (including compound herbal prescriptions, compound Western medicines and compound Chinese-Western medicines) based on the "system-system" (human system and the drug system) models (Luo, Wang, Fan, & Xie, 2018). Xi-jun Wang built a technical system of serum pharmacochemistry, which is widely used in the TCMs researches (Chen et al., 2018). Only by integrating the pharmacological activities (efficacy), network pharmacology (target) and metabolic fingerprints (component), the active components of the herbal prescriptions will be mined more thoroughly. Network pharmacology is a scientific method to construct and analyze network topology structure which is based on high-throughput omics, virtual computing and network database retrieval (Wang, Cui, et al., 2019). As to complex biological network analysis among "disease-disease", "target protein- drug" and "drug-drug", some experiments may be used to verify and evaluate the corresponding efficacy, adverse reactions and the action mechanism. However, the weakness of manpower, materials and finances is unable to satisfy numerous pharmacological experiments *in vivo/ in vitro* which conducting a large number of compounds/targets. Meanwhile, the target prediction and search tools are an important supplement (Tanoli et al., 2018), e.g. swiss target prediction server (<http://www.swisstargetprediction.ch>) (Gfeller et al., 2014) and PharmMapper server (<http://lilab.ecust.edu.cn/pharmmapper/>) (Wang, Shen, et al., 2017). The latter uses statistical methods to identify the targets, which is capable to predict the candidate targets for a given small molecule through the reverse pharmacophore mapping method. This molecular docking technology has also been widely used in the study of interaction between proteins and small molecules, which is based on energy matching calculation and molecular docking site configuration simulation. Additionally, SVM, RF and other methods have been used to predict the interactions between compounds and proteins. Some examples

can be used to explain their roles in TCMs, e.g. the candidate targets prediction of active ingredients from Yujin Fang (Tao et al., 2013); Autodock virtual screening to study HMG-CoA reductase inhibitors in compound Danshen preparations (Gai, Zhang, Ai, & Qiao, 2010). These research results indicated that the less cost and more reliability can be obtained through *in-silico* approaches. It is worthwhile pointing out that the compounds studied should not only come from a series of prediction values, but also from real metabolic fingerprint. Some scholars (Zhou, Yan, He, Hong, & Cao, 2019) have investigated compound-target-disease on herb-pair *Chuanxiong Rhizoma-Xiangfu Rhizoma* by means of 'metabolic fingerprint-network target' approach. Only through the multidimensional integration of "chemical fingerprint- metabolic fingerprint- network target- biological effect- TCMs efficacy", can the research results of 'material basis and Q-markers' be more accurate and reliable.

7. Conclusion

In this paper, chemometric tools in the chromatography- MS related fields are systematically reviewed, which will bring a profound effect on the 'material basis and Q-markers' exploration from TCMs. These smart tools have been widely used in design-modeling -optimization, calibration in *rt* and *m/z* dimensions, resolution of co-eluted peaks, fingerprint-data grouping, fingerprint-efficacy modeling, and multidimensional-data integration, etc. After data procession, the precise qualitative and quantitative results in chemical (metabolic) fingerprints and metabolic trajectories can ensure the reliability of 'material basis and Q-markers' results. Certainly, these tools can also be introduced into the omics and fingerprint- efficiency modeling researches. When the significant components that affect the drug properties/efficacy needs to be found, pattern recognition and variable selection react on fingerprint related data-sets. Usually the data only obtained by chromatography-related technology cannot disclose the material basis or Q- markers in TCMs fully. Only through the deep integration of multi-dimensional data, the mysterious herbs can be clarified from multiple levels and multiple perspectives. With the

development of chemometric tools, the drug properties/efficacy-related components and Q-markers will be identified more accurately, laying a solid foundation for the internationalization and modernization of TCMs.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Acknowledgements

This work is financially supported by Hunan 2011 Collaborative Innovation Center of Chemical Engineering & Technology with Environmental Benignity and Effective Resource Utilization, Hunan Province Natural Science Fund (No. 2016JJ4085, 2020JJ4569), the Key Project of Hunan Provincial Education Department (18A055), the Open Project Program of the Chongqing TCM Key Laboratory for Metabolic Disease (Chongqing Medical University). The studies meet with the approval of the university's review board.

References

- Allard, P. M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., et al. (2016). Integration of molecular networking and *In-Silico* MS/MS fragmentation for natural products dereplication. *Analytical Chemistry*, 88, 3317–3323.
- Alvarez-Rivera, G., Ballesteros-Vivas, D., Parada-Alfonso, F., Ibañez, E., & Cifuentes, A. (2019). Recent applications of high resolution mass spectrometry for the characterization of plant natural products. *TrAC Trends in Analytical Chemistry*, 112, 87–101.
- Amos, R. I. J., Haddad, P. R., Szucs, R., Dolan, J. W., & Christopher, A. P. (2018). Molecular modeling and prediction accuracy in Quantitative Structure-Retention Relationship calculations for chromatography. *TrAC Trends in Analytical Chemistry*, 105, 352–359.
- Aspromonte, J., Wolfs, K., & Adams, E. (2019). Current application and potential use of GC×GC in the pharmaceutical and biomedical field. *Journal of Pharmaceutical and Biomedical Analysis*, 176, 112817.
- Babushok, V. I. (2015). Chromatographic retention indices in identification of chemical compounds. *TrAC Trends in Analytical Chemistry*, 69, 98–104.
- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. *PLS-DA. Analytical Methods*, 5(16), 3790.
- Bian, X. Q., Zhang, Q., Zhang, L., & Chen, J. (2017). A grey wolf optimizer-based support vector machine for the solubility of aromatic compounds in supercritical carbon dioxide. *Chemical Engineering Research and Design*, 123, 284–294.
- Bilbao, A., Gibbons, B. C., Slys, G. W., Crowell, K. L., Monroe, M. E., Ibrahim, Y. M., et al. (2018). An algorithm to correct saturated mass spectrometry ion abundances for enhanced quantitation and mass accuracy in omic studies. *International Journal of Mass Spectrometry*, 427, 91–99.
