**Introduction**

Hyphenation of efficient separation methods such as high performance liquid chromatography (HPLC) or capillary electrophoresis (CE) with detection methods that provide structural or functional information is a powerful tool for the analysis of biological macromolecules. In particular, characterization of complex molecules such as proteins, peptides and carbohydrates necessitates the use of combined analytical methods to assign molecular structures to the peaks occurring during chromatography or electrophoresis. The combination of HPLC and mass spectrometry (LC–MS) is particularly powerful and a mainstay method in most laboratories dealing with the analysis of biomacromolecules.

Recent years have seen a shift from hypothesis-driven biological research towards the global analysis of the cell components, tissues or living organisms. The main goal of this research is to gain an understanding of the interplay between molecules at different levels within a biological system in an effort to generate new hypotheses that can be tested and will, hopefully, lead to new mechanistic insights. Systems biology, as this kind of research is often named, relies on powerful analytical methodologies such as DNA and protein arrays, two-dimensional gels, HPLC and mass spectrometry. In this article an overview of recent LC–MS developments in the areas of proteomics and glycoprotein analysis will be given. LC–MS is becoming a viable alternative to two-dimensional gel electrophoresis in proteomics and plays a prominent role in the analysis of the carbohydrate portion of glycoproteins as biopharmaceuticals.

**Developments of LC–MS in Proteomics**

With the completion of the Human Genome Project, scientific interest is turning increasingly to the task of converting DNA sequence information into knowledge that will potentially improve human medicine and healthcare. The surprising discovery that the human genome contains less genes than anticipated stresses the importance of proteins as the primary actors in biological systems. Analysis of a proteome, the entire set of proteins expressed by an organism, cell or organelle, is highly complex and under constant flux, in response to various stimuli, over time. Proteomics has to deal with the separation of highly complex protein mixtures. In addition, low-abundance and high-abundance proteins differ in concentration over more than five orders of magnitude, hence highly abundant proteins can mask those of low abundance. Proteome analysis has classically been performed by separating proteins with two-dimensional gel electrophoresis, which remains the most widespread method today. More recently, HPLC-based separation methods have entered the stage of proteomics based on a number of methodological and technical advances. While entire proteins are difficult to separate and to recover quantitatively by HPLC, the method is often superior to 2D gels when it comes to low-molecular weight proteins (<20 kDa) and peptides. To apply HPLC efficiently to a proteome, it is advantageous to digest the proteins prior to their separation by HPLC. This approach has been taken by several groups with initial encouraging results. One of the main challenges in doing so is that from a couple of thousand proteins that may be present in a given sample hundreds of...
thousands of peptides are generated — a mixture that no single separation dimension can master. Reduction of this complexity by specific sample pretreatments such as tagging peptides that contain certain amino acids renders the separation problem less daunting but may also lead to a loss of information. The problem at hand and the experimental design will ultimately define whether such an approach is adequate. To deal with highly complex mixtures, it is necessary to combine different separation dimensions in an orthogonal fashion to increase the separation space. Early theoretical considerations of Giddings et al. have laid the foundations for this reasoning. Giddings calculated that two perfectly orthogonal separation dimensions should provide a peak capacity that is the product of the peak capacities of the individual dimensions. While many different chromatographic dimensions can be combined, it is the sequence of ion-exchange with reversed-phase HPLC that has found most widespread use because of the high binding capacity and robustness of commercially available ion-exchangers and the high resolution and on-line compatibility of reversed-phase HPLC with mass spectrometry. Furthermore, ion-exchange and reversed-phase HPLC can be considered to have a high degree of orthogonality, because they are predominantly based on ionic or hydrophobic properties. Such ‘two-dimensional’ HPLC systems have been applied to the analysis of the proteome of the yeast *Saccharomyces cerevisiae*, the genome of which is completely sequenced, to protein complexes isolated from mammalian cells and to biofluids such as urine. Cation-exchangers offer a fairly general approach because they bind almost all peptides present in a mixture under acidic conditions of about pH 3. Anion-exchangers have to be used under fairly basic conditions to capture most peptides where chemical reactions such as deamidations may occur. It is possible to pack a ‘hybrid’ column consisting of a cation exchange part preceding a reversed-phase packing in addition to the more commonly employed approach to having two separate columns that are connected via a switching valve. Adsorbed peptides are eluted from the cation-exchanger using a salt gradient or steps of increasing salt concentration onto the reversed-phase column, where they are further separated and desalted. This combination of columns is well-suited for the analysis of complex peptide mixtures and can be directly coupled to a mass spectrometer with an electrospray ionization source. Most modern HPLC systems allow this configuration to be implemented without the need for major hardware modifications.

