Liquid Chromatography Coupled to Mass Spectrometry

Implementation of Chemometric Optimization and Selected Applications

MY MOBERG
Dissertation presented at Uppsala University to be publicly examined in Room B:21, BMC, Uppsala, Thursday, September 28, 2006 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Liquid chromatography (LC) coupled to mass spectrometry (MS) offers highly selective and sensitive analysis of a wide variety of compounds. However, the use of hyphenated experimental set-ups implies that many parameters may have an effect on the studied response. Therefore, in order to determine optimized experimental conditions it is of vital importance to incorporate systematic procedures during method development. In this thesis, a generic stepwise optimization strategy is proposed that aims at high chromatographic quality, as well as high mass spectrometric response. The procedure comprises (i) screening experiments to identify the most important parameters, (ii) LC studies to ensure sufficient chromatographic separation, (iii) extended infusion experiments in order to maximize precursor signal(s), and in the case of tandem MS (iv) extended infusion experiments to determine optimal conditions for collision induced dissociation and when applicable also ion trap settings. Experimental design and response surface methodology is used throughout the procedure.

Further, the general applicability of LC-MS is demonstrated in this thesis. Specifically, a novel quantitative column-switched LC-MS method for ferrichrome, ferrichrysin and ferricrocin determination is presented. Using the method it was shown how the siderophore content varies with depth in podzolic soil profiles in the north and south of Sweden. The parallel approach using LC coupled to both inductively coupled plasma (ICP) mass spectrometry, and electrospray ionization (ESI) tandem MS is also evaluated as a tool to identify unknown siderophores in a sample. Additionally, different trypsin digestion schemes used for LC-ESI-MS peptide mapping were compared. By multivariate data analysis, it was clearly shown that the procedures tested induce differences that are detectable using LC-ESI-MS. Finally, the glutathione S-transferase catalyzed bioactivation of the prodrug azathioprine was verified using LC-MS.

Keywords: liquid chromatography, mass spectrometry, tandem mass spectrometry, electrospray ionization, inductively coupled plasma, chemometrics, optimization, multivariate data analysis, design of experiments, response surface model, siderophore, azathioprine, peptide

My Moberg, Department of Physical and Analytical Chemistry, Analytical Chemistry, Box 599, Uppsala University, SE-75124 Uppsala, Sweden

© My Moberg 2006

ISSN 1651-6214
ISBN 91-554-6612-5
urn:isbn:se:uu:diva-7071 (http://urn.kb.se/resolve?urn=urn:isbn:se:uu:diva-7071)
Papers included in this thesis

This thesis is based on the following papers, which are referred to in the text by their Roman numerals as indicated below:


VII. Daniel Bäckström, Per J. R. Sjöberg, Jonas Bergquist, Rolf Danielsson, **My Moberg**. Comparison between different trypsin digestion procedures used for LC-ESI-MS peptide mapping. Manuscript


Reprints were made with kind permissions from the publishers.

**Author’s contribution**

I performed all experiments and wrote the Papers I, II, IV and VI. Paper III was executed in close collaboration with Eva Nilsson. I was responsible for planning and writing the Paper VII. I was consulted during method development in Paper V, and I also critically reviewed the content of this paper. In Paper VIII I was responsible for the LC-MS analyses, and I also wrote these parts of the paper.
Contents

1 Introduction ................................................................................................................. 9

2 Mass spectrometry ..................................................................................................... 11
  2.1 Electrospray ionization (ESI) .............................................................................. 13
  2.2 Collision induced dissociation (CID) .................................................................. 16
  2.3 Speciation analysis with ICP-MS ........................................................................... 17

3 Hyphenation of LC and MS ..................................................................................... 19
  3.1 Method development in LC-ESI-MS and LC-ESI-MS/MS ................................. 21
    3.1.1 Optimizing criteria for LC and MS hyphenated methods .............................. 25
    3.1.2 Chemometric tools for optimization ............................................................ 27
      3.1.2.1 Simplex routines .................................................................................. 28
      3.1.2.2 Genetic algorithms (GAs) ..................................................................... 29
      3.1.2.3 Multiple linear regression (MLR) .......................................................... 30
      3.1.2.4 Artificial neural networks (ANNs) ......................................................... 30
    3.1.3 A proposed stepwise optimization strategy ..................................................... 32
      3.1.3.1 Identifying important parameters ........................................................... 33
      3.1.3.2 Ensuring liquid chromatographic quality ............................................... 33
      3.1.3.3 Obtaining high sensitivity for a precursor ion ......................................... 34
      3.1.3.4 Obtaining high sensitivity for a product ion ............................................ 35
      3.1.3.5 What is gained by the stepwise optimization strategy? ............................ 35
    3.1.4 Interpretation of LC-MS and LC-MS/MS data .................................................. 36
      3.1.4.1 Chemometric tools for fingerprinting/classification ................................. 36

4 Selected applications using LC coupled to MS ......................................................... 38
  4.1 Drug analysis ........................................................................................................ 38
  4.2 Siderophore analysis ............................................................................................ 40
  4.3 Protein and peptide analysis ................................................................................ 44

5 Concluding remarks and future outlooks ................................................................. 47

6 Acknowledgements .................................................................................................... 48

7 Swedish summary ...................................................................................................... 50
  7.1 Introduktion till LC-MS ...................................................................................... 50
  7.2 Utfordra delarbeten ............................................................................................. 51
  7.3 Utikter för framtiden och avslutande reflektioner .................................................. 52

8 References ................................................................................................................ 53
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>APCI</td>
<td>Atmospheric pressure chemical ionization</td>
</tr>
<tr>
<td>APPI</td>
<td>Atmospheric pressure photo ionization</td>
</tr>
<tr>
<td>Aza</td>
<td>Azathioprine</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>CE</td>
<td>Capillary electrophoresis</td>
</tr>
<tr>
<td>CEC</td>
<td>Capillary electrochromatography</td>
</tr>
<tr>
<td>CID</td>
<td>Collisional induced dissociation</td>
</tr>
<tr>
<td>DESI</td>
<td>Desorption electrospray ionization</td>
</tr>
<tr>
<td>DOC</td>
<td>Dissolved organic carbon</td>
</tr>
<tr>
<td>DoE</td>
<td>Design of experiment</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>EPI</td>
<td>Enhanced product ion scan</td>
</tr>
<tr>
<td>FTICR</td>
<td>Fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>GSH</td>
<td>Glutathione</td>
</tr>
<tr>
<td>ICP</td>
<td>Inductively coupled plasma</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>LMMOA</td>
<td>Low molecular mass organic acid</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix assisted laser desorption ionization</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple linear regression</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass-to-charge ratio</td>
</tr>
<tr>
<td>NPLC</td>
<td>Normal-phase liquid chromatography</td>
</tr>
<tr>
<td>PARAFAC</td>
<td>Parallel factor analysis</td>
</tr>
<tr>
<td>PC</td>
<td>Principal component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Principal component regression</td>
</tr>
<tr>
<td>PGC</td>
<td>Porous graphitic carbon</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial least squares</td>
</tr>
<tr>
<td>PTMs</td>
<td>Post-translational modifications</td>
</tr>
<tr>
<td>Q</td>
<td>Quadrupole</td>
</tr>
<tr>
<td>RPLC</td>
<td>Reversed-phase liquid chromatography</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>TOF</td>
<td>Time-of-flight</td>
</tr>
</tbody>
</table>
Conventions

Scalars are represented by lower-case characters, \( x \).
Vectors are represented by bold lower-case characters, \( \mathbf{x} \).
Matrices are represented by bold upper-case characters, \( \mathbf{X} \).
Running indices are indicated by sub-scripts on character, e.g. \( x_i \).
1 Introduction

Over the last century mass spectrometry (MS) has emerged as an indispensable analytical technique. Today, liquid chromatography (LC) coupled to MS is frequently used in routine qualitative and quantitative analysis in the laboratories. Historically, the issues of concern in MS research have mainly been the development and refinement of the hardware employed, whereas now a lot of emphasis is put on data evaluation and interpretation. The possibility to discover new biomarkers using MS is a rapidly growing field of research. Biomarkers are compounds that potentially can be used for early diagnostic or disease/treatment surveillance purposes. The measurement should then convey information of the biological condition being tested. Sceptics, however, stress the difficulties of validating the analytical procedures involved.

Even though MS has become a mature technique, there are still limitations to the instrumentation available today. Especially in quantitative analysis, many researchers report unwanted increase or decrease of the MS signal due to other constituents in the sample, which potentially will impair the analysis. Therefore, the quest for more robust technical solutions and analytical methods will continue.

The potential of MS, and also LC-MS, is recognized in the huge amount of data produced. However, in order to fully appreciate the data gathered, proper techniques for data treatment and evaluation have to be incorporated. The application of different mathematical tools is therefore a prerequisite for the realization of the high potential of MS. Also, in order to ascertain optimized conditions for analyses, it is essential to identify and properly adjust the parameters most likely to affect the obtained results. The field of chemometrics offers the means for successful planning and evaluation of the experiments.

In this thesis, methodologies for optimization of LC-MS integrated methods are presented. A number of questions are put forward and discussed, such as: What are the optimization criteria, how do we implement appropriate optimization strategies/procedures, and how do we interpret the data obtained? Writing the thesis, my intention has been to present and explore different parameters associated with an LC-MS hyphenated experimental set-up, utilizing mainly electrospray ionization (ESI). Possible restrictions when it comes to choosing the setting of a specific parameter are discussed. A stepwise optimization strategy is also proposed, that hopefully will aid the
reader to optimize the performance of such an experimental set-up. The outlined strategy is based on the experiences drawn from Papers I, IV and VI. In Paper I an optimization strategy for isocratic LC-ESI-MS was developed, while in Paper IV the parameters affecting the amount of fragmentation in tandem mass spectrometry (MS/MS) experiments were studied. In Paper VI a strategy for LC-ESI-MS/MS optimization was reported.

In addition, the thesis comprises a number of specific applications of LC-MS, including analyses of both small (drugs), medium (siderophores) and moderately large (peptides) compounds. The enzyme-catalyzed bioactivation of the prodrug azathioprine (Aza) was studied in Paper VIII. In Papers II, III and VI LC-MS based methods for siderophore (strong iron-chelating compounds) analyses were developed. Different approaches were exploited including LC-ESI-MS (Paper II) and LC-ESI-MS/MS (Paper VI). The unique features of the inductively coupled plasma (ICP) were also investigated as a complementary ionization technique (Paper III). The method developed in Paper II was further applied to natural soil solution samples in Paper V. In Paper VII different protein digestion protocols used prior to peptide mapping with LC-ESI-MS were compared.
2 Mass spectrometry

The mass spectrometer works, after careful calibration, as a highly accurate analytical scale for ions. Schematically, the MS system consists of an ion source, a mass analyser and a detector. In the ion source gas phase ions are formed, which are separated in the mass analyser according to their mass-to-charge ratios (m/z) and subsequently detected.

The ability to separate ions based on their mass and charge was first described by J. J. Thompson in the late 19th century and early 20th century\(^1\), and has since then risen to a prime analytical detection technique. The steady increasing popularity of MS is explained by the diversity of applications, e.g. in atomic physics, reaction physics, reaction kinetics and all forms of chemical analysis.