- Bloembergen, T. G., Gerretzen, J., Wouters, H. J. P., Gloerich, J., van Dael, M., Wessels, H. J. C. T., et al. (2010). Improved parametric time warping for proteomics. *Chemometrics & Intelligent Laboratory Systems*, 104(1), 65–74.
- Bouwmeester, R., Martens, L., & Degroev, S. (2019). Comprehensive and empirical evaluation of machine learning algorithms for small molecule LC retention time prediction. *Analytical Chemistry*, 91(5), 3694–3703.
- Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics & Intelligent Laboratory Systems*, 38(2), 149–171.
- Bro, R., Andersson, C. A., & Kiers, H. A. L. (2010). PARAFAC2—Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics*, 13(3–4), 295–309.
- Cabeza, A., Sobrón, F., García-Serna, J., & Cocero, M. J. (2016). Simulation of the supercritical CO₂ extraction from natural matrices in packed bed columns: User-friendly simulator tool using Excel. *The Journal of Supercritical Fluids*, 116, 198–208.
- Cappellin, L., Biasioli, F., Fabris, A., Schuhfried, E., Soukoulis, C., Märk, T. D., et al. (2010). Improved mass accuracy in PTR-TOF-MS: Another step towards better compound identification in PTR-MS. *International Journal of Mass Spectrometry*, 290, 60–63.
- Carabajal, M., Teglia, C. M., Cerutti, S., Culzoni, M. J., & Goicoechea, H. C. (2019). Applications of liquid-phase microextraction procedures to complex samples assisted by response surface methodology for optimization. *Microchemical Journal*, 152, 104436.
- Chervin, J., Stierhof, M., Tong, M. H., Peace, D., Hansen, K., Urgast, D. S., et al. (2017). Targeted dereplication of microbial natural products by high-resolution MS and predicted LC retention time. *Journal of Natural Products*, 80(5), 1370–1377.
- Chen, X. J., Yang, M. L., Liu, W., You, P. T., Zhou, A. J., Liu, Y. W., et al. (2018). Studies on serum pharmacokinetics of effective parts of modified Xiaochaihu Tang for treatment of gastric ulcer. *China Journal of Chinese Materia Medica*, 43(8), 1692–1700.
- Chen, Y. Y., Tang, Y. P., Shang, E. X., Zhu, Z. H., Tao, W. W., Yu, J. G., et al. (2019). Incompatibility assessment of *Genkwa Flos* and *Glycyrrhizae Radix et Rhizoma* with biochemical, histopathological and metabolomic approach. *Journal of Ethnopharmacology*, 229, 222–232.
- Cox, J., & Mann, M. (2009). Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *Journal of American Society for Mass Spectrometry*, 20, 1477–1485.
- Dalmou, N., Bedia, C., & Tauler, R. (2018). Validation of the regions of interest multivariate curve resolution (ROIMCR) procedure for untargeted LC-MS lipidomic analysis. *Analytica Chimica Acta*, 1025, 80–91.
- de Boer, W. P. H., & Lankelma, J. (2014). Two-dimensional semi-parametric alignment of chromatograms. *Journal of Chromatography A*, 1345, 193–199.
- Djombou-Feunang, Y., Pon, A., Karu, N., Zheng, J., Li, C., Arndt, D., et al. (2019). CFM-ID 3.0: Significantly improved ESI-MS/MS prediction and compound identification. *Metabolites*, 9(4), 72.
- Doherty, W., Grégoire, D. C., & Bertrand, N. (2008). Polynomial mass bias functions for the internal standardization of isotope ratio measurements by multi-collector inductively coupled plasma mass spectrometry. *Spectrochim Acta B*, 63, 407–414.
- Dossin, E., Martin, E., Diana, P., Castellon, A., Monge, A., Pospisil, P., et al. (2016). Prediction models of retention indices for increased confidence in structural elucidation during complex matrix analysis: Application to gas chromatography coupled with high-resolution mass spectrometry. *Analytical Chemistry*, 88(15), 7539–7547.
- Doucette, A. A., & Chisholm, R. A. (2019). Molecular-formula determination through accurate-mass analysis: A forensic investigation. *Journal of Chemical Education*, 96(7), 1458–1464.
- Dührkop, K., Shen, H., Meusel, M., Rousu, J., & Böcker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences of the USA*, 112(41), 12580–12585.
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, 16, 299–302.
- Faber, N. M. (2001). The price paid for the second-order advantage when using the generalized rank annihilation method (GRAM). *Journal of Chemometrics*, 15(9), 743–748.
- Fu, H. Y., Hu, O., Zhang, Y. M., Zhang, L., Song, J. J., Lu, P., et al. (2017). Mass-spectra-based peak alignment for automatic nontargeted metabolic profiling analysis for biomarker screening in plant samples. *Journal of Chromatography A*, 1513, 201–209.
- Fu, L., Ding, H., Han, L., Jia, L. W., Yang, C., Zhang, Y., et al. (2019). Simultaneously targeted and untargeted multicomponent characterization of Erzhi Pill by offline two-dimensional liquid chromatography/quadrupole-Orbitrap mass spectrometry. *Journal of Chromatography A*, 1584, 87–96.
- Furbo, S., Hansen, A. B., Skov, T., & Christensen, J. H. (2014). Pixel-based analysis of comprehensive two-dimensional gas chromatograms (Color Plots) of petroleum: A tutorial. *Analytical Chemistry*, 86(15), 7160–7170.
- Gai, W., Zhang, Y. L., Ai, L., & Qiao, Y. J. (2010). Screening of HMG-CoA reductase inhibitors from composite *Salvia miltiorrhiza* using Autodock. *Chinese Journal of Natural Medicines*, 8(1), 51–56.
- Gan, F., Ruan, G., & Mo, J. (2006). Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemometrics and Intelligent Laboratory Systems*, 82(1–2), 59–65.
- Ganzera, M., & Sturm, S. (2018). Recent advances on HPLC/MS in medicinal plant analysis—An update covering 2011–2016. *Journal of Pharmaceutical and Biomedical Analysis*, 147, 211–233.
- Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., & Zoete, V. (2014). Swiss target prediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Research*, 42(W1), W32–W38.
- Goudarzi, N., Shahsavani, D., Emadi-Gandaghi, F., & Arab Chamjangali, M. (2014). Application of random forests method to predict the retention indices of some polycyclic aromatic hydrocarbons. *Journal of Chromatography A*, 1333, 25–31.
