Open Horizon Scientific Review

Vol. 02 No.02 2024

Chemical Fingerprinting of Natural Products Using Advanced Analytical Techniques

Bilal Ahmad Chaudhry[1]

[1] Department of Chemistry, Quaid-i-Azam University, Islamabad, Pakistan

ABSTRACT

Natural products such as medicinal plants, essential oils, foods, honey, resins and marine extracts are complex mixtures that contain hundreds to thousands of chemically diverse constituents whose composition varies with species, geography, harvest time and processing. Chemical fingerprinting aims to capture this complexity in a characteristic and reproducible profile that can be used for identification, quality control, authentication and discovery of bioactive compounds. This article reviews the principles and workflow of chemical fingerprinting and the major analytical platforms used to generate fingerprints, including high-performance liquid chromatography, gas chromatography, mass spectrometry and its hyphenated forms, nuclear magnetic resonance spectroscopy and vibrational and optical spectroscopies. Sample preparation, chemometric data analysis, method validation and standardization are discussed as essential components of a reliable fingerprinting strategy. Applications in herbal medicine, food authentication and drug discovery are highlighted, together with current limitations such as matrix effects, reproducibility and compound identification. Emerging trends including molecular networking, ambient ionization and artificial intelligence are outlined as directions for future development.

Key words : *chemical fingerprinting, natural products, mass spectrometry, chemometrics, metabolomics*

Introduction:

Natural products have been a source of medicines, flavors, fragrances, dyes and nutrition throughout human history, and they remain a major inspiration for modern drug discovery. A large proportion of approved small-molecule drugs are natural products, derivatives of natural products or compounds inspired by them. In many regions, including South Asia, traditional systems of medicine continue to rely on herbal preparations whose safety and efficacy depend on the identity and quality of the plant material. Unlike synthetic drugs, which usually contain a single defined active ingredient, natural products are multi-component systems in which the biological effect may arise from the combined action of many constituents.This chemical complexity creates serious challenges. The content of secondary metabolites in a plant depends on genetics, soil, climate, altitude, stage of growth, harvesting practice, drying, storage and extraction method, so two samples with the same botanical name may differ widely in composition. Adulteration, substitution with cheaper species and contamination are common problems in the herbal and food markets. Quantifying one or two marker compounds is often insufficient to guarantee quality, because it cannot reveal whether the overall chemical profile is consistent with that of an authentic reference.Chemical fingerprinting addresses this problem by recording a profile of the many

constituents of a sample, using an analytical technique with sufficient selectivity and reproducibility, and by using the pattern as a characteristic signature of the material. A fingerprint may consist of a chromatogram, a mass spectrum, an NMR spectrum or an infrared spectrum, and it can be compared with that of a reference sample by visual inspection, similarity measures or multivariate statistics. The concept is recognized by pharmacopoeias and regulatory bodies as a tool for quality control of herbal medicines.The present article reviews the use of advanced analytical techniques for the chemical fingerprinting of natural products. It describes the fundamental concepts, sample preparation, the principal analytical platforms, the chemometric methods used to interpret the data and the requirements for validation. It then discusses applications, challenges and future developments, and ends with a short summary.

Concept and Scope of Chemical Fingerprinting:

A chemical fingerprint is a characteristic pattern of signals that represents the chemical composition of a sample in a way that can be compared with other samples. The principle is analogous to a human fingerprint: the information does not require identification of every compound, but the pattern as a whole identifies the origin and quality of the material. Two approaches are commonly distinguished. In targeted analysis, a defined set of known compounds is quantified with high accuracy. In untargeted fingerprinting or metabolomics, as many signals as possible are recorded without prior selection and are analyzed statistically to find patterns and differences.A good fingerprint must meet several requirements. It should be representative of the major classes of constituents, selective enough to distinguish related species or origins, reproducible between runs, laboratories and instruments, and practical in terms of time and cost. The choice of technique therefore depends on the purpose of the analysis, whether it is rapid screening for authenticity, detailed characterization for research or quantification of bioactive markers for regulatory control. No single technique captures all constituents, which differ widely in polarity, volatility, molecular size and abundance, so orthogonal methods are often combined.The scope of fingerprinting covers many types of natural material. Medicinal plants and their extracts, essential oils, spices, teas, honey, edible oils, fruit juices, resins and propolis, microbial fermentation broths and marine organisms are all analyzed by these methods. Fingerprints can be used to identify botanical species, determine geographical origin, detect adulteration, monitor batch-to-batch consistency, follow changes during processing or storage and relate chemical composition to biological activity, which links analytical chemistry directly with pharmacology and food science.

