We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

6,600 Open access books available
177,000 International authors and editors
195M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

Liquid chromatography approaches account for a wide range of analytical applications that really have impact on our daily life, especially for pharmaceutical, biological and food analysis. It is the undeniable fact, even for those not initiated in the world of science [1].

Technically speaking, it is used to separate out a mixture into its individual parts based on the interactions of the sample and the mobile and stationary phases. There are several different types of liquid chromatography, in which the mobile phase is a liquid and the separation can be performed either in a column or on a plane. It is noteworthy that HPLC is the most popular chromatographic technique presently employed. This acronym stands for high-performance (or high-pressure) liquid chromatography, featuring the utilization of pumps to pass a relatively high pressurized liquid solvent through a column packed with very small particles.

Although HPLC is a versatile separation tool, the coupling of liquid chromatographic analysis with mathematical methods has been increasingly proved to be an economical alternative to resolve any problematic situation without using sophisticated instruments in modern laboratories. In this introductory chapter, the state-of-the-art developments and challenges of the application of mathematical methods in liquid chromatography are briefly described in a context to fit nicely with the scope of this book, specifically aimed at the exploitation of chemometrics and data analysis in chromatographic science.
2. Mathematical methods in liquid chromatography

2.1. Method development and validation

It was recognized that method development in HPLC analysis is often a troublesome and bewildering task for the newcomer approaching the problem in a very much “trial and error” manner [2]. This is because chromatographic separation of the analytes is a function of the controllable variables (e.g., mobile phase velocity and composition, column physico-chemical properties, separation temperature, detection wavelength, etc.) and the uncontrolled variables (noise, drift, column performance). Although the separation could be reasonably achieved with little consideration of the fundamental aspects of these variables, it would be optimized with a full consideration of their relevance.

In analytical chemistry, the optimization of an analytical response (called as the dependent variable) is to find the optimal settings or conditions for a number of experimental factors (called the independent variables) by using mathematical terms or rules. As early as 1979, Massart presented this issue at the 27th International Congress of Pure and Applied Chemistry, specifically addressing the methods unknown to the majority of analytical chemists such as information theory, autocovariance curves as applied in continuous analysis, multi-criteria analysis, feature reduction in pattern recognition and operations research [3]. At that time, method optimization was primarily performed via Dantzig’s simplex algorithm (or simplex method) [4]. For HPLC separations, however, there are significant disadvantages associated with simplex optimization because it is generally unable to assess the quality of a located optimum (e.g., the global optimum could not be obtained due to a change in peak elution orders in successive separations) and it requires a relatively large number of experiments to locate optimum separation conditions [5].

In a tutorial, Hayashi and Matsuda described the total optimization of a chromatographic process using analysis criteria such as (i) precision described by the Shannon mutual information (\( \phi \)) and (ii) efficiency denoted by the transmission speed (\( \nu \)) of the information (\( \phi \)) [6]. For demonstration, the authors showed the optimum condition of the HPLC analysis of an antipyretic mixture, corresponding to the maximal amount of (\( \phi \)) and/or (\( \nu \)) among all the operating conditions examined, e.g., mobile phase composition and velocity, organic modifier fraction, column length, detection wavelength and amount of internal standard.

To optimize mobile phases in isocratic RP-HPLC analysis, Cönenegrecht et al. guided the methodology to locate the global optima in terms of separation and analysis time of isoeluotropic and non-isoeluotropic ternary and quaternary eluent mixtures, using regression techniques combined with multi-criteria decision-making [7]. It was suggested that from the quadratic models, different separation criteria such as the resolution or selectivity of the worst separated pair of peaks could be calculated as well as the capacity factor of the last peak could be predicted as a measure of analysis time.

Statistical modeling of the chromatographic process shows that encountering two or more partially overlapped peaks is highly probable in a chromatogram (e.g., see Figure 1). In the 1990s, online multiwavelength absorptiometric detection was often used to generate the additional
data required to facilitate peak-purity assessment in LC [9]. Using photodiode array technology, there was still no single algorithm capable of fulfilling all the requirements of peak-purity assessment methods, especially for structurally similar impurities co-eluted with a parent compound. In this case, mass spectrometry (as a reliable and robust LC detector) should be used with multivariate statistical analysis to overcome the problem of spectral discrimination with a higher degree of confidence.

Computerized-assisted optimization (optimization in silico) is an attractive tool (i.e., robust and reliable) for LC method development (e.g., see Figure 2). It could be applied for the prediction of analyte retention and optimization of separations in ion chromatography, based on retention time and peak width modeling data. Provided that numerous isocratic and/or gradient steps in ion chromatographic analysis of very complex mixtures involve a large input in time, in silico approaches can offer rapid algorithms for prediction of retention time [11].

In 1992, Carlson [12] drew the attention of chemists on the use of statistical experimental design with a tutorial emphasizing on the steps to be taken before conducting any multivariate design of a screening experiment.