- Graczyk, D. G., McLain, D. R., Tsai, Y., Chamberlain, D. B., & Steeb, J. L. (2019). Correcting nonlinearity and mass-bias in measurements by inductively coupled plasma quadrupole mass spectrometry. *Spectrochimica Acta Part B*, 153, 10–18.
- Gröger, T. M., Käfer, U., & Zimmermann, R. (2020). Gas chromatography in combination with fast high-resolution time-of-flight mass spectrometry: Technical overview and perspectives for data visualization. *TrAC Trends in Analytical Chemistry*, 122, 115677.
- Gu, Y. Y., Shi, L., Zhang, D. D., Huang, X., & Chen, D. Z. (2018). Metabonomics delineates allergic reactions induced by Shuang-huang-lian injection in rats using ultra performance liquid chromatography-mass spectrometry. *Chinese Journal of Natural Medicines*, 16(8), 628–640.
- Guan, Z. B., Wang, M., Cai, Y., Yang, H. M., Zhao, M., & Zhao, C. J. (2018). Rapid characterization of the chemical constituents of Sijunzi decoction by UHPLC coupled with Fourier transform ion cyclotron resonance mass spectrometry. *Journal of Chromatography B*, 1086, 11–22.
- Guo, L., Gong, M., Wu, S., Qiu, F., & Ma, L. (2020). Identification and quantification of the quality markers and anti-migraine active components in *Chuanxiong Rhizoma* and *Cyperus Rhizoma* herbal pair based on chemometric analysis between chemical constituents and pharmacological effects. *Journal of Ethnopharmacology*, 246, 112228.
- He, M., Hong, L., Yang, Z. Y., Yang, T. B., & Zeng, J. (2018). Bioactive assay and hyphenated chromatography detection for complex supercritical CO₂ extract from Chaihu Shugan San using an experimental design approach. *Microchemical Journal*, 142, 394–402.

- He, M., Hong, L., & Zhou, Y. (2019a). Multi-scale Gaussian/Haar wavelet strategies coupled with sub-window factor analysis for an accurate alignment in nontargeted metabolic profiling to enhance herbal origin discrimination capability. *Journal of Separation Science*, 42(11), 2003–2012.
- He, M., Nie, J., Wu, H., & Liang, Y. (2015). Accurate mass determination of platycosides prepared by macroporous resins using high-resolution mass spectroscopy and optimization methods. *Chromatographia*, 78, 683–693.
- He, M., Peng, G., Xie, F., Hong, L., & Cao, Q. (2019b). Liquid chromatography - high resolution - mass spectrometry with ROI strategy for non-targeted analysis of the *in vivo*/ *in vitro* ingredients coming from *Ligusticum chuanxiong* hort. *Chromatographia*, 82(7), 1069–1077.
- He, M., Wu, H., Nie, J., Yan, P., Yang, T. B., Yang, Z. Y., et al. (2017). Accurate recognition and feature qualify for flavonoid extracts from Liang-wai Gan Cao by liquid chromatography-high resolution-mass spectrometry and computational MS/MS fragmentation. *Journal of Pharmaceutical and Biomedical Analysis*, 146, 37–47.
- He, M., Yan, J., Cao, D., Liu, S., Zhao, C., Liang, Y., et al. (2013). Identification of terpenoids from *Ephedra* combining with accurate mass and *in-silico* retention indices. *Talanta*, 103, 116–122.
- He, M., Yan, P., Yang, Z. Y., Zhang, Z. M., Yang, T. B., & Hong, L. (2018). A modified multiscale peak alignment method combined with trilinear decomposition to study the volatile/ heat-labile components in *Ligusticum chuanxiong* hort-*Cyperus rotundus* rhizomes by HS-SPME-GC/MS. *Journal of Chromatography B*, 1079, 41–50.
- He, M., Yan, P., Yang, Z. Y., Ye, Y., Cao, D., Hong, L., et al. (2018). Multi-analytical strategy for unassigned peaks using physical/mathematical separation, fragmental rules and retention index prediction: An example of sesquiterpene metabolites characterization in *Cyperus rotundus*. *Journal of Pharmaceutical and Biomedical Analysis*, 154, 476–485.
- He, M., Yang, Z. Y., Yang, T. B., Ye, Y., Nie, J., Hu, Y., et al. (2017). Chemometrics-enhanced one-dimensional/ comprehensive two-dimensional gas chromatographic analysis for bioactive terpenoids and pthalides in Chaihu Shugan San essential oils. *Journal of Chromatography B*, 1052, 158–168.
- He, S., Fang, S., Liu, X., Zhang, W., Xie, W., Zhang, H., et al. (2016). Investigation of a genetic algorithm based cubic spline smoothing for baseline correction of Raman spectra. *Chemometrics and Intelligent Laboratory Systems*, 152, 1–9.
- Hibbert, D. B. (2012). Experimental design in chromatography: A tutorial review. *Journal of Chromatography B*, 910, 2–13.
- Hong, L., Li, Y., He, M., Zhao, C., & Li, M. (2020). An algorithm to calibrate ionic isotopes using data mining strategy in hyphenated chromatographic datasets from herbal samples. *Journal of Chromatography A*, 1613, 460668.
- Huang, L., Xie, D., Yu, Y., Liu, H., Shi, Y., Shi, T., et al. (2018). TCMID 2.0: A comprehensive resource for TCM. *Nucleic Acids Research*, 46(D1), D1117–D1120.
- Izadmanesh, Y., Garreta-Lara, E., Ghasemi, J. B., Lacorte, S., Matamoros, V., & Tauler, R. (2017). Chemometric analysis of comprehensive two dimensional gas chromatography-mass spectrometry metabolomics data. *Journal of Chromatography A*, 1488, 113–125.
- Jaramillo, R., & Dorman, F. L. (2018). Retention time prediction in thermally modulated comprehensive two-dimensional gas chromatography: Correcting second dimension retention time modeling error. *Journal of Chromatography A*, 1581–1582, 116–124.
- Jaramillo, R., & Dorman, F. L. (2019). Retention time prediction of hydrocarbons in cryogenically modulated comprehensive two-dimensional gas chromatography: A method development and translation application. *Journal of Chromatography A*, 2019–460696.