Different types of mass analysers exist on the market today, making use of separate principles for determination of the m/z of the ion investigated. In this thesis time-of-flight (TOF)\(^{III, VII}\) and quadrupole (Q)\(^{I, VI, VIII}\) mass analysers were used, and therefore deserve a closer description. In a TOF\(^2\) mass spectrometer, the time (t) is measured for an ion to travel a pre-defined length (d), i.e. the length of the flight tube. Small “packages” of ions, with essentially the same kinetic energy, are released into the flight tube and as low mass and/or highly charged ions travel faster than high mass and/or singly charged ions, the various ions reach the detector at different times. The time measures are subsequently used to calculate the corresponding m/z according to:

\[
\frac{m}{z} = t^2 \left( \frac{2Ve}{d^2} \right),
\]

where \(e\) is the charge of an electron and \(V\) is the applied acceleration voltage.

The mass resolving properties of the Q mass analyser is more instructively described as a variable mass band-pass filter, since only ions in a narrow mass range are allowed to pass the analyser at a time\(^3\). Physically the quadrupole consists of four rods. An alternating positive and negative potential is applied to the rods, making the ions accelerate either towards or away from the centre-axis of the rods. By putting the two pair of rods 180° out of phase discrimination towards unwanted ions can be obtained. Ions that are
too heavy (for the particular settings) will experience an averaged applied electric field, i.e. the DC (direct current) setting, and hence their trajectory paths will be directed away from the centre of the rods, while ions that are too light are accelerated into the rods before the polarity is changed. Commercially available instruments are single and serially coupled Q set-ups. In single Q instrumentation, ions entering the mass spectrometer can be analysed with respect to m/z, whereas serially coupled Qs offers the opportunity to generate and subsequently analyse fragments of these ions by the utilization of a collision cell. The ions fragmented in tandem MS operation are called precursor or parent ions and the generated fragment ions are called product or daughter ions. However, in order to facilitate further discussion in this thesis, the ions entering the mass spectrometer are exclusively referred to as precursor ions, regardless single or tandem MS operation, and the generated fragment ions are referred to as product ions. For a more detailed description of the mechanisms of fragmentation see Section 2.2.

Recently, a hybrid instrument called the Q TRAP™ was made commercially available⁴. In this mass spectrometer the ion path is configured as a triple quadrupole instrument, with the possibility to use the third Q as a linear ion trap. Linear ion traps have a relatively high ion storage capacity due to the absence of a quadrupole field along the z-axis, i.e. the ions are focused to a line rather than to a point. Ions admitted into the partly pressurized linear quadrupole ion trap (due to collision gas leakage from the collision cell) undergo a series of collisions effectively reducing their axial energy prior to encountering the end electrodes, thereby enhancing trapping efficiency and thus also sensitivity⁴. The Q TRAP™ instrument was used in Papers III, IV, VI and VIII (see Figure 1 for a graphical presentation).

Other mass analysers include e.g. magnetic and electric sector instruments and ion traps. Particularly accurate measures of m/z can be obtained using a Fourier Transform Ion Cyclotron Resonance (FTICR)⁵ ⁶ mass analyser, through the utilization of a very strong magnet, typically 3-9.4 Tesla⁶. For a more comprehensive review on MS instrumentation the reader is referred to literature⁷ ⁸.
A prerequisite for MS analysis is, however, that the compounds to be "weighed" are charged and present in the gas phase and hence a lot of emphasis has been paid to the development of different ionization techniques suitable for incorporation with MS detection. As an evidence for the importance of applicable ion sources, John B. Fenn and Koishi Tanaka were in 2002 rewarded with the Nobel price in chemistry for their pioneering work employing ESI and laser desorption ionization, respectively, for the analysis of large biomolecules.

Ionization of compounds can be performed at atmospheric pressure or in vacuum. There are essentially three different kinds of ionization techniques at atmospheric pressure, namely atmospheric pressure chemical ionization (APCI), atmospheric pressure photo ionization (APPI) and ESI (for details on ESI, see Section 2.1). In short, ionization using APCI and APPI takes place in the gas phase, whereas in ESI it is mainly considered to take place in the liquid phase. In APCI a corona discharge is incorporated to initiate ionization, while photons are responsible for the initiation using APPI. In ionization under vacuum, the most frequently employed techniques are matrix assisted laser desorption ionization (MALDI) and electron ionization. Even though the number of available ionization techniques has increased rapidly since the first introduction of the mass spectrometer, there are still new techniques and variants of already existing techniques presented today, one of the most promising being desorption electrospray ionization (DESI). In DESI, an electrospray emitter is used to create gas phase solvent ions and charged droplets, which are directed towards a solid, liquid or frozen sample at atmospheric pressure. The projectiles hitting the sample surface generate ions that are desorbed into the gas phase.

It is outside the scope of this thesis to make a detailed comparison between different ionization techniques. However, since optimization of parameters affecting the ESI process is reported (Papers I and VI), this particular technique is given a more comprehensive portrayal as outlined below. Also, the unique features of the ICP were exploited in Paper III and therefore the technique is discussed in Section 2.3.

2.1 Electrospray ionization (ESI)

As pointed out in the previous section, gas phase ions are analysed in the mass spectrometer and hence ionization of analytes and evaporation of solvents and additives are crucial steps in the MS analysis. The first description of electrospray was made by Zeleny in 1917, but the mechanisms of ESI as a sample introduction technique was first described by Dole et al. in the 1960s. John B. Fenn has earned the merit of being the first person to couple ESI to mass spectrometric detection in his work together with Yamashita.
Employing ESI\textsuperscript{21-23}, a continuous stream of liquid is fed through a capillary to the ion source, and by applying a voltage an electric field is obtained between the capillary tip and the counter-electrode, i.e. the mass spectrometer entrance. A graphical presentation of the electrospray ion source is given in Figure 2. The presence of the field will lead to a charge separation at the meniscus of the liquid at the capillary tip. An excess of ions with opposite charge compared to the counter-electrode will be formed at the surface. The meniscus will expand due to ion repulsion and a cone will protrude from the capillary. If the applied field is strong enough a fine jet emerges from the cone tip, and from the cone-jet small droplets are released with a net charge. Ideally, the charged droplets shrink as the solvent evaporates from the surface. According to the charge residue model suggested by Dole et al.\textsuperscript{18} droplet fission will occur when the repulsion between the charges on the surface becomes too high, yielding new smaller droplets. As these droplets shrink the charge repulsion, once again, will force the droplets to fission, and the procedure is continued until gas phase ions are formed. The ion evaporation model introduced by Iribarne and Thomson\textsuperscript{24}, on the other hand, implies that ions are released directly into the gas phase from the droplet surface, due to the charge repulsion. Probably both these mechanisms take place, and also co-exist, in the electrospray, the charge residue model being more dominate for large molecules while the ion evaporation model is probably a more accurate description of gas phase ion generation of smaller molecules\textsuperscript{25,27}.

![Figure 2. Schematic picture of the electrospray process in positive mode, with the positive net charge indicated. Picture by Andreas Dahlin.](image)
As ions are transported from the capillary tip to the mass spectrometer, a resulting current can be measured. In an equation derived by Pfeifer and Hendricks\textsuperscript{28} the estimated current is a function of the permittivity of solvent and vacuum, the surface tension, the conductivity of the solution, the applied electric field and the volumetric flow rate. The electric field is, in turn, a result of the applied voltage, the radius of the capillary emitter and the distance between the emitter and the counter plate. Other alternative equations have also been put forward in literature to explain the obtained current and mass spectrum intensities employing ESI, using theoretical and/or empirical approaches\textsuperscript{29-33}.

The amount and type of ions that enter the mass spectrometer can be explained by liquid-phase chemistry to a great extent, as the gas phase ions generated are a result of the charge separation obtained in the liquid. The formation of ions from acids and bases are governed by deprotonation or protonation reactions, while adduct formation is often obtained between neutrals and various ions (typically with protons or sodium ions). The charge separation will also be accompanied by electrochemical reactions\textsuperscript{34} that may also produce analyte ions. In addition to liquid phase based reactions, formation of ions can also take place in the gas phase\textsuperscript{21,35,36}.

Using ESI, the analyte has to incorporate charge, thus chemical properties such as pK\textsubscript{a} of the analyte will be very important for the determination of appropriate experimental conditions. Also, the solute has to be distributed to the droplet surface, and not situated in the interior of the droplet, in order to end up in the gas phase, hence the surface activity\textsuperscript{30,31,37} and the electrophoretic mobility\textsuperscript{30,31} of the solute are important. Sometimes derivatization of the analyte is performed in order to increase the sensitivity employing ESI, for instance by the incorporation of a permanently charged moiety\textsuperscript{38}. Other constituents and their relative concentrations in the solution will also affect the amount of analyte ions entering the mass spectrometer\textsuperscript{29,39,40}. Ion suppression of the signal of interest is a common phenomenon in ESI when non-volatile buffers and additives are used\textsuperscript{41,42}.

Evidently, many parameters often manipulated by the mass spectrometrist will affect the overall spray performance, and as a consequence proper considerations should be taken when ESI ionization is applied. The parameters to consider include the composition of the spayed solution, the applied ESI voltage, the distance between the emitter and the mass spectrometer entrance, and the flow rate of the sprayed solution.

Different configurations of ESI ion sources exist, sustaining a stable spray at different flow rates. So called pure electrospray is achieved when the solution is sprayed directly from the emitter without any aid, while the term ion spray\textsuperscript{43} is often used in correspondence to pneumatically assisted electrospray. Pure electrospray is best suited for low flow rates below 10 \(\mu\)L/min, whereas ionspray can use flow rates up to 200 \(\mu\)L/min\textsuperscript{44}. If still higher flow rates are to be sprayed some kind of heating is generally
incorporated in the ion source. The pressure of the assisting gas using ion spray, i.e. the nebulizer gas setting, and the heating temperature are thus possible additional parameters affecting the spray performance. More vendor-specific instrumental parameters also exist, such as the counter-flow directed drying gas (curtain gas) used to effectively reduce the amount of neutrals entering the mass spectrometer in Sciex instrumentations (Applied Biosystems, Ontario, Canada).

2.2 Collision induced dissociation (CID)

Tandem mass spectrometry (MS/MS) has been established as a useful technique for ion structure elucidation and in analyses of complex samples. Compared to single MS operation, MS/MS generates data of higher specificity, a feature that is exploited in quantitative and qualitative analyses in a wide variety of application areas. Over the last decade, MS/MS has become an important tool for peptide sequencing, structure determination and investigation of dissociation mechanisms of biomolecules. Collision induced dissociation (CID), as presently practised, derives its origin in works by Barber and Elliott and by Futrell et al. about meta-stable ions in sector instruments. Other pioneering studies in this research area were performed by Jennings and McLafferty.

Typically in CID experiments, a precursor ion is selected in the first mass analyser. The precursor is subsequently subjected to an inert collision gas in a confined compartment, called the collision cell, resulting in fragmentation of the precursor ion. The idea of a differentially pumped collision cell was first explored in the 1950s and 1960s. The generated product ions are analysed with a second mass analyser. A two-step process is generally adopted to simplify the description of the overall CID reaction, also called collision activated dissociation/decomposition. First, the precursor ion is excited internally through collisional activation, as part of the kinetic energy of the ion is converted into internal energy. Second, if enough internal energy is added the precursor ion breaks into fragments, as it undergoes unimolecular dissociation. This interpretation is readily supported by the typical CID reaction times observed.