Despite these advances in separation science, we are still a long way from performing a comprehensive analysis on peptide mixtures of some hundred thousand components. Two-dimensional HPLC systems have reached peak capacities of about 5000 showing that further challenges lie ahead. Integrating a sample pretreatment step into the two-dimensional HPLC system further facilitates direct analysis of complex biological mixtures such as serum or urine. The example of human haemofiltrate (comparable to urine) shows that individual peptides can be sufficiently enriched for sensitive detection by mass spectrometry. New developments in mass spectrometry instrumentation such as the recently described ion mobility mass spectrometer may help to increase the separation capacity of LC–MS further by combining three of the most powerful separation principles; namely the chromatographic principle (HPLC), the electrophoretic principle (ion mobility MS) and separation based on the mass-to-charge ratio (standard mass analysers). New instruments combining orthogonal separation principles in an integrated way will ultimately provide biologists and medical researchers with tools to penetrate the various proteomes deeper and to discover novel biological mechanisms and mediators. Novel stationary phases for HPLC based on monoliths from polymers or silica as well as targeted sample pretreatments are also developments that may improve the efficiency of LC–MS for proteomics.

Separating proteins and peptides is a prerequisite for successful proteome analysis but it is often not sufficient because quantitative data must be generated to compare one state of a biological system with another (e.g., healthy tissue with diseased tissue). For 2D gels this is done by protein staining and image analysis using, for example, fluorescent dyes. This is the most tedious part of 2D gel analysis in proteomics because image analysis and comparison between gels requires significant experience and time. Furthermore, quantitative image analysis requires highly reproducible gels. Having said this, quantitative comparative proteome analysis is far from being easy using LC–MS because of the vast amount of data and the complexity of the samples. A high level of reproducibility is always required but this might be more easily achieved with an instrumental technique such as HPLC rather than with 2D gels. Based on experiences in the area of quantitative metabolite analysis and bioanalysis, researchers have developed stable isotope labelling techniques that allow a quantitative comparison between different samples. The principle is simple: one sample is labelled with a reagent containing light atoms such as hydrogen, while the sample it is to be compared with is labelled with a reagent containing heavy stable isotopes such as deuterium. Upon mixing of the two samples each analyte should ideally be present in the ‘light’ and the ‘heavy’ form thus serving as its own internal standard. In this way variations in, for example, recovery or ionization efficiency of individual peptides should be corrected for. Recent work indicates that the picture may be more complex because of some chromatographic resolution between the labelled peptides but this depends on the label that is used. Furthermore, application of the ‘internal standard’ technique requires that the response curve of the LC–MS system is linear over the range investigated, a parameter that is often not evaluated in the published literature. It is still too early to say how this relative quantification method compares with the more established image-analysis methods after 2D gel electrophoresis, but initial data show that further improvements are needed, especially with respect to extending the dynamic range. It remains important to have enough separation capacity in the analytical system to avoid too much overlap between peptides, especially if on-line identification of the regulated peptides by MS–MS is required.