Sample Collection and Preparation Strategies:

The reliability of a fingerprint begins with the sample. Representative sampling, correct botanical identification, documentation of the plant part, harvest conditions and storage, and the deposition of voucher specimens are necessary to ensure that analytical results can be interpreted. Drying methods, particle size and moisture content influence the extraction and must be standardized. Enzymatic activity and oxidation can alter the chemical profile quickly after harvest, so fresh material is sometimes frozen in liquid nitrogen or lyophilized to preserve

labile compounds.Extraction should recover as broad a range of constituents as possible in a reproducible way. Maceration, Soxhlet extraction, ultrasound-assisted extraction, microwave-assisted extraction, pressurized liquid extraction and supercritical carbon dioxide extraction are widely used, and the solvent is selected according to the polarity of the target compounds. Methanol and ethanol-water mixtures extract a wide range of polar and semi-polar metabolites, whereas hexane and dichloromethane extract lipophilic substances. For volatile constituents, techniques such as hydrodistillation, headspace sampling and solid-phase microextraction are preferred because they avoid solvent interference.Clean-up and fractionation reduce matrix complexity and protect instruments. Solid-phase extraction removes interfering pigments, lipids and sugars, and liquid-liquid partitioning separates compounds by polarity. For untargeted work, however, extensive clean-up must be avoided because it can remove constituents of interest and introduce variability. The use of internal standards, pooled quality control samples and randomized injection order is recommended so that variation caused by sample handling and instrument drift can be distinguished from true biological differences.

Chromatographic Fingerprinting Techniques:

High-performance liquid chromatography with ultraviolet or diode array detection is the most widely used technique for fingerprinting of non-volatile natural products such as flavonoids, phenolic acids, alkaloids, saponins and terpenoids. Reversed-phase columns with gradient elution of water and acetonitrile or methanol containing a small amount of acid provide good separation of compounds of intermediate polarity. The resulting chromatogram, with the retention times and relative peak areas of the common peaks, constitutes the fingerprint, and the diode array detector adds spectral information that supports peak identity and purity.Ultra-high-performance liquid chromatography, using columns packed with particles smaller than two micrometers and operating at high pressure, improves resolution, speed and sensitivity, and it reduces solvent consumption. Hydrophilic interaction chromatography complements reversed-phase separation by retaining very polar compounds such as sugars, amino acids and glycosides, and two-dimensional liquid chromatography increases peak capacity for highly complex extracts. Detectors such as evaporative light scattering and charged aerosol detectors allow the detection of compounds lacking a chromophore, including saponins and many lipids.Gas chromatography is the method of choice for volatile and semi-volatile compounds such as those in essential oils, aromas and fatty acid methyl esters. Capillary columns with non-polar or polar stationary phases separate hundreds of components, and retention indices calculated relative to a homologous series of alkanes allow identification independent of instrumental conditions. Comprehensive two-dimensional gas chromatography couples columns of different selectivity to resolve co-eluting compounds and produces structured chromatograms in which chemical classes occupy characteristic regions. Thin-layer chromatography and its high-performance version remain valuable for low-cost screening, particularly in laboratories with limited resources, since many samples can be analyzed in parallel and the plate image provides a visual fingerprint.