Using Q instruments, low energy collisions (typically less than 100 eV) are used compared to sector instruments where collision energies as high as 8 keV are used. Vibrational excitation of a few tenths of an electron-volt is often brought about in a single low energy collision. This level of internal excitation is insufficient for most CID reactions. However, as a result of multiple collisions, the total internal energy is augmented to the point where CID is possible. The history, basic principles and various aspects of collisional activation of polyatomic ions have been discussed in a number of articles.
Physically in a mass spectrometer, CID reactions are generally obtained in a collision cell, in an ion trap or in the orifice-skimmer region, also called up-front CID. The amount of energy transferred during the excitation step will depend upon the size and pressure of collision gas, the kinetic energy of the precursor ion, and the reaction time between the precursor ion and the collision gas\textsuperscript{56,57}. These parameters are not all directly accessible to the mass spectrometrist, but rather indirectly changed via other parameters, and as a consequence, not easily optimized. In my opinion there is a lack of literature for the common MS user dealing with determination of optimal CID conditions. Thus, the work in Paper IV was performed in order to show how rather simple mathematical tools can be utilized to understand and visualize fragmentation behaviour using the enhanced product ion scanning function (EPI)\textsuperscript{58} of the Q TRAP\textsuperscript{TM} linear ion trap\textsuperscript{4} instrument.

Table 1. Explained variances for the models relating the signal intensities for various product ions of the proton adduct of bombesin to different CID conditions.

<table>
<thead>
<tr>
<th>Product ion m/z 396</th>
<th>m/z 509</th>
<th>m/z 566</th>
<th>m/z 680</th>
<th>m/z 808</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2-value (%)</td>
<td>91.2</td>
<td>82.6</td>
<td>84.0</td>
<td>85.1</td>
</tr>
</tbody>
</table>

As is indicated by the R2-values (the explained variances) in Table 1, the ability to model fragmentation behaviour using different CID conditions was very satisfying. The approach was further employed for optimization of CID conditions for LC-ESI-MS/MS analysis of siderophores in Paper VI.

2.3 Speciation analysis with ICP-MS

In many structural investigations, the ability to study specific elements, isotopes or oxidation states is beneficial or even crucial. A wide variety of compounds containing trace elements exist in biological systems. Examples are biosynthesized molecules with metal-carbon bonds such as selenoamino-acids binding Hg, and complexes between peptides, proteins, nucleobases, nucleosides and metals such as Cd, Cu, Zn, and Fe\textsuperscript{59}. Also, lipids and carbohydrates can coordinate metals\textsuperscript{59}. Many peptide-metal complexes exhibit regulatory, storage, catalytic, and transport functions in biochemical processes and are therefore essential\textsuperscript{60}. Other metal-containing anthropogenic environmental contaminants are, on the other hand, toxic for biological species.

The term speciation analysis is defined as the analytical activity of identifying and quantifying one or more chemical species of an element in a sample\textsuperscript{61}. Typically, the term speciation is used in reference to the analysis of the various forms of a specific element, i.e. the unbound element and the
different molecules containing or complexing the element, in a sample. One very powerful technique to study trace elements is ICP coupled to MS detection, and applications can be found in the field of environmental analysis, biochemistry, medicine and industrial chemistry. The ICP-technique offers sensitive elemental analysis of almost all elements in the periodic table. However, due to high background signals, measurements of H, C, N and O are problematic.

In the ICP-spectrometer, the sample is introduced either through self-aspiration or pumped to a nebulizer (pneumatic or ultrasonic), where an aerosol is formed. A spray-chamber is often incorporated in connection to the nebulizer to separate large droplets from small. The aerosol is then transported to the hot plasma (typically 5000-9000 K) consisting of a mixture of electrons, argon atoms and argon ions. The plasma is sustained through the utilization of an induction coil generating an oscillating electric and magnetic field. In the plasma the aerosol is vaporized, leaving the different components desolvated and susceptible for atomization and ionization. In ICP-MS, the formed ions are subsequently extracted and analysed with a mass analyser, i.e. a sector field, a Q, or a TOF analyser. More thorough discussions on the merits of different ICP-MS configurations together with useful references can be found elsewhere62-66.

One limitation of the ICP-MS technique is that the identification of a chemical species containing a particular element of interest can be troublesome. The parallel analyses of a sample with ESI-MS and ICP-MS enable the detection of both an organic and an inorganic moiety of a molecule67. The complementary information gained with ESI-MS can be especially useful in exploratory studies and chemical structure elucidation. This parallel ionization set-up is an attractive approach, which has been realized in the analysis of various compounds containing an ICP-detectable element as summarized by Wind and Lehmann68. Although ESI-MS also can be used for speciation69, ICP-MS often show lower limits of detection/quantification for hetero-atoms (elements other than C or H)70.
3 Hyphenation of LC and MS

The mass spectrometer is a highly selective detector and one could argue that direct MS analysis of raw, untreated, samples should be sufficient for sample component characterization/identification or quantification. But as the complexity of the sample increases the need for an additional step in the analysis is not only justified but may also be necessary. As in all analytical techniques, the composition of the matrix may affect the studied response, specifically the signal intensity of the monitored m/z may decrease or increase due to other ions simultaneously generated in the ion source. This phenomenon is frequently observed when ESI is applied\textsuperscript{71-75}. Due to the high vacuum and the advanced components used in a mass spectrometer, undesired constituents in the samples might also contaminate or even damage the instrument. As a consequence, clean-up of the sample prior to MS analysis is often performed. The sample preparation procedure can be executed either off- or on-line in numerous ways, depending on characteristics of the compound of interest and the overall sample composition. Commonly used sample preparation techniques include liquid-liquid extraction, solid phase extraction and different kinds of filtering. A very attractive procedure is to evaluate the potency of different sample clean-up methods on-line, as described by Bonfiglio\textsuperscript{76}. An alternative on-line evaluation approach has been developed by Kaufmann et al.\textsuperscript{77}. Both these approaches use a post-column infusion system to study matrix interferences. Using the former approach, a pump delivers a constant amount of analyte(s) into a T-connection where it is mixed with the LC effluent. Blank sample extracts are injected onto the LC system using the conditions chosen for separation and the selected MS response is monitored. A possible interfering compound contained in the sample extract is detected as a variation (suppression or enhancement) in the studied response. Using the latter approach, a high-frequency segmented addition of analyte is proposed instead. The obtained signals with and without analyte are superimposed, and any change in the spike height of these superimposed signals is attributed to matrix interferences.

Another way to reduce unwanted matrix effects is to incorporate a preceding separation technique into the experimental set-up. In addition, the inclusion of a separation technique will ideally also provide a selectivity enhancement. Thus, it is most beneficial if this technique relies on inherently different separation criteria, i.e. is orthogonal to MS. Gas chromatography (GC) was used in combination with MS already in the early 1950s, and since
both GC and MS work with chemical species in the gas phase their hyphenation is rather straightforward. Over the years, GC-MS has become a very valuable method, especially in the field of environmental analyses. One limitation working with GC is, however, the inability to analyse thermolabile and non-volatile compounds. Hence, due to the broad range of compounds amenable to LC, this technique was soon recognized as a valuable complement to GC. Another very profitable feature of LC lies in its flexible selectivity through the utilization of different chromatographic supports. Although all modes of LC have been tried together with MS (more or less successfully), reversed-phase liquid chromatography (RPLC) using hydrocarbon bonded phases, especially C18-phases, are most frequently employed today. One issue when coupling LC to MS is, however, the high amount of gas produced when the liquid from the LC-column is transferred to the gas phase. Early solutions to this problem were transporting devices with concomitant elimination of the solvent and molecular separators. The review by Abian78 provides a historical overview of the coupling of GC and LC with MS.

Today, liquid based separation methods coupled on-line to MS most often employ ionization at atmospheric pressure. The preferred ionization technique depends on the nature of the analyte, the separation method chosen and the accessibility of different ion sources in the laboratory. Each ionization technique has its own merits in hyphenation with LC, as described in various articles44,79-87. However, most applications of LC coupled to MS utilize ESI, owing its popularity to its sensitivity and ability to ionize a large variety of analytes. In contrast to other ionization techniques, multiply charged analytes can easily be formed with ESI making it possible to analyse large molecules as proteins and nucleic acids. Small polar to medium polar analytes are also readily ionized with this technique, making LC-ESI-MS useful in drug, metabolite and environmental studies. Other separation techniques used together with MS include capillary electrophoresis (CE) and electrochromatography (CEC). Various ways of combining more than one orthogonal separation technique with MS have also been suggested in literature, for instance ion exchange chromatography-RPLC-MS, LC-CE-MS and LC-ion mobility-MS88,89.

The hyphenation of LC and ICP-MS can be exploited in speciation analyses of complex samples containing hetero-atoms. In comparison with ESI, the ionization using an ICP-plasma suffers less from interferences caused by the sample matrix and/or the buffer additives. Hence, size exclusion, ion exchange and ion pair chromatography are readily coupled to ICP-MS90-92. One shortcoming of the ICP-technique is its low tolerance towards organic modifiers, such as methanol and acetonitrile, which are commonly added to the mobile phase in both RPLC and normal-phase liquid chromatography (NPLC). The organic solvent vapour will cool the plasma, possibly making it unstable or even extinguished. Heat of vaporization, surface tension and
viscosity are characteristics of the solvent that will have a considerable effect on the aerosol formation in a pneumatic nebulizer\textsuperscript{93}. The high gas load of small droplets formed with solvents having low surface tension and low heat of vaporization can be a potential problem. Thus, the utilization of RPLC and NPLC becomes restricted, and consequently special attention has to be paid during method development to assure long-term stability. The general hyphenation of LC and ICP-MS has been discussed by Sutton and Caruso\textsuperscript{90}, Larsen\textsuperscript{94}, Ponce de León et al.\textsuperscript{95} and Montes-Bayón et al.\textsuperscript{91}. The importance of different hyphenated techniques in the analysis of metal-biomacromolecular complexes is reported by Szpunar and Łobiński\textsuperscript{60}.

3.1 Method development in LC-ESI-MS and LC-ESI-MS/MS

In most quantitative LC-MS methods, low detection limits and reproducible results are desirable features. In order to achieve these requirements certain precautions have to be taken during method development. Knowledge about the chemistry and physics affecting the obtained results is crucial, and this knowledge has to be translated into selection of the right experimental conditions. As described earlier, there are several parameters affecting an LC-ESI-MS or LC-ESI-MS/MS run. Essentially, there are three different regions to consider: the LC system, the ESI ion source and the mass spectrometer as depicted in Figure 3.

![Figure 3. Schematic presentation of the three different regions of an LC-ESI-MS/(MS) hyphenated experimental set-up; the LC separation system, the ESI ion source and the mass spectrometer.](image)

The type and dimension of the separation column, the flow rate, the mobile phase composition during the chromatographic run and the applied temperature are parameters affecting the retention of a solute\textsuperscript{96}. Specifically,
in isocratic elution RPLC at ambient temperature, the selectivity of the mobile phase is determined mainly from the pH and the amount and type of organic modifier. To some extent the buffer type and ionic strength may also affect the selectivity.

When acids and bases are analysed, the pH of the mobile phase will be an important parameter during method development since it determines to what extent the solute is charged. The retention using reversed-phase stationary phases mainly implies hydrophobic interaction, therefore the possible charge on the analyte will have a profound impact on the obtained retention. To evaluate the change in retention due to the degree of ionization of the solute, a common recommendation is to try two different pH-values; one at pH = pKᵦ − 1.5 of the solute and one at pH = pKᵦ + 1.5. Worth mentioning in this context, is the change of pH that occurs when a buffer is mixed with an organic solvent. The change of pH will depend on the amount of organic solvent added and the type of solvent. Also, the pKᵦ-value of the solute is influenced by its surrounding environment, i.e. the type and amount of organic solvent and the ionic strength of the mobile phase. The buffer capacity, i.e. the ability to resist a pH-change with the addition of protons, is determined by pH, buffer pKᵦ and concentration. In order to ensure high buffer capacity, a buffer with a pKᵦ close to the desired pH should be chosen. In ESI it is important to use volatile buffers, hence the number of suitable buffers having a high buffer capacity at the desired pH becomes restricted. The buffers commonly used for RPLC separations together with ESI are ammonium acetate, ammonium bicarbonate, ammonium carbonate, ammonium formate and trifluoroacetic acid. The concentrations of the buffers used in LC ranges from about 10 to 50 mM. However, as high buffer concentrations causes ion suppression of the analyte in ESI, the maximum tolerated buffer concentration in LC-ESI-MS and LC-ESI-MS/MS is approximately 20 mM.