Nowadays, the multivariate optimization techniques have been increasingly employed in the field of analytical chemistry, in particular the surface response methodologies (e.g., see Figure 3). For preliminary assessment of experimental factors, the two-level full factorial design was often used, whereas the central composite design was most utilized for determination of critical conditions using quadratic models [14]. With the greater availability of statistical software and overall capacity of modern instrumentation, the popularity of experimental design has not been surprisingly increasing. In chromatographic analysis, it is devoted to showing lack of significant effects in robustness studies for method validation (e.g., Plackett-Burman designs) and identifying significant factors in response to optimization for method
Figure 2. Overlaid DryLab1 and actual two-dimensional separation of the model methamphetamine seizure sample [10].

Figure 3. 3D surface response plot of desirability for optimization of HPLC factors: peak area, tailing factor (5%) and number of theoretical plate [13].
development (e.g., fractional factorial designs and their extensions such as central composite designs, Box-Behnken and Doehlert designs). It is noticed that D-optimal designs are now becoming more popular and particularly useful when the factor space is not uniformly accessible (when combinations of solvent composition and solute concentration are not possible) and/or there is a constraint on the total number of experiments that can be done [15].

Doehlert uniform shell designs have been implemented in LC analysis for the development and improvement of sample preparation by exploring up to four factors and the optimization of specific instrumental parameters by exploring up to three factors [16].

Recently, chemometric approaches, i.e., the Quality by Design paradigm for method development and the Six Sigma practice as a quality indicator in chromatography, have been acknowledged as precious complements to HPLC practices [17]. Although Design of Experiment (DOE) application in chromatography gives excellent separation desirability, global satisfaction is still unattainable. The treatment of Quality by Design paradigm, initiated by USFDA and ICH, can locate the global optimal condition of a robust HPLC method with fewer method transfer or failures and issues by establishing a comprehensive design space as a function of DOE philosophy. On the other hand, Six Sigma is a set of statistical techniques and tools used to indicate a process yield in terms of the percentage of defect-free outputs, i.e., in a Six Sigma process, 99.99966% of all outputs are statistically expected to be free of defects. The characteristics of a chromatographic process may be thus amenable to Six Sigma overall process control, e.g., reduction in organic solvent used.

In an effort to improve LC analysis, it is often worth trying to replace a column with either an identical or a totally different one. This could be effectively done, based on the classification of chromatographic stationary phases by a number of approaches, i.e., radar plots, hierarchical cluster analysis, principal component analysis, calculation of a distance factor from a reference column based on Pythagorean Theorem, and two-dimensional diagrams plotting one property versus another [18]. Moreover, chromatographic resolution and analysis time could be further enhanced by modulation of the stationary phase through a serial coupling of columns. Full benefit of such coupling is only achieved through the application of interpretive (i.e., based on model) optimization of columns (nature and length) and solvent (isocratic or gradient mobile phase) [19].

Basically, a chemometric approach can help systematize the knowledge of factors affecting retention in RP-HPLC. By this means, quantitative structure (reversed-phase)–retention relationships (QSRRs) could rationalize RP-HPLC retention mechanism by employing molecular descriptors expected to model fundamental intermolecular interactions and solvatochromic parameters of similar potency for retention prediction [20]. In literature, different modeling methodologies were reported for QSRR approaches for RP-HPLC, e.g., principal component analysis (e.g., see Figure 4) and decision trees; these methodologies include artificial neural networks, partial least squares, uninformative variable elimination partial least squares, stochastic gradient boosting for tree-based models, random forests, genetic algorithms, multivariate adaptive regression splines, and two-step multivariate adaptive regression splines [22]. In thin layer chromatography (TLC) analysis, QSRR studies were employed to illustrate the relationship between retention and lipophilicity of solutes [23]. Nonetheless, not all physicochemical descriptors correlate strongly with retention data, and there is no need to display retention data in the form of an equation, given a small number of compounds involved [24].
Hydrophilic interaction liquid chromatography (HILIC) is an alternative HPLC mode that utilizes hydro-organic mobile phases with a high organic content and a hydrophilic stationary phase for separating predominantly polar compounds and charged substances in complex matrices. With an enormous growth in the application of HILIC, computer-aided modeling (most importantly the QSRR modeling strategy) proved to be useful in understanding the retention mechanism, classification of stationary phases, prediction of retention times, optimization of chromatographic conditions, and interaction effects of chromatographic factors [25].

In addition to QSRR studies, molecular simulation methods for RP-HPLC modeling (such as molecular dynamics and Monte Carlo) are able to elucidate not only the retention mechanism for different analytes but also (i) the structure and dynamics of the bonded phase and its interface with the mobile phase and (ii) the interactions of analytes with the bonded phase. While the former can provide information on chain dynamics and transport properties, the latter is uniquely suited for the investigation of phase and sorption equilibria underlying RPLC retention [26].