- Ji, H., Xu, Y., Lu, H., & Zhang, Z. (2019). Deep MS/MS-aided structural-similarity scoring for unknown metabolite identification. *Analytical Chemistry*, 91(9), 5629–5637.
- Jiang, M., Kulsing, C., Nolvachai, Y., & Marriott, P. J. (2015). Two-dimensional retention indices improve component identification in comprehensive two-dimensional gas chromatography of saffron. *Analytical Chemistry*, 87(11), 5753–5761.
- Jiang, M. (2019). Facile approach for calculation of second dimensional retention indices in comprehensive two dimensional gas chromatography with single injection. *Analytical Chemistry*, 91(6), 4085–4091.
- Jiang, W., & Erve, J. C. L. (2012). Spectral accuracy of a new hybrid quadrupole time-of-flight mass spectrometer: Application to ranking small molecule elemental compositions. *Rapid Communications in Mass Spectrometry*, 26(9), 1014–1022.
- Jiang, Z., Zhao, C., Gong, X., Sun, X., Li, H., Zhao, Y., et al. (2018). Quantification and efficient discovery of quality control markers for *Emilia prenanthoidea* DC. by Fingerprint-Efficacy Relationship Modelling. *Journal of Pharmaceutical and Biomedical Analysis*, 156, 36–44.
- Kang, J., Choi, M. Y., Kang, S., Kwon, H. N., Wen, H., Lee, C. H., et al. (2008). Application of a 1H nuclear magnetic resonance (NMR) metabolomics approach combined with orthogonal projections to Latent Structure-Discriminant Analysis as an efficient tool for discriminating between Korean and Chinese herbal medicines. *Journal of Agricultural and Food Chemistry*, 56(24), 11589–11595.
- Karri, R. R., Sahu, J. N., & Meikap, B. C. (2020). Improving efficacy of Cr (VI) adsorption process on sustainable adsorbent derived from waste biomass (sugarcane bagasse) with help of ant colony optimization. *Industrial Crops and Products*, 143, 111927.
- Kassidas, A., Macgregor, J. F., & Taylor, P. A. (1998). Synchronization of batch trajectories using dynamic time warping. *Aiche Journal*, 44(4), 864–875.
- Kelly, K., Brooks, S., & Bell, S. (2019). The effect of mass spectrometry tuning frequency and criteria on ion relative abundances of cathinones and cannabinoids. *Forensic Chemistry*, 12, 58–65.
- Kiers, H. A. L., ten Berge, J. M. F., & Bro, R. (2010). PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics*, 13(3–4), 275–294.
- Krishnan, S., Vogels, J. T. W. E., Coulier, L., Bas, R. C., Hendriks, M. W. B., Hankemeier, T., et al. (2012). Instrument and process independent binning and baseline correction methods for liquid chromatography-high resolution-mass spectrometry deconvolution. *Analytica Chimica Acta*, 740, 12–19.
- Kulkarni, O., & Kulkarni, S. (2018). Process parameter optimization in WEDM by Grey Wolf Optimizer. *Materials Today: Proceedings*, 5(2), 4402–4412.
- Kumar, K., & Mishra, A. K. (2015). Multivariate curve resolution alternating least square (MCR-ALS) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets: Comparing certain ways of arranging TSFS-based three-way array. *Chemometrics and Intelligent Laboratory Systems*, 147, 66–74.
- Lau, H., Liu, S. Q., Tan, L. P., Lassabliere, B., Sun, J., & Yu, B. (2019). A systematic study of molecular ion intensity and mass accuracy in low energy electron ionisation using gas chromatography-quadrupole time-of-flight mass spectrometry. *Talanta*, 199, 431–441.
- Li, K., Li, J., Su, J., Xiao, X., Peng, X., Liu, F., et al. (2018). Identification of quality markers of Yuanhu Zhitong tablets based on integrative pharmacology and data mining. *Phytomedicine*, 44, 212–219.
- Li, M., & Wang, X. R. (2019b). Peak alignment of gas chromatography-mass spectrometry data with deep learning. *Journal of Chromatography A*, 1604, 460476.
- Li, Y., Zhang, Y., Wang, Y., Li, Y., Yang, F., Zhang, P., et al. (2019d). A strategy for the discovery and validation of toxicity quality marker of Chinese medicine based on network toxicology. *Phytomedicine*, 54, 365–370.
- Li, Y. Z., Xie, Y., He, Y. F., Hou, W. B., Liao, M. L., & Liu, C. X. (2019a). Quality markers of traditional Chinese medicine: Concept, progress, and perspective. *Engineering*, 5(5), 888–894.
- Li, Y., Kuhn, M., Gavin, A. C., & Bork, P. (2019c). Identification of metabolites from tandem mass spectra with a machine learning approach utilizing structural features. *Bioinformatics*, pii, btz736.
- Li, Z. Q., Liu, J., Li, Y. Z., Du, X., Li, Y. F., Wang, R. H., et al. (2018). Identify super quality markers from prototype-based pharmacokinetic markers of Tangzhiqing tablet (TZQ) based on *in vitro* dissolution/ permeation and *in vivo* absorption correlations. *Phytomedicine*, 45, 59–67.
- Liang, Y. Z., & Kvalheim, O. M. (1992). Heuristic evolving latent projections: Resolving two-way multicomponent data. *Analytical Chemistry*, 64(8), 936–946.
- Liang, Y. Z., Xie, P., & Chan, K. (2004). Quality control of herbal medicines. *Journal of Chromatography B*, 812(1–2), 53–70.
- Liang, Y. Z., Xie, P. S., & Chan, K. (2010). Perspective of chemical fingerprinting of Chinese herbs. *Planta Medica*, 76(17), 1997–2003.
- Liang, Y. Z., Wu, H. L., & Yu, R. Q. (2016). *Handbook of analytical chemistry*. 10. Chemometrics. Chemical Industry Press.
- Liao, M., Shang, H., Li, Y., Li, T., Wang, M., Zheng, Y., et al. (2018). An integrated approach to uncover quality marker underlying the effects of *Alisma orientale* on lipid metabolism, using chemical analysis and network pharmacology. *Phytomedicine*, 45, 93–104.