When it comes to choosing amongst different organic solvents, the so-called solvent-selectivity triangle offers excellent means of identifying solvents with different selectivity. Commonly used solvents are placed in the triangle according to their i) acidity, ii) basicity and iii) dipolar properties. From the solvent-selectivity triangle it is easily realized that methanol, acetonitrile and tetrahydrofuran display different solvent features. A reasonable approach is thus to try out all these three solvents or mixtures thereof during LC method development. In order to promote the formation of small droplets and aid evaporation during the ESI process, the organic solvent(s) used in the mobile phase(s) should have relatively low surface tension and heat of vaporization. Generally, higher ESI responses are obtained using higher percentages of organic solvent. However, very high amount of organic solvent (above 80 %) might actually result in a decreased ESI response. In order to force the analyte ions to distribute to the droplet surface, low solvent dielectric constant is also desirable. Moreover, other
chemical properties of the organic solvent such as the proton affinity will affect the production of ions in the gas phase\textsuperscript{35}. Methanol and acetonitrile have favourable chemical properties for ESI, and are the most frequently used organic solvents for LC-ESI-MS.

If gradient elution, i.e. programmed mobile phase composition, is applied, the number of experimental parameters increases even further. The start composition and the gradient shape, in terms of the slope or multiple slopes (using segmented gradients) or, with some pumps, possible curvature, have also to be determined. In order to facilitate the selection of the mobile phase composition, different generic strategies have been outlined in literature\textsuperscript{110-112}. The use of retention models to describe retention behaviour of the various solutes using different chromatographic conditions can also facilitate the LC method development. Retention models have been developed to predict solute elution behaviour employing different modes of LC, for example in reversed-phase\textsuperscript{113-116}, normal-phase\textsuperscript{113} and micellar\textsuperscript{117} liquid chromatography.

The column dimension and properties of the packing material, such as type and size, will determine the appropriate flow rate and injection volume\textsuperscript{118}. For a specific column, and in comparison between different packing materials, the influence of the flow rate on the chromatographic quality is usually visualized in a plot with reduced plate height as a function of reduced linear velocity\textsuperscript{119}. Optimal velocity is then determined as the velocity giving the lowest reduced plate height\textsuperscript{119}.

The effluent from the column is led to the second region, the electrospray ion source, where gas phase ions of the analytes are generated and transferred to the mass spectrometer. The ionization process is affected by a number of parameters, such as the mobile phase composition, the flow rate and the ESI voltage, as discussed in Section 2.1. Moreover, the response from a detector can be either proportional to the concentration of the analyte reaching the detector (concentration sensitive behaviour), or proportional to the amount of analyte reaching the detector per unit time (mass-flow sensitive behaviour). If the sensitivity of a hyphenated LC method is optimized, the column dimension should be chosen according to the behaviour of the detector system employed. Hence, large bore columns are preferred if a mass-flow sensitive detector is used and the sample volume is unlimited. ESI-MS often shows mixed behaviour, i.e. mass-flow sensitive at low flow rates and concentration sensitive at higher flow rates\textsuperscript{118}. It is therefore advisable to establish the expected behaviour at different flow rates for the specific ion source at hand. However, if the sample volume is limited it becomes necessary to miniaturize the chromatographic system, and different low-flow interface configurations have been constructed and evaluated\textsuperscript{118,120-122}.

The third region is identified between the entrance of the mass spectrometer and the detector. In this region the potential difference, also denoted
the declustering potential (on the Q TRAP™), in the orifice-skimmer region will affect the degree of declustering and up-front fragmentation of ions. In LC-ESI-MS/MS the parameters associated with CID in the collision cell, as described in Section 2.2, also have to be considered.

For obvious reasons, the tuning of hyphenated systems becomes more complicated as the number of parameters possibly having an effect on the studied response increases (see Table 2 for a summary). Also, as is evident from the discussion above, the chromatographic quality as well as the ESI process are affected by the mobile phase composition and flow rate, but probably with different optimal settings which makes method development more intricate. Questions to be dealt with include: How should the results be judged, i.e. what are the optimization criteria? What parameters will have a significant effect on the response? How should the optimal parameter settings for a particular criterion be determined? What strategies should be employed in the case of several optimization criteria?

Table 2. Experimental parameters possibly affecting the performance of an on-line LC-ESI-MS/MS analysis.

<table>
<thead>
<tr>
<th>Region</th>
<th>Influence on</th>
<th>Typical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC system</td>
<td>LC separation</td>
<td>Packing material, column dimension, flow rate, buffer type, pH and concentration, organic solvent type and concentration, gradient shape, temperature</td>
</tr>
<tr>
<td>ESI ion source</td>
<td>Ionization efficiency</td>
<td>Flow rate, buffer type, pH and concentration, organic solvent type and concentration, applied voltage, positioning of the ESI emitter, nebulizer and curtain gas flow, heating</td>
</tr>
<tr>
<td>Mass spectrometer</td>
<td>Up-front CID and declustering</td>
<td>Declustering potential</td>
</tr>
<tr>
<td></td>
<td>CID in the collision cell</td>
<td>Collision gas type and pressure, collision energy, reaction time</td>
</tr>
<tr>
<td></td>
<td>Trapping efficiencya</td>
<td>Linear ion trap fill time, scan rate</td>
</tr>
</tbody>
</table>

a Using the Q TRAP™ (Applied Biosystems, MDS SCIEX, Toronto, Canada).
Often, LC-MS hyphenated methods are optimized by varying one parameter at a time. In some instances the parameters associated with the mass spectrometer are the only ones considered. With few exceptions such procedures will fail to provide experimental conditions giving optimum performance, since no interactions between the parameters are taken into account. In fact it has been stated that at least two-thirds of the processes in the chemical industries indeed are affected by interactions\textsuperscript{123}. To increase the rate of success during method optimization, and probably also reduce the number of experiments, more systematic strategies should be applied. The field of chemometrics offers such means. In The handbook of Chemometrics and Qualimetrics: Part A\textsuperscript{124} the following definition of chemometrics is given: “Chemometrics is a chemical discipline that uses mathematics, statistics and formal logic (a) to design or select optimal experimental procedures; (b) to provide maximum relevant chemical information by analyzing chemical data; and (c) to obtain knowledge about chemical systems.” From this statement it is evident that chemometric techniques can be incorporated in all the different steps of an analysis, i.e. in planning and evaluation of sampling and sample pre-treatment, in the development of the analytical method, and also in the final result interpretation. The papers presented in this thesis show how chemometric methods can be used at various levels, e.g. for sample pre-treatment evaluation (Papers III and VII), for method optimization (Papers I, IV and VI), and for data evaluation and interpretation (Papers IV, V and VII).

Criteria and chemometric methods used for optimization purposes of LC-MS coupled systems are discussed in the following part of this chapter, together with available evaluation tools for interpretation of the obtained data. Also, in Section 3.2, I propose a strategy that facilitates the development of robust, sensitive and selective LC-ESI-MS and LC-ESI-MS/MS methods.

3.1.1 Optimizing criteria for LC and MS hyphenated methods

Optimization means quantitative improvement. Thus, in order to optimize an analytical method we first have to define the response that we want to quantitatively improve.

A general measure of performance in chromatography is elusive. The goal of separation is often to obtain sufficient resolution of all solutes in a sample in a reasonable length of time. If a single response is measured, like the UV-absorbance at a specific wavelength, completely overlapped peaks can not be tolerated. If several responses are measured simultaneously, like with MS or diode array detection, a certain degree of overlap in chromatographic separation can be compensated for as long as the detection method offers additional selectivity amongst the investigated analytes.
There is no universally accepted measure of chromatographic performance, but the measure first outlined by Morgan and Deming is frequently mentioned in literature. This measure is generally referred to as the chromatographic response function (CRF) and is calculated as:

\[ \text{CRF} = \sum_{i=1}^{j} \log(p_i), \]

where \( p_i \) is the peak separation of the \( i \)-th pair of adjacent peaks and \( j \) is the total number of pairs of adjacent peaks. However, numerous optimization criteria flourish in literature based on selectivity or resolution descriptors with or without considerations of asymmetrical peak shapes.

Often both adequate separation and short analysis time are desirable features of an LC-method, and hence different quantitative measures incorporating more than one quality descriptor have been developed. In multi-criteria decision making, different quality aspects are judged individually and quantitatively. Using a combined criterion called a utility function (\( U \)) a quantitative measure of \( m \) criteria (\( y_j \)) can be weighted (\( w_j \)) and summed using:

\[ U = \sum_{j=1}^{m} w_j y_j \]

One difficulty using utility functions is the a priori determination of the weights. Also, there is a potential risk of allowing an unacceptable value of one criterion with a low weight if a criterion with a high weight takes a high value. Instead of summing, Bourguignon and Massart suggested to multiply the different criteria. Every criterion is then scaled between 0 (unacceptable) and 1 (optimal) prior to multiplication of the so-called desirabilities. As the product is calculated, there is no risk that an unacceptable result of one criterion still gives a high global response. A review on different responses and response functions used in LC and CE can be found in a paper written by Siouffi and Phan-Tan-Luu.

In this thesis retention factors were used as chromatographic optimization criterion in Paper I. The goal was to obtain values higher than one. In Paper VII the calculated differences in retention times between any two adjacent chromatographic peaks were scaled (using a sigmoidal transfer function), summed and subsequently used as a global chromatographic response.

\(^*\)For two peaks, \( p \) is given by the depth of the valley drawn perpendicular to the time axis from a straight line connecting the two peak maxima to the lowest point in the valley divided by the distance from the connection line through the valley to the baseline.
Optimization of an MS response includes methods to improve the abundance of a specific precursor or product ion, or yielding as comprehensive or selective measurements as possible. Maximization of the signal or the signal-to-noise ratio (S/N) of a particular ion can be used for quantitative optimization, but choosing qualitative optimization criteria are less trivial. In the identification/characterization of peptides and proteins, the obtained sequence coverage can be used to transfer a qualitative criterion into a numerical response, as was done in Paper VII.

3.1.2 Chemometric tools for optimization

Multi-parameter optimization can be performed in two different ways, sequentially or simultaneously. In sequential optimization a few initial experiments are conducted. Guided by the results a search routine, usually the simplex procedure (see Section 3.1.2.1), directs the subsequent experiments step by step towards optimal conditions, i.e. maximum response. It is then essential that there is a single response function to optimize, which can be evaluated for each experiment. If there are several local maxima within the experimental domain (the region of allowed parameter settings) there is also a risk that the search procedure fails to locate the global optimum.

In simultaneous optimization, on the other hand, a larger number of experiments are performed according to a certain plan, the experimental design (also denoted design of experiments or DoE). A mathematical model is made that relates the response to the parameter settings, and the relationships can be visualized as response surfaces or contour plots. There are several different modelling techniques available and the choice of a specific model depends on the nature of the acquired data. Classical response surface modelling with a second-order polynomial is the most commonly used technique for optimization. The optimum is then derived from the fitted quadratic function.