2.2. Data analysis

Despite of being widely applied to the analysis of real-world samples, one-dimensional chromatography cannot always guarantee sufficient resolving power for separation of target compounds. As a result, multidimensional chromatography is proposed for delivering heightened separation performance for complex and difficult substances [27]. A conventional heart-cutting
multidimensional chromatographic technique uses multiple columns, each with a different stationary phase, positioned at right angles to each other to selectively choose elements unthor-oughly processed from the first column and transfer them to the second column for better sepa-
ration. In contrast, in a comprehensive multidimensional chromatography, the entire sample is subjected to separation in both dimensions with the following requirements: (i) any two com-
ponents separated in the first dimension must remain separated in the second dimension and (ii) elution profiles from both dimensions are preserved [28]. In principle, the implementation of
two-dimensional chromatography could be realized by combining the chromatographic sepa-
rations in space ($LC^t \times LC^x$) and/or in time ($LC^t$). Although column-based combinations ($LC^t \times LC^x$) are the most actively pursued, the combination $LC^t \times LC^x$ can provide remarkable time-saving
because it allows simultaneous second-dimension separations of all the fractions resolved in
the first-dimension separations [29]. It was shown that fast, comprehensive two-dimensional
liquid chromatography could offer the huge advantage over one-dimensional one with regard
to a tremendous potential increase in resolving power. This is because the peak capacity of the
comprehensive system approaches the product of the peak capacities of the first- and second-
dimension separations when combining highly orthogonal separation mechanisms [30]. This
chromatographic technique (particularly coupled with mass spectrometry detection) has been
increasingly applied for the analysis of food samples of both relatively low and highly chal-
 lenging complexities [31–33], complex Chinese herbal medicines (e.g., see Figure 5) [35, 36],
synthetic polymers and biopolymers [37]. Data processing of comprehensive two-dimensional
chromatography is an area of great expansion with reference to data acquisition and handling,
peak detection (i.e., two-concept algorithm, inverted watershed algorithm and multiway che-
mometric methodologies (such as parallel factor analysis model, target finder algorithms and
multivariate curve resolution with alternating least squares)), deconvolution of overlapping
peaks and data analysis software [38].

Modern instruments can generate a variety of second- and higher-order measurement data
that makes multiway analysis a subject of high-level interest for the analytical community [39].
Separation quality of hyphenated chromatography—(multichannel) spectroscopy instruments—
could be reliably and essentially evaluated by using chemometric methods for two-way data
treatment to determine the number of chemical components, elution sequence and peak purity
[40]. Identification and quantification of banned substances or substances with a specified maxi-
mum limit could be chromatographically performed by constructing calibration of a systematic
analytical procedure with chromatographic n-way data, with commonly used techniques such
as n-way partial least squares, multivariate curve resolution and parallel factor analysis [41].
Numerous works on quantitative analysis were reported for second-order data (matrices from
unidimensional chromatography with multivariate detection or from two-dimensional chroma-
tography) or third-order data (three-dimensional data arrays from two-dimensional chromatog-
raphy with multivariate detection), with special attention paid to processing algorithms to cope
with the ubiquitous phenomenon of inter-run variation in time shift [42].

By definition, chromatographic fingerprinting is a quality control model that looks at the “com-
plete information” or comprehensiveness of the chromatogram and displays integrated quality
information. In recent years, its exploitation has really implied a methodological change on
the major paradigm of chromatography. By referring to the common peaks of chromatograms, chro-
matographic fingerprinting could be applied for identification and authentication of foodstuffs
Figure 5. Contour plots for *P. lobata* by using conventional comprehensive 2DLC (A) and heart-cutting + comprehensive 2DLC (B). Dark areas in plot (B) indicate heart-cutting sections. After removing the major compounds by heart-cutting, the minor constituents of the Chinese herbal medicine *Puerariae radix* (Ge-Gen) could be clearly observed on the 2D plot [34].

Figure 6. MRM-MS-based detection of a putative biomarker peptide belonging to proliferating cell nuclear antigen (ATPLSLTTLSMSADVPLVEVK) that is not identifiable in the MCF-7 breast cancer cells’ extract by data-dependent analysis [47].
Specific chemometric tools are used in such studies, aiming at (i) verification of the similarity between chromatographic signals, (ii) resolution of overlapped chromatographic signals, and (iii) classification of samples regarding diverse criteria.

Modern combination of liquid chromatography with mass spectrometry, in particular matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS), has quickly established LC–MS as a popular tool used by the proteomics research community to search for new potential markers associated with pathological conditions over the last decade [46]. Many different chemometric approaches currently used for LC–MS spectrum processing in biomarker discovery are mainly for reducing the background noise and properly classifying the samples to the studied groups based on statistically significant features selected by chemometric algorithms (e.g., see Figure 6).

3. Conclusion

Based on the above overview, it is reasonable to state that mathematical methods provide effective means for all critical stages in LC analysis with a considerable degree of flexibility and adaptability to different types of practical problems. Nevertheless, despite the fact that they have been continually improved since the 1990s, most chromatographers continue to develop LC procedures manually on the basis of their own experience and chromatographic training. Therefore, it requires new more user-friendly chemometric tools that are better adapted to situation-driven development of chromatographic separation so that computer-assisted method development will become a routine procedure in chromatography laboratories.

Author details

Vu Dang Hoang
Address all correspondence to: hoangvd@hup.edu.vn
Department of Analytical Chemistry and Toxicology, Hanoi University of Pharmacy, Hanoi, Vietnam

References