Mass Spectrometry and Hyphenated Platforms:

Mass spectrometry provides high sensitivity and structural information, and its coupling to chromatography has become the central platform for fingerprinting. In liquid chromatography-mass spectrometry, electrospray ionization generates protonated or deprotonated molecules from polar and semi-polar compounds, and atmospheric pressure chemical ionization extends the range to less polar compounds. Gas chromatography-mass spectrometry with electron ionization yields reproducible fragmentation patterns that can be compared with large spectral libraries for identification of volatile compounds.High-resolution analyzers, including time-of-flight, Orbitrap and Fourier transform ion cyclotron resonance instruments, determine accurate masses with errors of a few parts per million and thus permit the assignment of molecular formulas. Tandem mass spectrometry produces fragment ions that reveal substructures and allow the distinction of isomers when combined with retention time and reference data. Ion mobility adds a separation based on molecular shape and size, which helps to resolve isomeric compounds that cannot be separated by chromatography alone. The graph below illustrates the approximate number of constituents that typical analytical platforms can profile in a single plant extract run, showing the much larger coverage achieved with high-resolution liquid chromatography-mass spectrometry.The large data sets generated by untargeted mass spectrometry are organized with the help of molecular networking, in which compounds with similar fragmentation spectra are connected into clusters that correspond to structural families. Public platforms allow researchers to compare spectra with community libraries and to annotate known compounds, and they have accelerated the discovery of new natural products and the dereplication of known ones. Ambient ionization techniques such as direct analysis in real time and desorption electrospray ionization allow rapid analysis of samples with minimal preparation, which is useful for screening of herbal products and foods.

Nuclear Magnetic Resonance Spectroscopy:

Nuclear magnetic resonance spectroscopy is a highly reproducible and inherently quantitative technique that detects all proton-containing compounds in a sample above a certain concentration without the need for separation or derivatization. The proton spectrum of an extract acts as a fingerprint in which signals from sugars, amino acids, organic acids, phenolics, terpenoids and lipids can be recognized in different spectral regions. Because signal intensity is directly proportional to the number of nuclei, relative or absolute quantification is possible with an internal standard, which has made quantitative NMR a recognized method for purity determination and the standardization of reference materials.Two-dimensional experiments such as correlation spectroscopy, heteronuclear single quantum coherence and heteronuclear multiple bond correlation resolve overlapping signals and provide connectivity information needed to identify unknown compounds directly in mixtures. NMR is particularly valuable for distinguishing structural isomers, for determining stereochemistry and for detecting compounds that ionize poorly in mass spectrometry. Its main drawback is lower sensitivity, which limits detection to compounds present at micromolar concentrations or above, and the high cost of instruments, although

cryogenic probes and higher magnetic fields have improved performance.Hyphenated forms of NMR, such as liquid chromatography coupled with solid-phase extraction and NMR, allow separated peaks to be accumulated and analyzed with high sensitivity, and the combination of NMR and mass spectrometry in an integrated workflow takes advantage of the complementary strengths of the two techniques. Studies of food and herbal materials show that NMR-based fingerprints can discriminate geographical origin, cultivar and processing conditions with high reliability, and they are increasingly used in regulatory laboratories for authenticity testing of juices, honey, edible oils and wine.

Vibrational and Optical Spectroscopic Methods:

Infrared, near-infrared and Raman spectroscopy provide rapid and non-destructive fingerprints based on molecular vibrations. The mid-infrared region contains strong absorption bands from functional groups such as hydroxyl, carbonyl, amide and aromatic groups, and the region below about 1500 reciprocal centimeters, called the fingerprint region, is highly characteristic of the complete composition of a sample. Attenuated total reflectance accessories allow solid and liquid samples to be measured without preparation, and a spectrum can be acquired in less than a minute.Near-infrared spectroscopy measures overtones and combination bands and penetrates deeper into samples, which makes it suitable for the analysis of intact powders, grains and tablets and for on-line process monitoring. Raman spectroscopy, which is complementary to infrared, is sensitive to non-polar and symmetric bonds, is little affected by water and can be applied through glass or plastic packaging. Surface-enhanced Raman scattering increases sensitivity by several orders of magnitude and has been used for the detection of pesticides and adulterants in foods and herbal products. Portable and handheld instruments allow fingerprinting in the field, at customs points and at the point of sale.Ultraviolet-visible spectroscopy and fluorescence spectroscopy provide simple and inexpensive fingerprints for compounds with chromophores or fluorophores, such as anthocyanins, carotenoids, chlorophylls and polyphenols. Excitation-emission matrices of fluorescence can be analyzed by multiway methods to distinguish oils and beverages. These spectroscopic methods do not identify individual compounds, but they offer speed and low cost, and when combined with chemometric models they are highly effective for screening and classification.