In this thesis the chemometric techniques of particular interest are; the simplex routine, genetic algorithms (GAs), multiple linear regression (MLR), artificial neural networks (ANNs) and the fuzzy correlation approach (see Section 3.3.1). These techniques are well described in the literature, see for example Handbook of Chemometrics and Qualimetrics, and only a short summary of their different features together with a description on how they were used in this thesis will be given below. Complementary summaries on optimization methods used in LC can be found in review articles. Chemometric approaches for optimization of ESI or MS/MS parameters also occur in literature, and include work by Vaidyanathan et al., Riter et al., Charles et al., Mazsaroff et al., Nilsson et al., Varesio et al., Jiao et al., and Paper IV. Generic chemometric optimization procedures for LC-ESI-MS and LC-ESI-MS/MS methods considering both LC quality and MS or MS/MS sensitivity occur.
more rarely, but include work performed by Seto et al.\textsuperscript{138}, and Papers I and VI.

3.1.2.1 Simplex routines

According to Morgan and Deming\textsuperscript{139}, the simplex algorithm first developed by Spendley et al.\textsuperscript{140} was introduced in analytical chemistry by R. R. Ernst in the late 1960s. The simplicity of the simplex algorithm makes it a very popular optimization tool, and nowadays various modifications of the original simplex routine exist and are put into practice. The simplex is a hill-climbing search routine utilized for optimization purposes in a wide variety of application areas. The simplex algorithm can be visualized as a geometric figure with \((n + 1)\) corners, where \(n\) is the number of parameters investigated. In Figure 4 the mathematics underlying the modified simplex approach\textsuperscript{141} is depicted. Firstly, experiments according to the start simplex, denoted \(b_{nw}\), are performed and evaluated. The experiment in \(b\) gave the best response, while \(n\) gave the next-to-best and \(w\) the worst response, respectively. The centroid of the face remaining when the experiment yielding the worst response is eliminated is denoted \(p\). The settings for the new experiment are calculated by projection and expansion from the vertex to be eliminated through \(p\). The new vertex, \(r\), is calculated as:

\[
r = p + k(p - w),
\]

where \(k = 1\). The response in \(r\) is evaluated and now three scenarios are possible\textsuperscript{124}:

1. If the response in \(r\) is better than in \(b\) a positive expansion is tried, typically by setting \(k = 2\) (in the equation above), giving the \(e_p\). If the response in \(e_p\) is still better than in \(r\), then the new simplex is \(b_{ne_p}\). If the response is worse in \(e_p\) compared to \(r\) the expansion has failed and \(b_{nr}\) is the new simplex.

2. If the response in \(r\) is between \(b\) and \(n\) the new simplex is \(b_{nr}\).

3. If the response in \(r\) is worse than next-to-best a contraction is performed. If \(r\) offers the worst solution a negative contraction, giving \(c_p\) (typically with \(k = -0.5\)), is used giving the new simplex \(b_{nc_p}\). If the solution is between \(n\) and \(w\) a positive contraction takes place, giving \(c_p\) (typically with \(k = 0.5\)), and the new simplex is \(b_{nc_p}\).

The new simplex is evaluated and a new \(r\) is determined with elimination of the experiment offering the worst solution, except for when a contraction has taken place, then the next-to-worse vertex is eliminated\textsuperscript{125}. The procedure is repeated, with the algorithm directing the simplex towards higher responses, until the optimal settings have been identified.
The simplex approach can actually be utilized for two different situations, either in sequential optimization as guidance in what experiments to perform, or in simultaneous optimization as a tool to find the settings giving the highest response in a mathematical model.

In Paper VI the modified simplex method was utilized to search for the optimal gradient shape and buffer concentration in previously determined empirical models describing the retention behaviour of 14 hydroxamate siderophores.

### 3.1.2.2 Genetic algorithms (GAs)

Evolutionary ideas and terminology pervade the genetic algorithm\(^{142,143}\). The algorithm works with a population of individuals (sets of parameter settings), each being a candidate solution to the optimization problem. The individuals may “recombine” (exchange values) and “mutate” (change randomly) resulting in successively new generations of individuals. The probability to recombine varies; the better solution the individual offers, the higher is the probability to recombine.

Genetic algorithms are distinguished from many other optimization tools by the simultaneous investigation of many search trajectories. For this reason they can be applied to responses (and response models) with several optima, where a gradient-search method like the simplex could arrive at a local optimum.

In Paper I a genetic algorithm was used to find the optimal S/N ratio according to a trained artificial neural network. Each individual was the parameter settings of an experiment taken as the network input. The network outputs, i.e. the corresponding predicted S/N ratios, were used for
determination of recombination probabilities. Different initial populations were tried to confirm the optimal results.

3.1.2.3 **Multiple linear regression (MLR)**
Familiar to all analytical chemists is the procedure of establishing a linear (or maybe quadratic) mathematical relationship between a response, for instance the area of a chromatographic peak, and the concentration of the compound of interest. The philosophy behind this procedure can be expanded from the univariate situation (one single variable) to the multivariate situation, i.e. with several variables and possibly several responses. With MLR\(^{144}\) a linear relation is established between the independent variables \(X\) and one or more responses \(Y\):

\[
Y = X\beta + E,
\]

where \(Y\) is the response matrix, \(\beta\) is the matrix of regression coefficients and \(E\) is the error, or residual, matrix. The model matrix \(X\) contains the design variables (the parameters) and additional terms of interest, for instance interaction terms and/or quadratic terms. The regression coefficients are determined from:

\[
\beta = (X^TX)^{-1}X^TY,
\]

where \(X^T\) is the transposed matrix of \(X\). One shortcoming of MLR is that the number of experiments must exceed the number of responses. Other multivariate calibration techniques which cope for situations when the number of responses is higher than the number of experiments are principal component regression (PCR), partial least squares (PLS) and artificial neural networks (ANNs). An alternative approach is to do variable selection prior to MLR modelling\(^{145}\).

In Paper I the ability to model the S/N response of precursor ions was compared for MLR and ANNs, as discussed below. MLR was also utilized to determine optimal conditions for gallium exchange during sample preparation in Paper III. The effects of different instrumental parameters on the obtained fragmentation were determined using MLR models in Paper IV. Further, in Paper VI, MLR models relating the precursor ion intensity to the ESI conditions, as well as relating the product ion intensity to the CID and ion trap conditions were calculated.

3.1.2.4 **Artificial neural networks (ANNs)**
The construction and function of artificial neural networks (ANNs) are often described in terms of the human brain. ANNs are not really artificial brains, but the metaphor helps us to understand the activities of the neural networks. The brain consists of about hundred billion neurons, and between the neu-
rons there are synapses where signals can be transferred from neuron to neuron. In fact, each neuron may have more than 1000 synapses \(^{146}\). A multi-layer feed-forward neural network \(^{147}\) consists of operating units referred to as nodes or neurons. The nodes are arranged in layers, see Figure 5 for a graphical presentation. Each node is connected to all nodes in the next layer, and the information travels from layer to layer as determined by the connection weights \((w)\) and the transfer function of the nodes. There are different types of layers: input layer, hidden layer(s) and output layer. The number of nodes in the input and the output layers equals the number of variables and responses, respectively. The number of hidden layers and nodes has to be selected to match the complexity of the problem at hand.

![Diagram of an artificial neural network](image)

**Figure 5.** Schematic presentation of an artificial neural network.

The network has to be “trained” with a set of known values for \(X\) and \(Y\). The \(X\) values are introduced to the input layer and the network presents a response. This response is compared with the true \(Y\) value (the target) and the weights in the network are changed in order to decrease the error. The training procedure is repeated until the error reaches an acceptable level and the network can be used for prediction.

The neural network is very flexible as a modelling tool and it may adapt to virtually any relationship between \(X\) and \(Y\). The flexibility, however, increases the risk of overfitting calibration data, which in turn impairs the prediction ability. One way to control this is to use a separate test set of known \(X\) and \(Y\) to check the performance of the trained network. Another disadvantage in comparison with a regression model is that it is not obvious how to localize the optimum. One alternative is to use a search method like the simplex, but the possibility of multiple optima makes this method less straightforward.

As already mentioned, the capability of ANNs to model S/N of different ions obtained from ESI-MS experiments was investigated in Paper I. The models were further explored in LC-MS method optimization. Three different applications were studied, namely estriol in a mixture of estrogens,
ibuprofen in urine, and morphine together with codeine. Comparisons between the prediction performance with MLR and ANN are summarized in Table 3.

Table 3. Root mean squared errors of prediction for the test sets using MLR and ANN, respectively, concerning the prediction of S/N as a function of different ESI conditions.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Levels in the experimental design</th>
<th>MLR</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estriol</td>
<td>5</td>
<td>10.5</td>
<td>7.9</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>5</td>
<td>14.1</td>
<td>7.7</td>
</tr>
<tr>
<td>Morphine</td>
<td>3</td>
<td>13.1</td>
<td>16.2</td>
</tr>
</tbody>
</table>

Using MLR, a quadratic function was fitted to the calibration results. However, the designs used for estriol and ibuprofen comprised experiments at five levels, and this extra information could be utilized in the neural networks. Hence, prediction errors were lower for ANN compared to MLR for estriol and ibuprofen. For morphine, on the other hand, experiments were performed only at three levels and less extra information could thereby be incorporated into the neural networks.

3.2 A proposed stepwise optimization strategy

When several optimizing criteria are involved in the development of an analytical method, the practical procedure is less obvious. With the stepwise approach presented here the multi-decision problem is simplified, as only one type of response is considered at a time. The optimization strategy proposed aims at good chromatographic quality as well as high sensitivity in single or tandem MS operation (in terms of high signal or high S/N). The possibility to model several responses related to different goals, with no a priori selection for optimization, makes the response surface methodology a natural choice in the present context. The relatively large number of parameters, however, makes it necessary to select and handle groups of parameters in a stepwise manner. A flowchart of the proposed optimization strategy is presented in Figure 6. The different steps are: (i) screening experiments to identify the most important parameters, (ii) LC studies to ensure sufficient chromatographic separation, (iii) extended infusion experiments in order to maximize precursor signal(s), and in the case of tandem MS (iv) extended infusion experiments to determine optimal CID conditions and when applicable also ion trap settings. In the following text the various steps
will be discussed in general terms and in connection to the results obtained from Papers I, IV and VI. The parameters associated with an integrated LC-ESI-MS/MS method have been summarized previously, see Table 2 Section 3.1.

3.2.1 Identifying important parameters

In Paper I it was shown how screening experiments could facilitate determination of the most important parameters influencing the S/N ratio obtained for a specific precursor ion in a mass spectrometer. Reduced factorial designs were used for this purpose. However, the conclusions drawn from such experiments must be handled with caution since fractional factorial designs give information about the main effects but without considering all possible interactions. Hence, knowledge about the system is always important and should not be disregarded at this stage. Also, if the performance of the method is threatened to be impaired by other components (than the analyte) in the sample, the chromatographic quality has to be prioritized. It is not certain, not even likely, that high chromatographic quality and high MS sensitivity imply the same experimental settings. However, if sufficient separation is ensured, the number of parameters to investigate in more detail can possibly be reduced based on the results from screening experiments, and thus valuable time can be saved.