- Lin, Y., Zong, Z., Tian, S., & Lin, J. (2018). A new baseline correction method for near-fault strong-motion records based on the target final displacement. *Soil Dynamics and Earthquake Engineering*, 114, 27–37.
- Linder, M., & Sundberg, R. (1998). Second-order calibration: Bilinear least squares regression and a simple alternative. *Chemometrics & Intelligent Laboratory Systems*, 42(1–2), 159–178.
- Liu, P., Shang, E., Zhu, Y., Qian, D., & Duan, J. (2018). Volatile component interaction effects on compatibility of *Cyperus Rhizoma* and *Angelicae Sinensis Radix* or *Chuanxiong Rhizoma* by UPLC-MS/MS and response surface analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 160, 135–143.
- Liu, X., Zhang, Z., Liang, Y., Sousa, P. F. M., Yun, Y., & Yu, L. (2014). Baseline correction of high resolution spectral profile data based on exponential smoothing. *Chemometrics and Intelligent Laboratory Systems*, 139, 97–108.
- Liu, S., Liang, Y. Z., & Liu, H. T. (2016). Chemometrics applied to quality control and metabolomics for traditional Chinese medicines. *Journal of chromatography B*, 1015–1016, 82–91.
- Liu, C. X., Cheng, Y. Y., Guo, D. A., Zhang, T. J., Li, Y. Z., Hou, W. B., et al. (2017). A new concept on quality marker for quality assessment and process control of Chinese medicines. *Chinese Herbal Medicines*, 9(1), 3–13.
- Liu, F.-J., Jiang, Y., Li, P., Liu, Y.-D., Yao, Z.-P., Xin, G.-Z., & Li, H.-J. (2020). Untargeted metabolomics coupled with chemometric analysis reveals species-specific steroidal alkaloids for the authentication of medicinal *Fritillariae Bulbus* and relevant products. *Journal of Chromatography A*, 1612, 460630.
- Liu, H., Yin, Z., Zhang, Q., Li, X., Tang, K., Liu, J., et al. (2019). Mathematical modeling of bovine hides swelling behavior by response surface methodology for minimization of sulfide pollution in leather manufacture. *Journal of Cleaner Production*, 237, 117800.
- Liu, W. L., Zhang, X. L., Fan, S. Q., Zhu, J. P., Liang, H. H., Zhang, Y. T., et al. (2018). A novel concept of Q-markers: Molecular connectivity index. *Phytomedicine*, 45, 36–40.
- Lochmuller, C. H. (1995). Predicting retention in LC. *Analytical Chemistry*, 67(21), 660A.
- Luan, F., Xue, C., Zhang, R., Zhao, C., Liu, M., Hu, Z., et al. (2005). Prediction of retention time of a variety of volatile organic compounds based on the heuristic method and support vector machine. *Analytica Chimica Acta*, 537(1–2), 101–110.

- Luo, G. A., Wang, Y. M., Fan, X. M., & Xie, Y. Y. (2018). Research strategy and practice from clinical reality, targeting at signal pathway for the innovative compound drug-The sixth discussion on the proposal of holistic systems medicine. *Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*, 20(7), 11–32.
- Ma, Y., Kind, T., Vaniya, A., Gennity, I., & Fahrman, J. F. (2015). An *in silico* MS/MS library for automatic annotation of novel FAHFA lipids. *Journal of Cheminformatics*, 7, 53.
- Matsuo, T., Tsugawa, H., Miyagawa, H., & Fukusaki, E. (2017). Integrated strategy for unknown EI-MS identification using quality control calibration curve, multivariate analysis, EI-MS spectral database, and retention index prediction. *Analytical Chemistry*, 89(12), 6766–6773.
- Matyushin, D. D., Sholokhova, A. Y., & Buryak, A. K. (2019). A deep convolutional neural network for the estimation of gas chromatographic retention indices. *Journal of Chromatography A*, 1607, 460395.
- Mazur, D. M., Zenkevich, I. G., Artaev, V. B., Polyakova, O. V., & Lebedev, A. T. (2018). Regression algorithm for calculating second-dimension retention indices in comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, 1569, 178–185.
- Mihaleva, V. V., Vorst, O., Maliepaard, C., Verhoeven, H. A., de Vos, R. C. H., Hall, R. D., et al. (2008). Accurate mass error correction in liquid chromatography time-of-flight mass spectrometry based metabolomics. *Metabolomics*, 4, 171–182.
- Mitchell, B. C., & Burdick, D. S. (2010). Slowly converging parafac sequences: Swamps and two-factor degeneracies. *Journal of Chemometrics*, 8(2), 155–168.
- Miti, M., Tošić, S., Pavlović, A., Mašković, P., Kostić, D., Mitić, J., et al. (2019). Optimization of the extraction process of minerals from *Salvia officinalis* L. using factorial design methodology. *Microchemical Journal*, 145, 1224–1230.
- Mohammad Munawar, T., Surya Prakash, D. V., & Vangalapati, M. (2018). Development of response surface methodology for optimization of parameters and quantitative analysis of chebulinic acid from composition of medicinal herbs by HPLC. *Saudi Journal of Biological Sciences*, 26(7), 1809–1814.
- Mokhtari, L., & Ghoreishi, S. M. (2019). Supercritical carbon dioxide extraction of trans-anethole from *Foeniculum vulgare* (fennel) seeds: Optimization of operating conditions through response surface methodology and genetic algorithm. *Journal of CO2 Utilization*, 30, 1–10.
- Mousavi, L., Tamiji, Z., & Khoshayand, M. R. (2018). Applications and opportunities of experimental design for the dispersive liquid-liquid microextraction method – A review. *Talanta*, 190, 335–356.
- Narendran, S. T., Meyyanathan, S. N., & ReddyKarri, V. V. S. (2019). Experimental design in pesticide extraction methods: A review. *Food Chemistry*, 289, 384–395.
- Navarro-Reig, M., Jaumot, J., van Beek, T. A., Vivó-Truyols, G., & Tauler, R. (2016). Chemometric analysis of comprehensive LC×LC-MS data: Resolution of triacylglycerol structural isomers in corn oil. *Talanta*, 160, 624–635.