LC–MS is becoming a viable alternative to two-dimensional gel electrophoresis in proteomics and plays a prominent role in the analysis of the carbohydrate portion of glycoproteins as biopharmaceuticals.
Combination of two-dimensional HPLC with stable isotope labelling is more amenable to automation and integration with mass spectrometry than two-dimensional gel electrophoresis, a feature that is very attractive to groups applying proteomics to large series of samples. At this point in time, quantification of complex peptide mixtures from proteomic studies is still in its beginnings and more experimental work is needed both at the labelling and the mass spectrometry level. New adaptations of existing database search algorithms and MS control software promise to make this approach more accessible to the average user of LC–MS. Targeted affinity-based analytical systems will probably become more prevalent in the future as researchers start focusing their attention on certain classes of proteins such as kinases, phosphorylated proteins or proteases, and thus reduce sample complexity. The recent success in deciphering multiprotein complexes such as the spliceosome using affinity-tagged proteins has shown the potential of this approach.\textsuperscript{29,30}

**LC–MS in the Analysis of Endogenous Peptides**

Next to proteomics a new term ‘peptidomics’ has been coined to differentiate the analysis of endogenous peptides and low-molecular weight proteins from the more classic proteomics approach, which relies on proteolytic digestion to generate peptides from larger proteins.\textsuperscript{31} Indeed, the analysis of body fluids or tissue extracts for endogenous peptides has a long tradition with major discoveries such as insulin. Peptidomics poses further challenges to both separation methodology and mass spectrometry as the structural variety of components is even higher because of processing events that can no longer be predicted as in the situation of proteolytic digests of proteins where the cleavage site is given by the choice of the enzyme. One goal of peptidomics is the discovery of molecular markers that allows the diagnosis of multifactorial diseases, such as various forms of cancer, chronic inflammatory disorders such as chronic obstructive pulmonary disease (COPD), cardiovascular and metabolic disfunctions.\textsuperscript{32–38} Many of these studies have been performed using surface-enhanced laser desorption mass spectrometry (SELDI-MS), a combination of protein enrichment on a chip with detection by mass spectrometry. LC–MS is a promising alternative to this technique because of its higher separation efficiency in combination with better quality mass spectra. Combining HPLC with on-line MALDI-MS\textsuperscript{39,40} may ultimately also deliver the throughput that is needed to perform clinical studies. A general difficulty with both proteomics and peptidomics is that high abundance proteins often obscure the lower abundance ones. For example, albumin is present in serum at a concentration of 40 mg/mL (~0.7 mM) while cytokines circulate at pg/mL levels. This means that an analytical method for the analysis of cytokines in the presence of albumin must be able to handle proteins over a range of nine orders of magnitude. No separation method nor any detection method spans such a wide range. Most low-abundance proteins and peptides are therefore analysed with highly specific immunological methods that rely on antibodies. While this is a viable approach if the analyte is known, it is not a suitable method if samples need to be compared in an effort to identify new biomarkers or biomarker patterns. Sample pretreatment is therefore critical if low-abundance markers shall be found in a sea of high-abundance proteins, no matter what the separation method.

HPLC offers a number of possibilities when it comes to sample pretreatment. Because of the modular nature of most modern HPLC instruments and the availability of sophisticated liquid-handling devices, it is possible to incorporate sample pretreatment steps into an analytical scheme. Such steps range from simple size-exclusion steps to approaches in which size-exclusion is combined with chromatographic adsorption in so-called restricted access materials (RAMs).\textsuperscript{41,42} Pretreatment of serum with a RAM cartridge having a hydrophobic internal reversed-phase surface and a hydrophilic non-adsorbing external surface removes most of the albumin from serum and leaves additional capacity for lower molecular weight analytes (Figure 1). This allows the HPLC system to reach a higher loading capacity for peptides and a better peak capacity. Alternatively, highly specific elimination strategies for certain high-abundance proteins are being developed. These are mostly based on specific antibodies immobilized on chromatographic media which can be used in an on-line integrated or off-line fashion. Recent results from our group indicate that these procedures are efficient and selective although further studies are needed to confirm the initial findings, especially with respect to the robustness of the affinity matrices.