Chemometric and Multivariate Data Analysis:

The data produced by fingerprinting techniques consist of hundreds or thousands of variables for each sample, and their interpretation requires chemometric methods. Data preprocessing is the first stage and includes baseline correction, alignment of retention times, peak detection, normalization and scaling. For spectroscopic data, standard normal variate transformation, multiplicative scatter correction and derivatives reduce physical effects such as light scattering. The choice of preprocessing strongly influences the results and should be documented so that the analysis can be reproduced.Unsupervised methods explore the data without prior knowledge of sample classes. Principal component analysis reduces the dimensions of the data and reveals clusters, trends and outliers in a score plot, and hierarchical cluster analysis groups samples

according to similarity. Supervised methods build models that relate the data to known classes or properties. Partial least squares discriminant analysis and orthogonal projections to latent structures identify variables responsible for the separation between groups, while linear discriminant analysis, support vector machines and random forests classify unknown samples, and partial least squares regression predicts quantitative properties such as content of a marker compound.Models must be validated carefully to avoid overfitting, which is a particular risk when the number of variables greatly exceeds the number of samples. Cross-validation, permutation testing and, most importantly, validation with an independent set of samples are required. Similarity measures such as the correlation coefficient and the cosine of the angle between fingerprints, as recommended in pharmacopoeial guidance, provide a simple evaluation of consistency for quality control. Deep learning models are increasingly applied to spectra and chromatograms and can learn features automatically, but they need large and well-curated data sets and their interpretation is more difficult.

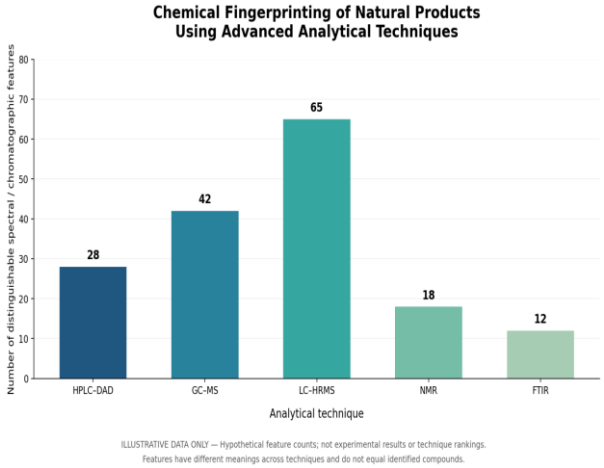
Method Validation, Standardization and Authentication:

For a fingerprinting method to be accepted for quality control or regulatory use, it must be validated. Specificity, precision, repeatability, intermediate precision, stability of the sample solution and robustness to small changes in conditions are normally assessed, and for quantitative methods linearity, accuracy, limits of detection and quantification are also determined. In fingerprint analysis, precision is often evaluated through the relative standard deviation of retention times and relative peak areas of common peaks, and through the similarity of replicate fingerprints.Standardization of procedures between laboratories is a major challenge. Differences in instruments, columns, mobile phases and software can change retention times and intensities, so the use of reference materials, system suitability tests and pooled quality control samples is essential. Minimum reporting standards proposed by the metabolomics community specify the information that must be provided about samples, preparation, data acquisition and processing, and levels of confidence in compound identification, thereby allowing results from different laboratories to be compared and reproduced.Authentication of natural products is one of the most important uses of validated fingerprints. A reference database of authentic samples covering natural variability must be built, and a decision rule for accepting or rejecting an unknown sample must be defined with known error rates. Chemical fingerprints are increasingly combined with DNA barcoding, which identifies the species but not the quality or content of active compounds, to give a complete picture of identity and quality. Such integrated approaches are recommended for high-value and frequently adulterated products such as saffron, honey, ginseng and essential oils.