- Navarro-Reig, M., Jaumot, J., & Tauler, R. (2018). An untargeted lipidomic strategy combining comprehensive two-dimensional liquid chromatography and chemometric analysis. *Journal of Chromatography A*, 1568, 80–90.
- Ni, Y., Mei, M., & Kokot, S. (2012). One- and two-dimensional gas chromatography-mass spectrometry and high performance liquid chromatography-diode-array detector fingerprints of complex substances: A comparison of classification performance of similar, complex *Rhizoma Curcumae* samples with the aid of chemometrics. *Analytica Chimica Acta*, 712, 37–44.
- Palmblad, M., Mills, D. J., Bindschedler, L. V., & Cramer, R. (2007). Chromatographic alignment of LC-MS and LC-MS/MS datasets by genetic Algorithm Feature Extraction. *Journal of the American Society for Mass Spectrometry*, 18(10), 1835–1843.
- Parastar, H., Jalali-Heravi, M., & Tauler, R. (2012). Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation induced shifting and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 117, 80–91.
- Prebihalo, S. E., Berrier, K. L., Freye, C. E., Bahaghighat, H. D., Moore, N. R., Pinkerton, D. K., et al. (2018). Multidimensional gas chromatography: Advances in instrumentation, chemometrics, and applications. *Analytical Chemistry*, 90(1), 505–532.
- Qi, D., Miao, A., Cao, J., Wang, W., Chen, W., Pan, S., 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 Chemistry*, 265, 189–199.
- Qian, F., Wu, Y., & Hao, P. (2017). A fully automated algorithm of baseline correction based on wavelet feature points and segment interpolation. *Optics & Laser Technology*, 96, 202–207.
- Ràfols, P., del Castillo, E., Yanes, O., Brezmes, J., & Correig, X. (2018). Novel automated workflow for spectral alignment and mass calibration in MS imaging using a sputtered Ag nanolayer. *Analytical Chimica Acta*, 1022, 61–69.
- Rajadurai, P., & Sankaranarayanan, S. (2012). A Fast Quad-Tree based two dimensional hierarchical clustering. *Bioinformatics and Biology Insights*, 6(6), 265–274.
- Randazzo, G. M., Bileck, A., Danani, A., Vogt, B., & Groessl, M. (2019). Steroid identification via deep learning retention time predictions and two-dimensional gas chromatography-high resolution mass spectrometry. *Journal of Chromatography A*, 2019–460661.
- Russo, S., Li, G., & Villez, K. (2019). Automated model selection in principal component analysis: A new approach based on the cross-validated ignorance score. *Industrial & Engineering Chemistry Research*, 58(30), 13448–13468.
- Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., & Neumann, S. (2016). MetFrag relaunched: Incorporating strategies beyond *in silico* fragmentation. *Journal of Cheminformatics*, 8, 3.
- Sadygov, R. G., Maroto, F. M., & Hühmer, A. F. (2006). ChromAlign: A two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces. *Analytical Chemistry*, 78(24), 8207–8217.
- Sammon, J. W. (1969). *A nonlinear mapping for data structure analysis*. IEEE Computer Society.
- Sanchez, E., & Kowalski, B. R. (1990). Tensorial resolution: A direct trilinear decomposition. *Journal of Chemometrics*, 4(1), 29–45.
- Sawale, P. D., Patil, G. R., Hussain, S. A., Singh, A. K., & Singh, R. R. B. (2020). Development of free and encapsulated Arjuna herb extract added vanilla chocolate dairy drink by using response surface methodology (RSM) software. *Journal of Agriculture and Food Research*, 2, 100020.
- Sawall, M., Harbou, E., Moog, A., Behrens, R., Schröder, H., Simoneau, J., et al. (2018). Multi-objective optimization for an automated and simultaneous phase and baseline correction of NMR spectral data. *Journal of Magnetic Resonance*, 289, 132–141.
- Sharif, K. M., Rahman, M. M., Azmir, J., Mohamed, A., Jahurul, M. H. A., Sahena, F., et al. (2014). Experimental design of supercritical fluid extraction – A review. *Journal of Food Engineering*, 124, 105–116.
- Shawky, E., El Newehy, N. M., Beltagy, A. M., Abd-Elhaseeb, M. M., Omran, G. A., & Harraz, F. M. (2018). Fingerprint profile and efficacy-associated markers of *Nigella sativa* oil for geographical origin determination using targeted and untargeted HPTLC - multivariate analysis. *Journal of Chromatography B*, 1087–1088, 108–117.
- Sodeifan, G., Ardestani, N. S., Sajadian, S. A., & Ghorbandoost, S. (2016). Application of supercritical carbon dioxide to extract essential oil from *Cleome coluteoides* Boiss: Experimental, response surface and grey wolf optimization methodology. *The Journal of Supercritical Fluids*, 114, 55–63.
- Soepangkat, B. O. P., Norcahyo, R., Effendi, M. K., & Pramujati, B. (2019). Multi-response optimization of carbon fiber reinforced polymer (CFRP) drilling using back propagation neural network-particle swarm optimization (BPNN-PSO). *Engineering Science and Technology*, 23, 700–713.
- Stanstrup, J., Neumann, S., & Vrhovšek, U. (2015). PredRet: Prediction of retention time by direct mapping between multiple chromatographic systems. *Analytical Chemistry*, 87(18), 9421–9428.
- Steinmann, D., & Ganzer, M. (2011). Recent advances on HPLC/MS in medicinal plant analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 55(4), 744–757.
- Stoll-Werian, A., Flierl, L., Rienitz, O., Noordmann, J., Kessel, R., & Prammann, A. (2019). Absolute isotope ratios – Analytical solution for the determination of calibration factors for any number of isotopes and isotopologues. *Spectrochim Acta B*, 157, 76–83.
- Sun, H., Zhang, A. H., Yang, L., Li, M. X., Fang, H., Xie, J., et al. (2019). High-throughput chinmedomics strategy for discovering the quality-markers and potential targets for Yinchenhao decoction. *Phytomedicine*, 54, 328–338.
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information & Computer Sciences*, 43(6), 1947–1958.
- Tan, B. B., Liang, Y. Z., & Yi, L. Z. (2010). Identification of free fatty acids profiling of type 2 diabetes mellitus and exploring possible biomarkers by GC-MS coupled with chemometrics. *Metabolomics*, 6(2), 219–228.