**Data Handling and Analysis**

Data handling and analysis is another critical aspect of proteomics in general and LC–MS analysis of proteins and peptides in particular. An LC run with on-line acquisition of mass spectra may easily produce 200–400 MB of data files with thousands of MS and MS–MS spectra. Considering that a two-dimensional HPLC run comprises at least 10 regular runs, one is rapidly in the gigabyte range when analysing just one sample. Data reduction and comparison are therefore of great importance to extract information out of a vast amount of data. Comparing large data sets such as two-dimensional gel images or LC–MS runs requires rapid and efficient ways of focusing on the significant differences while being able to
detect and reduce the influence of experimental variations and trends that may be caused by phenomena not related to the study object. While it is beyond the scope of this article to discuss the various developments in this area, it is necessary to highlight a few approaches that are currently being implemented into mass spectrometry data analysis software. The basic requirement for a successful LC–MS run is that the mass spectrometer can be triggered to switch from MS to MS–MS mode to acquire partial sequence information from the separated peptides based on a meaningful measured parameter. Partial sequence information is often needed to identify the peptide or the protein from which it was derived by database searching. While triggering based on a given precursor ion exceeding a set threshold is of interest, it is far from sufficient in the situation of highly complex protein digests even if two-dimensional HPLC is used and further improvements are needed. There are a number of different algorithms that allow the searching of protein or DNA sequence databases based on MS and/or MS–MS data (see Proteomics Tools at http://www.expasy.ch). Although protein identification based on mass spectrometric data is now fairly well established for tryptic digests, in which all peptides end in a lysine or arginine residue, it is still tricky with peptide mixtures from body fluids or other natural sources. Further refinement of the software tools in combination with an increasing spectral quality will improve automatic spectra interpretation and allow fully integrated automatic LC–MS on proteomics and peptidomics samples to be run.

**Future Outlook on LC–MS in Proteomics**

Biological systems do not generally react to perturbations by regulating just a few components but through changes in many interacting pathways. It is becoming clear that, for example, multifactorial diseases cannot be diagnosed based on single biomarkers but that patterns of markers are necessary for a prognosis as well as to follow therapeutic interventions. Chip-based protein analytical systems as well as LC–MS approaches may provide the technological means to do so in the future. Recognizing patterns is pivotal to successful proteomics or peptidomics experiments. Pattern recognition has been the subject of intensive study in mathematics and informatics but has not yet found widespread applications in the biosciences. Recently, several efforts have shown that integration of analytical data obtained from a biological system at various levels can lead to a better understanding of how it reacts to a given perturbation. Proteomic data possibly generated by LC–MS constitute one level of such an analysis. The upcoming field of the large-scale analysis of metabolites (so-called ‘metabolomics’) also relies heavily on LC–MS as an analytical method. Metabolomics is now being integrated into the analysis of biological systems and provides additional insights.

It is without doubt that LC–MS as a method will see further growth in the areas of proteomics and peptidomics. New mass spectrometers and the associated software facilitate automated spectra acquisition and analysis while search algorithms adapt to the new methodological advances such as stable isotope labelling. The completion of sequence databases with additional organisms whose genomes are fully sequenced will provide the blueprint for successful protein and peptide identification. New developments in the area of chromatographic separation media, such as monolithic columns, will open more possibilities for combining different dimensions in an orthogonal fashion to enlarge the separation space. Last, but not least, chromatography instrumentation is continuously being developed and it is now possible to run \( \mu \text{L}/\text{min} \) flow-rates routinely while more recent developments even allow below the \( \mu \text{L}/\text{min} \) flow rate to be worked. All this is not without difficulty, because miniaturized systems are always more prone to technical problems rendering them less robust than their macroscopic counterparts. Sample preparation becomes more important as the dimensions of the analytical system are being diminished but exquisite sensitivities can be achieved. Advanced proteomics technologies thus require interdisciplinary specialists that can understand the biological questions and define the most appropriate analytical approach. We expect that LC–MS and two-dimensional gel electrophoresis will remain two complementary separation techniques with many applications in proteomics. Sample pretreatment will always be a critical point in any analytical scheme and it can be expected that there will be many specific approaches dependant on the problem. Namely the elimination of high-abundance proteins and the enrichment of proteins of interest are topics of continuous development. It is impressive to see how LC–MS has made an inroad into proteomics in recent years and this trend is likely to continue.