Applications, Challenges and Future Directions:

In herbal medicine, chemical fingerprints are used for identification of species, assessment of batch consistency and detection of adulterants, and they form part of quality standards in several pharmacopoeias. In food science, they are applied to verify the origin of olive oil, honey, tea, wine and spices, to detect dilution and substitution and to monitor freshness. In drug discovery, fingerprinting combined with bioassay-guided fractionation and dereplication

quickly distinguishes known compounds from novel ones, saving the time and cost of re-isolation. In ecology and agriculture, metabolite profiles reveal how plants respond to stress, climate and cultivation practice, which is of interest for crop improvement.Important challenges remain. Compound identification is still the bottleneck of untargeted analysis, as only a minor fraction of detected features can be assigned with confidence because libraries of authentic standards are incomplete. Matrix effects, ion suppression and instrument drift limit reproducibility, and the chemical diversity and wide concentration range of natural products exceed the dynamic range of any single method. The absence of universally accepted criteria for how similar a sample must be to a reference, and of harmonized databases, also hinders the regulatory use of fingerprints.Future developments will likely combine multiple platforms in integrated workflows, with data fusion methods that merge information from mass spectrometry, NMR and spectroscopy. Advances in ion mobility, miniaturized and portable instruments, ambient ionization and in silico prediction of spectra and retention times will improve identification and extend analysis to the field. Artificial intelligence and shared open databases are expected to speed up annotation and classification. For countries such as Pakistan, which have rich medicinal plant resources, establishing national reference libraries of fingerprints of local species would support quality assurance, protect consumers and promote the export of authentic herbal products.



Summary:

Chemical fingerprinting converts the complex composition of natural products into characteristic profiles that can be used for identification, quality control, authentication and discovery. Chromatographic methods, mass spectrometry, NMR and vibrational spectroscopies each contribute different strengths in terms of separation, sensitivity, structural information, quantification and speed, and their combination provides the most complete view. Careful sampling, reproducible preparation, chemometric analysis and rigorous validation are as important as the instrument itself. The main current limitations are compound identification, standardization between laboratories and the lack of harmonized reference databases. Progress in high-resolution instrumentation, molecular networking, portable devices and artificial intelligence will further strengthen the role of fingerprinting in the safe and sustainable use of natural products.

References:

- Wolfender, J.-L., Marti, G., Thomas, A., and Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. Journal of Chromatography A, 1382, 136-164.
- Kim, H. K., Choi, Y. H., and Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. Nature Protocols, 5, 536-549.
- Fiehn, O. (2002). Metabolomics: the link between genotypes and phenotypes. Plant Molecular Biology, 48, 155-171.
- Verpoorte, R., Choi, Y. H., and Kim, H. K. (2007). NMR-based metabolomics at work in phytochemistry. Phytochemistry Reviews, 6, 3-14.
- Xie, P., Chen, S., Liang, Y., Wang, X., Tian, R., and Upton, R. (2006). Chromatographic fingerprint analysis, a rational approach for quality assessment of traditional Chinese herbal medicine. Journal of Chromatography A, 1112, 171-180.
- Rinnan, A., van den Berg, F., and Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry, 28, 1201-1222.
- Trygg, J., Holmes, E., and Lundstedt, T. (2007). Chemometrics in metabonomics. Journal of Proteome Research, 6, 469-479.
- Wang, M., Carver, J. J., Phelan, V. V., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nature Biotechnology, 34, 828-837.
- Sumner, L. W., Amberg, A., Barrett, D., et al. (2007). Proposed minimum reporting standards for chemical analysis. Metabolomics, 3, 211-221.
- Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., and Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. Nature Reviews Drug Discovery, 20, 200-216.
- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. Journal of Natural Products, 83, 770-803.
- Ellis, D. I., and Goodacre, R. (2006). Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst, 131, 875-885.