3.2.2 Ensuring liquid chromatographic quality

The LC studies are performed to investigate the experimental domain with respect to chromatographic quality and time of analysis. The goal is to ensure retention of all analytes and to maintain sufficient separation between
different components in the sample mixture. Short time of analysis is also desirable. The chromatographic quality can be expressed by different measures, as discussed in Section 3.1.1. Results from LC experiments might determine the setting for the parameters affecting the chromatographic quality, or they may impose limits on the experimental domain for some parameters. As the complexity of the sample increases, all potential resolution from the LC system is required in order to preserve selectivity and reduce negative matrix effects. Consequently, the settings of parameters affecting the chromatographic quality, but possibly also the MS quality, sometimes have to be determined based solely on the obtained separation efficiency. In order to ensure sufficient separation in Paper VI, the parameters affecting chromatographic quality were optimized based on this criterion alone. Retention models were utilized to predict the retention behaviour of the various solutes employing different experimental conditions. The band-tracking model was used to predict the retention times of the various analytes investigated, which proved out to be a very powerful and also convenient strategy. Moreover, retention models calibrated from linear gradients could be utilized successfully for the prediction of retention behaviour employing segmented linear gradients.

3.2.3 Obtaining high sensitivity for a precursor ion

When sufficient chromatographic separation has been ensured, it is time to investigate parameters associated with the ion source in more detail. The number of parameters might have been reduced based on the results from the screening and/or LC experiments. As already implied, the accessible range of certain parameters might also have been affected by the results obtained from the LC studies. In any case, the investigation at this stage should be more thorough than for the screening experiments, hence the experimental designs should comprise more levels and experiments. Preferably, these experiments should be performed by direct infusion of standard solutions containing one analyte dissolved in the corresponding mobile phase at elution.

From the experimental results empirical models can be calculated. Even though slightly better prediction ability was obtained using ANNs compared to traditional MLR models in Paper I, this technique is associated with more elaborate calculations. With conventional designs the models are usually low-order polynomials which can be evaluated rather easily, whereas the properties of a neural network must be explored in a more laborious way. Thus, the simplicity of MLR will probably make this technique more attractive to most LC-ESI-MS(/MS) practitioners.

In Paper I, the S/N from only one analyte was considered for each application, and thus all parameters could be optimized according to merely to that analyte. In Paper VI, on the other hand, 14 different analytes were
included. For these situations, when many analytes are included, it can be beneficial to separate between parameters affecting the spray performance, and those affecting the amount of up-front fragmentation. Probably, the former is mainly determined from the mobile phase composition, whereas the latter is determined from the structure of the specific analyte. Analytes eluting close together can then have identical ESI settings but individual declustering potential. In any case, the determination of the final optimal experimental conditions with respect to the ion source and DP is made from these experiments.

3.2.4 Obtaining high sensitivity for a product ion

In the case of tandem MS experiments the parameters affecting fragmentation also have to be optimized. Again, this is most easily executed by designed infusion experiments and empirical modelling of the obtained results. Most likely, the same behaviour is expected from different chemometric modelling techniques as in the case of modelling S/N for precursors. That is, ANNs could after proper validation probably provide more accurate predictions than MLR. However, the utilization of MLR gave satisfactory results both in Paper IV and VI, and thus neural networks were not further considered. From the empirical models, optimal CID and, when applicable, ion trap conditions can be determined.

The collision gas pressure is a slowly adjusting parameter. Hence, in LC-ESI-MS/MS applications with analytes eluting closely to one another, it is not possible to change the collision gas pressure between every analyte. A compromised setting must then be chosen for this particular parameter.

3.2.5 What is gained by the stepwise optimization strategy?

The total gain of a stepwise procedure is hard to elucidate, especially when different optimization criteria are used along the way. One attempt to give a measure of the accumulated gain was made in Paper VI. The different response improvements, defined as the calculated response at optimal conditions compared to conditions according to the middle of the design, were then multiplied. However, this accumulated gain only includes the increases in MS signal (from a precursor or a product ion) and no LC separation improvements are taken into account.

From the results reported, it is evident that both LC separation and MS sensitivity can be improved quite drastically by the implementation of a stepwise optimization procedure. Actually, one advantage of this approach is that the different optimization criteria can be dealt with separately, which facilitate the evaluation of the obtained results as only one criterion is judged at a time. Also, as smaller groups of parameters are considered in each step, the experimental procedure becomes less laborious. Moreover, many interac-
tions between experimental parameters have indeed been proven significant in the performed studies. Thus, the approach by varying one parameter at a time must be considered inadequate for this experimental set-up, and the implementation of experimental design is therefore a prerequisite for successful optimizations.

3.3 Interpretation of LC-MS and LC-MS/MS data

The information obtained from an LC-MS or LC-MS/MS run can be used for various purposes, for instance to quantify specific compounds in the sample or elucidate the structure of sample components. For quantitative purposes different calibration techniques (univariate or multivariate) can be employed. In addition, LC-MS profiles are also used for comparative or classification purposes. Multivariate chemometric techniques are important tools for such approaches.

3.3.1 Chemometric tools for fingerprinting/classification

The data produced with LC-MS can be visualized as contour plots (two-dimensional maps), with the separation in time on one axis and separation in m/z on the other axis. These maps are generally thought of as fingerprints, reflecting the sample composition. Much attention is put on the comparison between several fingerprints, obtained from replicates of the same sample or from different samples. For instance, in the field of medicine, comparison between healthy controls and patients with a disease can reveal possible biomarkers that can be used for diagnostic purposes. Possibilities, challenges, requirements and possible pit-falls for such an approach have recently been discussed in a special issue in Journal of Proteome Research. Other fields of application include pharmaceutical quality control and metabonomics.

Multivariate data analysis can unravel hidden structures in the huge amount of data typically generated with two-dimensional techniques, such as LC-MS. Principal component analysis (PCA) is the prime technique used to find these latent structures. Briefly, PCA decomposes a data matrix into a structure part and a noise part, which facilitates the visualization and interpretation of data. The original data are projected onto new variables (fewer than the original number), which describe most of the variation in the data set. This new variable space and the positions of the different samples in this space are described by loadings and scores, respectively. In a score plot, one score vector is plotted versus another. Similarly, one loading vector is plotted versus another in a loading plot. Score plots facilitate detection of possible outliers and identification of subgroups with similar responses within the data set. Loading plots can be used to identify important variables.
and the relation between them. Important issues during construction of PCA models are scaling of the data and estimation of the number of important principal components (PCs).

Although PCA is a very powerful technique, possible drifts in chromatographic retention and/or mass calibration can cause problems\textsuperscript{151}. Two different approaches to circumvent such problems are pre-alignment of the data before PCA analysis and implementation of peak-picking procedures and subsequent analysis on identified peaks rather than on chromatographic and mass spectrometric profiles. A review concerning data processing on different levels in comparative proteomics has been written by Listgarten and Emil\textsuperscript{152}. A generalization of PCA into higher order data is provided by parallel factor analysis (PARAFAC)\textsuperscript{153}.

In this thesis the approach developed at the department by Rolf Danielsson\textsuperscript{154,155} was utilized in the comparison between different fingerprints obtained from protein digests in Paper VII, see Section 4.3 for a more comprehensive discussion on the results from this study. The approach relies on the implementation of so called ‘fuzzy’ correlation. The idea is to use a dimmed or blurred version of correlation, when different fingerprints are compared. The fuzzy correlation is not as sensitive to variation in one dimension (or both), as ordinary PCA. Time binning of the data, i.e. summing responses together in defined time intervals (time bins) and thereby reducing the time resolution, can also be incorporated into the calculations in order to compensate for undesired time shifts in the data set.

In the fuzzy correlation approach, two runs are compared at a time, one being binned and one being both binned and blurred. The correlation matrix is calculated between all pairs of runs. Prior to further calculations the elements in the matrix are scaled to unity in the diagonal, resulting in a similarity matrix. The eigenvectors of the similarity matrix are subsequently multiplied with the square root of the corresponding eigenvalue to give fuzzy score vectors in conformity with PCA. Just like in ordinary PCA, score plots can be drawn to visualize similarities and differences between the runs in the data set. Complementary to scores in PCA are the loadings, describing the contribution from the original variables to the principal components. For the fuzzy correlation approach, the loadings for a PC is calculated by back correlating the score vector to the blurred versions of the runs. These similarity measures can be utilized to illustrate which m/z channels or chromatographic time bins that are responsible for most of the variations found in the data set.
4 Selected applications using LC coupled to MS

Even though much of the analytical equipment is the same, there is often a distinction made between analyses of small and large compounds. In this thesis both small (drugs), medium (siderophores) and moderately large molecules (peptides) have been studied with LC-MS. The work has been performed as to explore the capability of this hyphenated technique in different areas, rather than to focus on one specific application. In the following text the backgrounds of and the results from different projects are summarized.

4.1 Drug analysis

In the analysis of drugs and drug metabolites LC-MS has been established as a very useful technique\textsuperscript{156,157}. The applications involve the screening for potential new drugs, quality control/assurance measurements and detection/quantification of drugs in different biological matrices. In Paper I the drug ibuprofen was analysed in urine, and even though the main purpose of the paper was not to develop a new analytical method to detect ibuprofen in a biological matrix, the usefulness of LC-ESI-MS for drug analysis was still demonstrated.

In Paper VIII the bioactivation of the prodrug azathioprine (Aza) was studied. Aza is clinically used for immunosuppression after organ transplantation and in the treatment of autoimmune diseases\textsuperscript{158}. The activation mechanism suggests the release of 6-mercaptopurine (6-MP) from Aza, which is further metabolized to active metabolites with immune-modifier activity\textsuperscript{158}, see Figure 7. Even though Aza has been used in the field of medicine since the 1960s, the actual process of bioactivation has remained obscure. A chemical nucleophile, such as a thiol, can displace the imidazole group from Aza thereby releasing 6-MP. The most abundant low-molecular-weight thiol in the cell is glutathione (GSH)\textsuperscript{159}; a reasonable assumption is thus that GSH plays an important role in the biotransformation of Aza. It has been shown that a group of enzymes called glutathione S-transferases catalyze the conjunction of GSH and thiols\textsuperscript{160}. Since high levels of these enzymes are found in many tissues, such as liver, placenta, adrenal glands,
and gonads, their potential role to catalyze the bioactivation of Aza was tested. Analyses with LC-ESI-MS were performed to confirm the increasing concentrations of 6-MP and GS-imidazole accompanying the decreasing concentration of Aza, as the enzymatically catalyzed reaction continued. In order to ascertain the chemical structure of the dominating reaction product the mass spectrum for the corresponding chromatographic peak was inspected.

![Figure 7](Image)

Figure 7. Proposed metabolic pathway of Aza and mechanism of action\(^\text{158}\). 6-TGNs are believed to induce apoptosis, i.e. programmed cell death, while me-TIMP blocks cell proliferation through inhibition of adenosine triphosphate and guanosine triphosphate de novo biosynthesis\(^\text{158}\). Abbreviations are Aza, azathioprine; GSH, glutathione; GST, glutathione S-transferase; 6-MP, 6-mercaptopurine; HPRT, hypoxanthine-guanine phosphoribosyl transferase; TIMP, thioinosine 5’- monophosphate; 6-TGNs, 6-thioguanine nucleotides; XO, xanthine oxidase; 6-TU, 6-thiouric acid; TPMT, thiopurine S-methyl transferase; 6-MMP, 6-methyl-mercaptopurine.