**LC–MS in Carbohydrate Analysis**

Glycosylation, one of the most common post-translational modifications in proteins found in nearly all biological systems, consists of the covalent attachment of saccharides to specific amino acid residues in peptides. There are two main types of glycosidic linkages in glycoproteins: the \( \text{N} \)-glycosidic linkage via the amide nitrogen of an asparagine residue and the \( \text{O} \)-glycosidic linkage via the hydroxyl group of serine, threonine, tyrosine, hydroxylsine or hydroxyproline.

The carbohydrate components of glycoproteins play crucial roles in the modulation of their physicochemical properties. The development of recombinant-derived glycoproteins for therapeutic use has led to an increasing demand for methods to characterize their carbohydrate structures. They influence protein folding, oligomer assembly and secretion processes, as well as the clearance of glycoproteins from the bloodstream. Certain carbohydrate structures have also been found to be antigenic and regulatory authorities such as the US Food and Drug Administration are demanding increasingly sophisticated carbohydrate analysis as part of the product or process validation. Therefore, determination of the glycan structures is important for full characterization of a protein and subsequently for understanding its properties and functions.

Different strategies can be followed for the analysis of the carbohydrate chains attached to a glycoprotein. One of them consists of the chemical or enzymatic release of the oligosaccharides and their subsequent fractionation by chromatography or electrophoresis followed by the characterization of the separated fractions. The separation of glycans cleaved from a glycoprotein yields a fingerprint, also referred to as oligosaccharide profile or map.

In the pharmaceutical and biotechnology industry, mapping techniques for glycoprotein carbohydrate structures have been
The development of recombinant-derived glycoproteins for therapeutic use has led to an increasing demand for methods to characterize their carbohydrate structures.

used

- during initial characterization, for comparison of the glycosylation of the native and recombinant protein
- to separate and identify the oligosaccharide structures present
- to monitor consistency of glycosylation and identify changes that may have resulted from alterations in cell culture conditions or during the manufacturing process
- to monitor changes in glycosylation that occur as a result of expression in different cell lines.

As mentioned before, for glycoprotein oligosaccharide mapping, chromatographic or electrophoretic techniques are usually employed. High performance anion exchange chromatography with pulsed amperometric detection (HPAEC–PAD) is one of the most widely chosen techniques because it allows rapid and direct separation of underivatized samples. It separates oligosaccharides according to charge, size, monosaccharide composition and unit linkage. The great attractiveness of HPAEC stems from its compatibility with pulsed amperometry that provides a sensitive and selective detection method for molecules lacking chromophores or other easily detectable functional groups, such as carbohydrates. In contrast, most of the other methods used for oligosaccharide analysis based on HPLC or CE require derivatization for improving separation and/or detection. Different reagents, most of them fluorescent labels, such as 8-aminoanphthalene-1,3,6-trisulfonic acid, 50 1-phenyl-3-methylaminonaphthalene-1,3,6-trisulfonic acid, 48 2-aminoacridone, 49 different reagents, most of them fluorescent labels, such as 8-aminophthalene-1,3,6-trisulfonic acid, 50 1-phenyl-3-methylaminonaphthalene-1,3,6-trisulfonic acid, 48 2-aminoacridone, 49 8-aminoacridone, 49 and 8-aminophthalene-1,3,6-trisulfonic acid, 50 1-phenyl-3-methylaminonaphthalene-1,3,6-trisulfonic acid, 48 2-aminoacridone, 49 and 8-aminophthalene-1,3,6-trisulfonic acid, 50 1-phenyl-3-methylaminonaphthalene-1,3,6-trisulfonic acid, 48 2-aminoacridone, 49 are commonly used. Retention of native complex carbohydrates on reversed-phase LC columns is quite poor because of their high polarity but the use of porous graphitized carbon as an alternative chromatographic material has been shown to provide good resolution of glycans. 53