- Tanoli, Z., Alam, Z., Vähä-Koskela, M., Ravikumar, B., Malyutina, A., Jaiswal, A., et al. (2018). Drug Target Commons 2.0: A community platform for systematic analysis of drug-target interaction profiles. *Database (Oxford)*, 2018, bay083.
- Tao, W., Xu, X., Wang, X., Li, B., Wang, Y., Li, Y., et al. (2013). Network pharmacology-based prediction of the active ingredients and potential targets of Chinese herbal *Radix Curcumae* formula for application to cardiovascular disease. *Journal of Ethnopharmacology*, 145(1), 1–10.
- Taraji, M., Haddad, P. R., Amos, R. I. J., Talebi, M., Szucs, R., Dolan, J. W., et al. (2018). Chemometric-assisted method development in hydrophilic interaction liquid chromatography: A review. *Analytica Chimica Acta*, 1000, 20–40.
- Tian, T. F., Wang, S. Y., Kuo, T. C., Tan, C. E., Chen, G. Y., Kuo, C. H., et al. (2016). Web server for peak detection, baseline correction, and alignment in two-dimensional gas chromatography mass spectrometry-based metabolomics data. *Analytical Chemistry*, 88(21), 10395–10403.
- Torgrip, R. J. O., Åberg, M., Karlberg, B., & Jacobsson, S. P. (2010). Peak alignment using reduced set mapping. *Journal of Chemometrics*, 17(11), 573–582.
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: Computational MS/MS fragmentation and structure elucidation using MS-FINDER Software. *Analytical Chemistry*, 88(16), 7946–7958.
- van Stee, L. L. P., & Brinkman, U. A. T. (2016). Peak detection methods for GC×GC: An overview. *TrAC Trends in Analytical Chemistry*, 83, 1–13.
- Verdegem, D., Lambrechts, D., Carmeliet, P., & Ghesquière, B. (2016). Improved metabolite identification with MIDAS and MAGMA through MS/MS spectral dataset-driven parameter optimization. *Metabolomics*, 12(6), 1–16.
- Vergeynst, L., Van Langenhove, H., Joos, P., & Demeestere, K. (2013). Accurate mass determination, quantification and determination of detection limits in liquid chromatography-high-resolution time-of-flight mass spectrometry: Challenges and practical solutions. *Analytica Chimica Acta*, 789, 74–82.
- Veselinović, A. M., Velimirović, D., Kaličanin, B., Toropova, A., Toropova, A., & Veselinović, J. (2017). Prediction of gas chromatographic retention indices based on Monte Carlo method. *Talanta*, 168, 257–262.
- Vest Nielsen, N. P., Carstensen, J. M., & Smedsgaard, J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A*, 805(1–2), 17–35.

- Vignaduzzo, S. E., Maggio, R. M., & Olivieri, A. C. (2020). Why should the pharmaceutical industry claim for the implementation of second-order chemometric models—A critical review. *Journal of Pharmaceutical and Biomedical Analysis*, 179 112965.
- Walczak, B., & Wu, W. (2005). Fuzzy warping of chromatograms. *Chemometrics & Intelligent Laboratory Systems*, 77(1–2), 173–180.
- Wang, B., Fang, A., Shi, X., Ho Kim, S., & Zhang, X. (2011). DISCO2: A comprehensive peak alignment algorithm for two-dimensional gas chromatography time-of-flight mass spectrometry. *Bio-Inspired Computing and Applications*, 6840, 486–491.
- Wang, B. (2013). A two-stage peak alignment algorithm for two-dimensional gas chromatography time-of-flight mass spectrometry-based metabolomics. *Computational and Structural Biotechnology Journal*, 7(8) e201304002.
- Wang, F., Xiong, Z. Y., Li, P., Yang, H., Gao, W., & Li, H. J. (2017). From chemical consistency to effective consistency in precise quality discrimination of *Sophora flower-bud* and *Sophora flower*: Discovering efficacy-associated markers by fingerprint-activity relationship modeling. *Journal of Pharmaceutical and Biomedical Analysis*, 132, 7–16.
- Wang, X., Shen, Y., Wang, S., Li, S., Zhang, W., Liu, X., et al. (2017). PharmMapper 2017 update: A web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Research*, 45(W1), W356–W360.
- Wang, Y., & Gu, M. (2010). The concept of spectral accuracy for MS. *Analytical Chemistry*, 82, 7055–7062.
- Wang, Y., Kora, G., Bowen, B. P., & Pan, C. (2014). MIDAS: A database-searching algorithm for metabolite identification in metabolomics. *Analytical Chemistry*, 86(19), 9496–9503.
- Wang, Y. L., Cui, T., Li, Y. Z., Liao, M. L., Zhang, H. B., Hou, W. B., et al. (2019). Prediction of quality markers of traditional Chinese medicines based on network pharmacology. *Chinese Herbal Medicines*, 11(4), 349–356.
- Wang, Z., Kim, U., Liu, J., Cheng, C., Wu, W., Guo, S., et al. (2019). Comprehensive TCM molecular networking based on MS/MS in silico spectra with integration of virtual screening and affinity MS screening for discovering functional ligands from natural herbs. *Analytical and Bioanalytical Chemistry*, 411(22), 5785–5797.
- Wei, W., Hou, J., Yao, C., Bi, Q., Wang, X., Li, Z., et al. (2019). A high-efficiency strategy integrating offline two-dimensional separation and data post-processing with dereplication: Characterization of bufadienolides in *Venenum Bufonis* as a case study. *Journal of Chromatography A*, 1603, 179–189.
- Weusten, J. J. A. M., Derks, E. P. P. A., & Mommers, J. H. M. (2012). Sjoerd van der Wal, alignment and clustering strategies for GC×GC-MS features using a cylindrical mapping. *Analytica Chimica Acta*, 726, 9–21.
- Wold, S., & Christie, O. H. J. (1984). Extraction of mass spectral information by a combination of autocorrelation and principal components models. *Analytica Chimica Acta*, 165, 51–59.