From other measurements in Paper VIII it was concluded that the amount of 6-MP released through glutathione S-transferase catalyzed reactions could be responsible for more than 90% of the biotransformation of Aza. This finding had not been demonstrated before.
4.2 Siderophore analysis

Trace amounts of metals, like iron, are essential for growth and proliferation of cells. The low solubility of ferric ions under aerobic conditions makes the available concentration of iron(III) low (approximately $10^{-18}$ M at pH 7)\textsuperscript{[61]}). This is far below what is needed for living cells, as an example typically a concentration higher than $10^{-7}$ M iron(III) is required for normal growth of plants\textsuperscript{[62]}. Hence, active acquisition mechanisms are utilized by many bacteria, fungi and plants. Through the secretion of low molecular mass organic acids (LMMOAs), like oxalate and citrate, and very strong iron chelators called siderophores water-soluble complexes with iron(III) are formed. The concentration of LMNOAs in soil solutions ranges from 1-50 µM for di-/tricarboxylic acids to 0-1 mM for monocarboxylic acids\textsuperscript{[63]}, while the concentration of siderophores is considerably lower (low nano-molar)\textsuperscript{[64,165]}. Still, the very high complex formation constants of siderophores, approximately $10^{30}$ for trihydroxamates\textsuperscript{[66]}, make these compounds highly efficient as iron sequesters. The dominating functional groups of siderophores facilitating complex formation are hydroxamates, catecholates, and phenolates. While bacterial siderophores are structurally diverse, fungal siderophores are dominated by hydroxamate siderophores\textsuperscript{[66]}. Extensive reviews on different siderophores can be found elsewhere\textsuperscript{[166-168]}.

Siderophores and LMNOAs are believed to contribute to the formation of podzol profiles in soil\textsuperscript{[69]}. Podzols are characterized by distinct layers of different chemical composition at various depths in the soil, see Figure 8 for a graphical presentation. Podzols are found mainly in cold humid climates under forest and heath vegetation. Below the organic surface (the mor layer), an ash-grey weathered layer, called the eluvial horizon, and a reddish-brown layer, called the illuvial horizon, can be found. Further beneath this, in the C-horizon, unaffected minerals are situated. A more detailed definition of a podzol can be found in World reference base for soil resources\textsuperscript{[170]}. In literature mainly two different processes have been proposed to explain the formation of podzols: (i) downward transport of aluminium and iron through complex formation with organic acids, and (ii) silicate weathering followed by downward transportation of aluminium and silicon\textsuperscript{[171]}. Evidence for siderophore and LMNOA promoted dissolution of iron from soil minerals have been found performing experiments in the laboratory, see Kalinowski et al.\textsuperscript{[172]} for a summary, and attempts to describe the mechanisms behind the siderophore-mediated uptake of iron on a molecular level have also been published\textsuperscript{[173]}.
In addition, siderophores are compounds of interest in the field of medicine. The strong binding of iron by siderophores is exploited to treat metal overload diseases like β-thalassaemia, aplastic anemia and sickle cell anemia. Siderophores have also shown capabilities of stopping proliferation of aggressive tumour cells in clinical trials. Currently, desferrioxamine B is the clinically preferred siderophore, but various analogues have been developed showing potent iron-sequestering capabilities.\(^{174}\)

Evidently, the interest for siderophores takes many perspectives and covers many fields of science; hence analytical methods to study these compounds in soil solutions, cultured samples as well as in biological fluids are of great importance. As these compounds are often found in low concentration in complex matrices, the analytical methods employed for identification and quantification have to provide excellent sensitivity and selectivity. The hyphenation of LC and MS offers such means, which is demonstrated by the increasing number of papers published on this topic since the turn of the century\(^{164,175-178}\).

Paper II presents a novel approach to the determination of three structurally similar trihydroxamate siderophores belonging to the ferrichrome family, namely ferrichrome, ferrichrysin, and ferricrocin. The analytical method is based on a column-switched capillary LC system with ESI-MS detection. The separation is accomplished on a porous graphitic carbon (PGC) column using isocratic elution. In contrast to the brush-type surface of C18 materials, PGC exhibits unique selectivity explained by its layered flat surface\(^ {179}\). The column-switched technique enables sample clean-up and pre-concentration of the analytes on-line\(^ {180,181}\), and in comparison with other
earlier published methods a more selective and sensitive method was obtained. Figure 9 shows chromatograms obtained when injecting a standard onto the column-switched system.

Figure 9. Selected ion chromatograms of ferrichrome, ferrichrysin and ferricrocin.

A natural continuation of the work performed in Paper II was to employ the developed method for quantitative determinations of siderophores in real samples. In Paper V the method was utilized analysing natural field soil solution samples collected in the north and south of Sweden taken at different depths in podzol profiles. Quantitative analysis of individual ferrichromes in soil solution had only been reported once previously\textsuperscript{16}. Concentrations of 0.1–12 nM ferricrocin were found in all podzol horizons, i.e. mor layer, eluvial, illuvial and C-horizon, with lower concentrations in the lower horizons, i.e. the illuvial horizon and the C-horizon. Similarly, the highest concentration of ferrichrome, i.e. up to 2.1 nM, was found in the mor layer and the eluvial horizon and decreased with depth until below the detection limit. In contrast to this, no ferrichrysin was found in any of the samples in this study. The results were also put in relation to other findings at the locations, such as pH, concentration of dissolved organic carbon (DOC), and concentration of different LMMOAs. By PCA it was shown that correlation exists between the concentrations of DOC, citric, acetic, oxalic acid, and ferricrocin.

However, due to the complexity of natural samples, unknown siderophores are not easily detected with LC-ESI-MS. The elemental selectivity of
ICP-MS makes it suitable for analysis of siderophores in natural samples. Unfortunately, the main isotope of iron, $^{56}\text{Fe}$, suffers from severe spectral interference from ArO$^+$, which makes analysis of Fe problematic with ICP-TOFMS. However, as siderophores form complexes with other trivalent ions (e.g. aluminium, chromium, gallium, and indium) an alternative approach was developed in Paper III by exchanging iron for gallium in a reducing environment and subsequently monitoring the $^{69}\text{Ga}$ trace with LC-ICP-MS. The obtained Ga traces can thus be regarded as fingerprints of the trivalent metal binding compounds contained in the samples. The parallel analysis with LC-ICP-MS and LC-ESI-MS/MS was also evaluated as a tool for identification and structural elucidation of unknown siderophores in soil solutions.

Even if the PGC material offers unquestionable selectivity between the analytes studied in Papers II, III, and V, some problems were encountered trying to incorporate a larger number of siderophores into the method. Unacceptable peak shapes were obtained for some siderophores, and consequently C18-material was considered once again in Paper VI. Now, the material was purchased from a different vendor, and with the utilization of a gradient the separation was improved considerably. The stepwise optimization strategy described in Section 3.2 was used during method development. In Figure 10 selected ion chromatograms obtained at optimal chromatographic conditions for the various siderophores are displayed.

![Figure 10](image-url)

*Figure 10.* Selected ion chromatograms of 14 hydroxamate siderophores using a linear segmented gradient. Obtained selected ion chromatograms using optimal chromatographic conditions. Elution order; Fe(III) dimerum acid complex (A), ferrioxamine B (B), neocoprogen II (C), ferricrocin (D), ferrichrome (E), ferrichrysin (F), ferrioxamine E (G), linear fusaramine B (H), neocoprogen I (I), coprogen (J), ferrirubin (K), linear triacetylfusaramine B (L), ferrirhodin (M) and finally cyclic triacetylfusaramine C (N).
In fact, with the utilization of a segmented gradient, almost complete separation among the siderophores could be accomplished. Further, the optimized method included MS/MS detection of the various siderophores.

4.3 Protein and peptide analysis

Proteins, peptides and amino acids, the building blocks of all living species, attract a lot of interest in the scientific community today and have done so for quite some time. Nowadays, researchers in the field of life science are trying to explain biological function as well as dysfunction on biochemical grounds, and often peptides and proteins play important roles in these puzzles to solve. Often very subtle chemical modifications to a protein/peptide change the biochemical function of the molecule entirely. Post-translational modifications (PTMs) are such processing events that change the properties of a protein. The processing can imply either a proteolytic cleavage of, or a modification to one or several amino acid residues. PTMs play many different roles in biochemical processes. For example, they can be used to control the activity state of the protein, i.e. active or non-active. Common modifications include phosphorylation of tyrosine, serine and threonine residues, acetylation of the N terminal or lysine residues, and methylation of arginine residues\textsuperscript{182}. Various forms of glycosylation at asparagine (N-linked) or at serine or threonine (O-linked) compose another group of PTMs\textsuperscript{182}.

The so-called “bottom-up” approach, also referred to as “shotgun proteomics”, is a commonly used strategy to determine the protein content of a sample, for instance a biological fluid or a tissue. Then, the proteins are digested (cleaved) prior to analysis and the resulting peptides are used to elucidate the primary structure of the parent material. Digestion of a protein can be accomplished both chemically and enzymatically. Typically, cyanobromide or a strong acid, like hydrochloric acid, is used for chemical digestion, while e.g. trypsin and Glu-C are frequently used proteases (proteolytic enzymes). In order to increase the degree of success during digestion, i.e. decrease the amount of missed cleavages, it is important to denature, unfold, the protein. Denaturation of proteins can be obtained by exposure to high temperature, an organic solvent, or a high concentration of buffer, typically guanidine-hydrochloride or urea. Another important step included in the digestion schemes is the breakage of potential disulphide bridges between cysteine residues. An almost unlimited number of different protocols for enzymatic digestion exist in literature, making the choice of a particular one more or less arbitrary. As a result, the researcher often relies on whatever has become laboratory tradition.

The prime method for protein and peptide separation is LC, but also CE separation methods are common. The reader is encouraged to take part of
cited references for a further insight into the various aspects of protein and peptide LC separation.

Mass spectrometry has been established as a very important tool for proteomics since the breakthrough of ESI and MALDI in the 1980s. These “soft” ionization techniques facilitate the analysis of intact proteins, peptides and even non-covalent complexes. In single MS operation mode the mass of the intact protein or the masses of the peptides generated via chemical or enzymatic digestion can be determined, whereas tandem MS operations provide the opportunity of complete primary structure determination. The work performed in Paper IV shows the prospects of implementing chemometric techniques in the evaluation of tandem MS data generated from a peptide precursor in a linear ion trap. Not only can the empirical models be utilized to determine experimental conditions yielding maximum response for selected product ions, but the results can also be interpreted in terms of characteristic mechanisms of CID reactions of peptides.

The analysis of digested proteins in pharmaceutical or biological samples is often performed with LC-ESI-MS. Studying the literature, however, it is apparent that the digestion protocols applied differ significantly. Therefore, three different experimental procedures for tryptic digestion were compared in Paper VII in order to study possible induced variations in the obtained peptide samples originating from the different pre-treatments employed. A schematic presentation of the investigated protocols is outlined Figure 11. Three different proteins were investigated in parallel, namely bovine serum albumin (BSA), casein and fetuin. The protein digests were analysed with LC-ESI-MS.

**Figure 11.** Different tryptic digest protocols used prior to LC-ESI-MS peptide mapping in Paper VII.
Multivariate data analysis, as well as comparisons based on a more traditional measure, i.e. obtained sequence coverage, was performed for each protein. From the results it was clearly evident that the three protocols investigated, indeed display different behaviour when their LC-MS profiles were compared. Several causes can be responsible for this discrepancy, for instance unequal digestion efficiencies. Moreover, losses of particular peptides may occur e.g. during desalting using the Zip-Tip approach. Also, induced artefacts might differ between the protocols being investigated.

Figure 12. Score plots for the first two fuzzy PCA components (PC1 and PC2), with the experiments distributed around origo (+). The different digestion protocols are indicated (Δ) for the Zip-Tip approach, (□) for the size exclusion column approach, and (○) for the 50% acetonitrile approach.