After separation oligosaccharides can be characterized through coelution with standards or through exoglycosidase treatments and retention-time comparison. Empirical relationships between oligosaccharide structure and chromatographic retention have been described but for real identification, the separated oligosaccharides are frequently collected postcolumn for further structural analysis using mass spectrometry and/or NMR. 54 Alternatively, MS can be coupled on-line with any of these chromatographic techniques although in the situation of HPAEC the incorporation of specially designed membrane desalters is mandatory. 55 Figure 2 displays the HPLC-TIC chromatogram obtained after injection of two N-linked oligosaccharides in a porous graphitized column and Figure 3 demonstrates their corresponding spectra that provide structural information.

All these mapping procedures described until now, in spite of

**Figure 2:** HPLC-TIC chromatogram corresponding to the injection of two N-linked oligosaccharides in a porous graphitized carbon column (Hypercarb 5 μm, 1 mm i.d. × 100 mm length). Mobile-phase composition: (a) = 100 mM ammonium acetate pH 8.0, (b) = acetonitrile. Elution performed with an increasing gradient of 1% B per min. Flow-rate 20 μL/min. (unpublished data).

**Figure 3:** Spectrum corresponding to the two oligosaccharides (a,b) observed in the chromatogram in Figure 4. (a) is a mono-sialylated diantennary N-linked oligosaccharide and (b) is a disialylated biantennary fucosylated N-linked oligosaccharide.
their broad applicability, present an important drawback: the loss of sequence context for each family of glycoforms.

In this instance reversed-phase liquid chromatography is the technique of choice. C18 or carbon graphitized LC columns provide the best separation of the glycopeptides. Differentiation of N- and O-linked carbohydrates can be performed by submitting the sample to peptide N-glycosidase F digestion.

When coupled to ESI-MS, this technique allows location of the glycopeptides by monitoring selected oxonium sugar ions thus providing information on the attachment site 56, 57 (Figure 4). If good quality spectra are obtained (high-resolution instruments) structural information can be extracted. The mass shift observed following specific glycosidase digestions allows further confirmation of the structure.

Mass spectrometry has developed into a method often used to study glycopeptides and oligosaccharides. Initially MALDI-TOF was the most popular technique to analyse protein oligosaccharides but nowadays ESI-TOF is increasingly used.

Interpretation of mass spectrometric data can be laborious, but with the appearance of new computational tools is becoming easier. One example is GlycoMod, included in the proteomics suite of tools available on Expasy (www.expasy.ch/tools/glycomod). GlycoMod is linked to the information available in the SWISS-PROT database and enables the prediction of N- and O-linked oligosaccharides on glycopeptides using experimentally obtained mass data. It also allows a range of options of oligosaccharide structures and release and derivatization strategies to be included in the calculation.

Future Outlook on LC–MS in Carbohydrate Analysis

The detection, separation, identification and sequencing of sugars from glycoproteins are complex procedures that require different analytical strategies. The development of a generic oligosaccharide sequencer in a way similar to what is available for proteins and polynucleotides is still not a reality.

In the future, the combination of LC with MS will probably become the most widely used analytical tool for the analysis of protein-derived oligosaccharides, in spite of the restrictions of MS such as the impossibility to distinguish between monosaccharides with the same mass (glucose, mannose, galactose etc.) or the fact that different combinations of monosaccharides result in the same mass. To overcome these limitations, the combination of MS with several chromatographic techniques, as well as various steps of enzymatic degradation can be valuable.

References

In the future, the combination of LC with MS will probably become the most widely used analytical tool for the analysis of protein-derived oligosaccharides, in spite of the restrictions of MS...