- Wu, H., Li, G., Liu, S., Liu, D., Chen, G., Hu, N., et al. (2015). Simultaneous determination of six triterpenic acids in some Chinese medicinal herbs using ultrasound-assisted dispersive liquid-liquid microextraction and high-performance liquid chromatography with fluorescence detection. *Journal of Pharmaceutical and Biomedical Analysis*, 107(1), 98–107.
- Wu, H. L., Shibukawa, M., & Oguma, K. (1998). An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons. *Journal of Chemometrics*, 12(1), 1–26.
- Wu, X. M., Zuo, Z. T., Zhang, Q. Z., & Wang, Y. Z. (2018). Classification of Paris species according to botanical and geographical origins based on spectroscopic, chromatographic, conventional chemometric analysis and data fusion strategy. *Microchemical Journal*, 143, 367–378.
- Xiang, W., Suo, T. C., Yu, H., Li, A. P., Zhang, S. Q., Wang, C. H., et al. (2018). A new strategy for choosing “Q-markers” via network pharmacology, application to the quality control of a Chinese medical preparation. *Journal of Food and Drug Analysis*, 26, 858–868.
- Xiong, Y., Hu, Y. P., Li, F., Chen, L. J., Dong, Q., Wang, J. B., et al. (2018). Promotion of quality standard of Chinese herbal medicine by the integrated and efficacy-oriented quality marker of effect-constituent index. *Phytomedicine*, 45, 26–35.
- Xu, L., Shang, Z., Bo, T., Sun, L., Guo, Q., Qiao, X., et al. (2019). Rapid quantitation and identification of the chemical constituents in Danhong Injection by liquid chromatography coupled with orbitrap mass spectrometry. *Journal of Chromatography A*, 1606 460378.
- Yan, J., Zhan, Y., & Zhu, W. F. (2009). Application of support vector machine in TCM diagnosis. *Congress on Intelligent Systems*, 338–342.
- Yang, T., Yan, P., He, M., Hong, L., Pei, R., Zhang, Z., et al. (2018). Application of subwindow factor analysis and mass spectral information for accurate alignment of non-targeted metabolic profiling. *Journal of Chromatography A*, 1563, 162–170.
- Yang, W. Z., Zhang, Y. B., Wu, W., Huang, L. Q., Guo, D., & Liu, C. X. (2017). Approaches to establish Q-markers for the quality standards of traditional Chinese medicines. *Acta Pharmaceutica Sinica B*, 7(4), 439–446.
- Zeng, Z. D., Liang, Y. Z., Wang, Y. L., Li, X. R., Liang, L. M., Xu, Q. S., et al. (2006). Alternative moving window factor analysis for comparison analysis between complex chromatographic data. *Journal of Chromatography A*, 1107(1–2), 273–285.
- Zhang, A., Sun, H., Yan, G., & Wang, X. (2017). Recent developments and emerging trends of mass spectrometry for herbal ingredients analysis. *TrAC Trends in Analytical Chemistry*, 94, 70–76.
- Zhang, J., Zheng, C. H., Xia, Y., Wang, B., & Chen, P. (2017). Optimization enhanced genetic algorithm-support vector regression for the prediction of compound retention indices in gas chromatography. *Neurocomputing*, 240, 183–190.
- Zhang, P. J., Li, Y. M., Zhang, Y., Huang, W., Li, Y. B., Zhang, Y. J., et al. (2018). Application and prospect of toxicity quality markers of Chinese materia medica based on metabolomics. *Chinese Herbal Medicines*, 10(2), 108–116.
- Zhang, Q., Huo, M., Zhang, Y., Qiao, Y., & Gao, X. (2018). A strategy to improve the identification reliability of the chemical constituents by high-resolution mass spectrometry-based isomer structure prediction combined with a quantitative structure retention relationship analysis: Phthalide compounds in Chuanxiong as a test case. *Journal of Chromatography A*, 1552, 17–28.
- Zhang, C., Zheng, X., Ni, H., Li, P., & Li, H. J. (2018). Discovery of quality control markers from traditional Chinese medicines by fingerprint-efficacy modeling: Current status and future perspectives. *Journal of Pharmaceutical and Biomedical Analysis*, 159, 296–304.
- Zhang, W., Chang, J., Lei, Z., Huhman, D., Sumner, L. W., & Zhao, P. X. (2014). MET-COFEA: A liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation. *Analytical Chemistry*, 86(13), 6245–6253.
- Zhang, Z. M., Chen, S., & Liang, Y. Z. (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135(5), 1138.
- Zhang, Z. M., Liang, Y. Z., Lu, H. M., Tan, B. B., Xu, X. N., & Ferro, M. (2012). Multiscale peak alignment for chromatographic datasets. *Journal of Chromatography A*, 1223, 93–106.
- Zhao, J., Wu, H. L., Niu, J. F., Yu, Y. J., Yu, L. L., Kang, C., et al. (2012). Chemometric resolution of coeluting peaks of eleven antihypertensives from multiple classes in high performance liquid chromatography: A comprehensive research in human serum, health product and Chinese patent medicine samples. *Journal of Chromatography B*, 902, 96–107.
- Zheng, J., Wu, M., Wang, H., Li, S., Wang, X., Li, Y., et al. (2018). Network pharmacology to unveil the biological basis of health-strengthening herbal medicine in cancer treatment. *Cancers (Basel)*, 10(11), 461.
- Zheng, Y. B., Zhang, Z. M., Liang, Y. Z., Zhan, D. J., Huang, J. H., Yun, Y. H., et al. (2013). Application of fast Fourier transform cross-correlation and mass spectrometry data for accurate alignment of chromatograms. *Journal of Chromatography A*, 1286(8), 175–182.
- Zhou, Y., Yan, P., He, M., Hong, L., & Cao, Q. (2019). Hyphenated chromatography detection and compound-target-disease investigation on herb-pair *Chuanxiong Rhizoma - Xiangfu Rhizoma*. *Journal of Ethnopharmacology*, 243 112125.
- Zhu, L. X., Xu, J., Wu, Y., Su, L. F., Ching Lam, K. Y., Qi, E. R., et al. (2019). Comparative quality of the forms of decoction pieces evaluated by multidimensional chemical analysis and chemometrics: *Poria cocos*, a pilot study. *Journal of Food and Drug Analysis*, 27(3), 766–777.