Figure 12 shows the score plots obtained for the three proteins using the fuzzy correlation approach described in Section 3.3.1. The plots clearly indicate differences between the protocols in the case of fetuin and casein, as the experiments using one particular digestion procedure groups together. In the case of BSA less pronounced differences were observed between the digestion procedures being studied.
5 Concluding remarks and future outlooks

It is my firm conviction that LC-MS will continue to be a very important analytical tool for a long time to come. From being a technique merely used by specialists it has over the years become indispensable for the everyday activity in many laboratories within various disciplines. However, it is just now when researchers from different fields of expertise are starting to explore the merits of LC-MS, that we have the possibility to fully realize what this technique has to offer. And as the complexity of the scientific questions that we try to solve using LC-MS increases, the need for more sophisticated data analysis also increases. Therefore, the utilization of chemometrics becomes obvious, and also explains why chemometric techniques were first introduced in the analysis of low resolution mass spectra\textsuperscript{185-188}. Since that date, the application of MS and other spectroscopic techniques has been one of the driving forces for development of new tools for multivariate data analysis.

In this thesis it has been demonstrated how the performance of LC-MS hyphenated methods can be improved quite drastically by the implementation of chemometric optimization. And it is my hope that the generic stepwise procedure presented will contribute to a more strategic approach to the further exploration of LC-MS hyphenated methods. The accounted results in this thesis also prove the general applicability of LC-MS.

A significant part of this thesis is devoted to siderophores analyses, and thus their future fate also has to be commented upon. I think that the scientific interest in siderophores will continue, and other previously not investigated roles of these remarkable compounds will be further studied. For instance, my co-workers at Mid Sweden University are now looking into the function of siderophores in the Baltic Sea. In addition, the possible binding of other metals to siderophores is studied at my department. For the interested party, it can be noted that while there is still a need for new sensitive and selective analysis methods to detect unknown siderophores, recently, siderophores themselves have been used as tools in chemical analysis. For instance, desferrioxamine has been used to chelate iron in biological tissue\textsuperscript{189} and plant samples\textsuperscript{190} for the subsequent determination of the total concentration of Fe(III). Also, siderophore-doped sol-gels have been used to determine free concentrations of Fe(III) in water and blood samples\textsuperscript{191,192}. 
6 Acknowledgements

My supervisors, Karin Markides, Jonas Bergquist and Dan Bylund. Karin, for giving me the opportunity to perform my PhD studies at the department. Jonas, for encouragement during the writing of my thesis and for being my supervisor during the last years. Dan, whom introduced me to the field of LC-MS, for many inspiring discussions on LC-MS, chemometrics and related topics. Dan is also acknowledged for all support and kind considerations during all these years, and for continuing to be my supervisor even after leaving the department and Uppsala.

Anna, Jenny, Per, Rolf and Salim for reading and commenting on my thesis. Rolf is also acknowledged for valuable contributions to Papers I and VII.

My collaborators at BMC; Daniel Bäckström, Birgitta Eklund and Eva Nilsson. The more, the merrier!

The fruitful and educating collaboration with Ulla Lundströms group at Mid Sweden University in Sundsvall. Special thanks to Sofia Essén and Sara Holmström!

Arde, my former room mate, for many interesting talks about all-kinds of things but chemistry, and of course for all help and advice concerning LC-MS. Jakob, my present room mate, for nice company during my last year.

Former and present colleagues at the department, especially Aida, Anna and Jenny, for making this time very pleasant and memorable. Stefan and Magnus for invaluable co-work with Doktorandföreningen.

My friends, whom I appreciate more than I have time to hang out with. Anette, I will visit you and your family in Oslo very soon. Anki, who has known me for probably too long. Anna S., Fagersta is not that far away after all. Anna T. and Jenny S. for long discussions about the life of a graduate student and for becoming two very dear friends of mine. Jenny W. and Veronica for many nice times. Jenny L., for always being there as my friend and making me in a good mood, no matter what. Kram på er allihop!
Jamila, for helping us out and taking such good care of my children when needed.

My mother and father for kind considerations and for being the very best grandparents. My sister, for showing other ways to spend a good time.

Finally, I owe a lot of gratitude to my dear family. Karim, without whom I would not have succeeded with this work. You are my very best friend in the world and my life-long companion. Our wonderful children, Sam and Finn, whom make me so happy and proud. I love you all!

"Mamma på jobbet." Teckning av Sam Moberg 2006.
7 Swedish summary

7.1 Introduktion till LC-MS

I takt med att mer komplexa prover undersöks måste också mer sofistikerade analytiska metoder utvecklas. Många gånger räcker det inte med att en enskild teknik nyttjas, utan flera tekniker måste sammankopplas för att få fram önskad information. Innan komponenterna i ett komplext prov ska bestämmas kan de med fördel först separeras ifrån varandra. Vätskekromatografi (LC) är en separationsteknik som kan åstadkomma detta. Identifiering eller åtminstone karaktärisering av de separerade komponenterna kan göras med masspektrometri (MS). Komponenterna lösta i mobilfas överförs då till gasfas, exempelvis med hjälp av elektrospray jonisation (ESI), och detekteras därefter i en masspektrometer, där bestämningen görs utifrån förhållandet mellan massa (m) och laddning (z), d.v.s. m/z. Ihopkopplingen av LC och MS ger således upphov till data med separation både i tid och i m/z. För den intresserade finns ett stort antal övergripande artiklar som behandlar LC-MS att hitta i litteraturen.

7.2 Utförda delarbeten

I artiklarna I, IV och VI redovisas hur ett systematiskt användande av kemometriska verktyg kan underlätta utveckling av LC-ESI-MS- och LC-ESI-MS/MS-metoder med hög prestanda. Sammanfattningsvis föreslår jag en strategi som består av ett stegvist förfarande. Fyra delsteg kan identifieras, nämligen (i) sällningsförsök för att identifiera parametrar med stor inverkan, (ii) LC-experiment för att säkerställa tillräcklig kromatografisk separation, (iii) infusions-experiment för att maximera responsen för en moderjon, och slutligen i fallet av en LC-ESI-MS/MS-metod (iv) infusions-experiment för att maximera responsen för en eller flera dotterjoner. I optimeringsprocessen ska experimentell design användas för att planera vilka experiment som ska utföras, sedemera ska empirisk modellering av erhållna resultat användas för att kunna förutspå optimala betingelser.

Beroende av provets karaktär kommer olika mycket LC-optimering att föregå MS-optimeringen. Om riktigt svåra kromatografiska separationer ska lyckas kan det krävas att alla parametrar som påverkar separationen optimeras med avseende på denna enbart, för att sedan optima resterande parametrar med avseende på den valda MS-responsen. År däremot förestående separationsproblem av en enklare karaktär, så kan parametrar som påverkar så väl LC- som MS-responsen väljas mer utifrån den slutgiltiga responsen, d.v.s. MS-responsen.

Ett antal olika applikationer av LC-MS presenteras också i denna avhandling. Analyterna som studerats har haft varierande storlek, från små läkemedelsföreningar till betydligt större peptider från enzymatiskt klyvda protein.

I artikel VIII används LC-ESI-MS för att säkerställa den dominerande reaktionsprodukten vid en enzymkatalyserad reaktion mellan pro-läkemedlet azathioprine (Aza) och glutation (GSH). Utifrån data som presenteras i denna artikel är det tydligt att en avgörande del av bioaktiveringen av Aza sker via enzymkatalyserad reaktion med GSH, vilket tidigare inte visats.

En omfattande del av denna avhandling ägnas vidare åt LC-MS-baserade analyser av sideroforer. Sideroforer utsöndras från mikro-organismer, svampar och en del växter för att öka tillgängligheten av järn(III) som annars finns i för låg koncentration för att täcka många organismers behov. Sideroforer används också i medicinskt syfte för att behandla överdosering av järn(III) i kroppen.

En LC-ESI-MS-metod för kvantitativ bestämning av tre, strukturellt mycket lika, sideroforer utvecklades i artikel II. I artikel V användes denna metod för att analysera markvattenprover funna på olika djup i podsoliska jordprofiler i södra och norra Sverige. Kvantifiering av enskilda ferrikromer i markvattenprover hade bara rapporterats en gång tidigare i litteraturen\textsuperscript{164}. I artikel VI utvecklades en LC-ESI-MS/MS-metod för ett större antal sideroforer. Det föreslagna stegvisa optimeringsförfarandet användes för att
utveckla en selektiv metod med god känslighet. Kromatografisk separation av siderofofera åstadkoms med hjälp av en segmenterad gradient.

En alternativ jonisationsteknik, nämligen ett induktivt kopplat plasma (ICP), användes i artikel III för att studera den oorganiska delen av siderofofer. Tyvärr är det inte möjligt att analysera låga halter av järn med lågupplösande ICP-MS, varför det inbundna järnet först byttes ut mot gallium i en reducerande miljö innan analys med LC-ICP-MS utfördes. De insamlade gallium-profilerna kunde sedan användas som fingeravtryck av provets innehåll av siderofofer.

Slutligen har olika protokoll för att enzymatiskt klyva proteiner jämförts i artikel VII. Jämförelsen har baserats på resultaten från LC-ESI-MS-analysen av samtliga prover.

7.3 Utsikter för framtiden och avslutande reflektioner

I denna avhandling har resultat presenterats som demonstrerar hur LC-MS kan användas inom en mängd forskningområden, både som kvantitativ teknik för utvalda analyter och i explorativt syfte för att undersöka ett provs olika komponenter. Avhandlingen visar också hur man med relativt enkla matematiska verktyg kan underlätta och också förbättra utvecklingen av robusta, selektiva och känsliga LC-MS-metoder. Resultat som rapporteras visar dessutom hur viktigt det är att den kemiska informationen som finns i provet bibehålls till den slutgiltiga LC-MS-analysen för att relevanta tolkningar om provets natur ska kunna göras.

Det är min övertygelse att MS kommer att fortsätta vara en mycket frekvent använd analysverktyg långt framöver, mest beroende på dess selektiva detektion, dess generella tillämpning och i de flesta fall också goda känslighet. Eftersom LC-MS har blivit en mogen analyseteknik har fokus för forskningen delvis flyttats från utveckling av hårdvaran till efterbearbetning och utvärdering av insamlade data. För närvarande riktas ett stort hopp till MS som möjlig teknik att upptäcka så kallade biomarkörer, föreningar som bl.a. kan studeras i ett sjukdomsdiagnostiserande syfte. Det återvunna intresset för föreningar som innehåller både en organisk och en oorganisk del, gör också ICP-MS till en teknik vars antal utövare med säkerhet kommer att öka. Flera applikationer av ICP-MS kommer framför allt att rapporteras för analyser av biomolekyler innehållande ett eller flera detekterbara element.

Studier av siderofofers funktion och inverkan kommer också att fortgå, och användningen av MS kommer vara ett viktigt verktyg för att kunna säkerställa/undersöka strukturen hos dessa föreningar. Användningen av ICP-MS inom detta område kommer troligtvis också att utökas, exempelvis undersöks redan nu inbindningen av andra metaller till siderofofer med ICP vid avdelningen.
8 References


111. Ghrist, B. F. D.; Snyder, L. R. Design of optimized high-performance liquid chromatographic gradients for the separation of either small or large molecules III. An overall strategy and its application to several examples. J. Chromatogr. 1988, 459, 43-65.


190. Fernández, V.; Winkelmann, G. The determination of ferric iron in plants by HPLC using the microbial iron chelator desferrioxamine E. BioMetals 2005, 18, 53-62.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 201

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-7071